How Do In-video Interactions Reflect Perceived Video Difficulty?

Nan Li, Łukasz Kidziński, Patrick Jermann, Pierre Dillenbourg
CHILI Lab, EPFL, Switzerland
firstname.lastname@epfl.ch

Abstract: Lecture videos are the major components in MOOCs. It is common for MOOC analytics researchers to model video behaviors in order to identify at-risk students. Much of the work emphasized prediction. However, we have little empirical understanding about these video interactions, especially at the click-level. For example, what kind of video interactions may indicate a student has experienced difficulty? To what extent can video interactions tell us about perceived video difficulty? In this paper, we present a video interaction analysis to provide empirical evidence about this issue. We find out that speed decreases, frequent and long pauses, infrequent seeks with high amount of skipping and re-watching indicate higher level of video difficulty. MOOC practitioners and instructors may use the insights to provide students with proper support to enhance the learning experience.

Introduction

Massive Open Online Courses (MOOCs) have swept online education into the mainstream. Most popular MOOC providers, exemplified by Coursera, Udacity and edX, are featured with video lectures, quizzes, tutorials, discussion forums and wikis. With thousands of learners taking MOOCs, learning analytics is making a big leap forward. One typical usage of such big educational data is the attempt to model students' learning behaviors in terms of their social engagement (Brinton, Chiang, Jain, Lam, Liu & Wong, 2013) and video interactions (Kim, Guo, Seaton, Mitros, Gajos & Miller, 2014). Such models are then used for predicting students' dropout (Halawa, Greene, & Mitchell, 2014; Sinha, Li, Jermann & Dillenbourg, 2014) and performance (Jiang, Warschauer, Williams, O'Dowd & Schenke, 2014), analyzing demographics (Guo & Reinecke, 2014) and engagement (Kizilcec, Piech & Schneider, 2013) etc.

Since video lectures are the major means for delivering MOOC content, analyzing video interactions has received much research attention. Existing research involving video analysis typically only takes into account macro-level video activity features, such as the number of videos watched (Anderson, Huttenlocher, Kleinberg & Leskovec, 2014), engaging time (Guo & Reinecke, 2014) as well as the navigation styles (Guo, Kim & Rubin, 2014), except for few recent attempts that scale the analytics down to the click-level (Halawa, Greene & Mitchell, 2014; Kim, Guo, Seaton, Mitros, Gajos & Miller,2014; Sinha, Li, Jermann & Dillenbourg, 2014). Compared to macro-level video activity features, click-level in-video analysis allows MOOC instructors to closely examine how a student interacts with each video lecture, e.g. what types of video interactions are employed, when they happen and how intense they are. This information further reflects the students' learning status, such as encountering problems during watching the videos, which may provide opportunities for the MOOC practitioners to design proper interventions to support the students' learning experience.

However, existing MOOC research literature lacks click-level video analysis that helps us infer students' mind states, e.g. the perceived difficulty of different videos. Our research aims to fill this research gap by providing an empirical investigation of different types of video interactions. Our key research questions are:

(1) Do in-video interactions reflect students' perceived video difficulty?

(2) If yes, then how do video interactions of different types and intensities correlate with the difficulty?

By answering these questions, we will gain insights on the effects of video interactions, which has the potential to help the instructors identify the videos that a particular student may have trouble with. In the long run, detecting problems and providing support is likely to increase the students' engagement with MOOC and reduce the dropout rate.

In-Video Interaction Analysis

This paper involves analyzing in-video interactions. Such interactions typically only include limited types of actions, each of which is associated with a time span. Researchers have started investigating in-video interactions well before the MOOC era. In the mid-1990s, Dey-Sircar et al. applied Markov model to develop a tool for evaluating video system designs (Dey-Sircar, Salehi, Kurose & Towsley,1994). Research work (Shenoy & Vin, 1995) and (Li, Liao, Qiu & Wong, 1996) both apply Markov chains to formulate an effective way of fast-forward/rewind service. These early research works mainly focused on video systems and quality-of-service issues. Research that attempted to understand and model video click behaviors did not come to light until late 1990s, when 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 (Branch, Egan & Tonkin, 1999). In order to identify interesting video segments reflected by the users' video interactions, Syeda-Mahmood et al. designed a user study with MediaMiner system (Syeda-Mahmood & Ponceleon, 2001), where they trained a Hidden Markov Model to learn from the ground-truth browsing states explicitly indicated by users. The goal was to generate video previews that best represent interesting video segments. All of the above studies were conducted in the time when the control menu of the video players were restricted to only continuous interactions, lacking discontinuous interactions (i.e. allow jumping between discontinuous time positions) that are common in today's player controls, such as seeking forward/backward. These discontinuous interactions were only included for analysis in a recent study by Gopolakrishnan (Gopalakrishnan, Jana, Ramakrishnan, Swayne & Vaishampayan, 2011). However, compared to the continuous interactions, the duration of discontinuous interactions has different meanings, i.e. they refer to video player time instead of the actual time spent on the states, since these actions last for a negligible amount of time, but may result in skipping a large amount of video content.

Method

Our study is based on two undergraduate MOOCs offered in Coursera: “Reactive Programming (RP)”, which covers advanced topics in programing and ”Digital Signal Processing (DSP)”, which is an entry-level Electrical Engineering course. An in-video survey is placed at the end of each video (See Figure 1) during the enactment of the courses. 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 answered by the learners right after they finished watching the video, providing ground-truth knowledge that allows us to reveal the hidden relationship between the video interaction and the perceived video 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”.

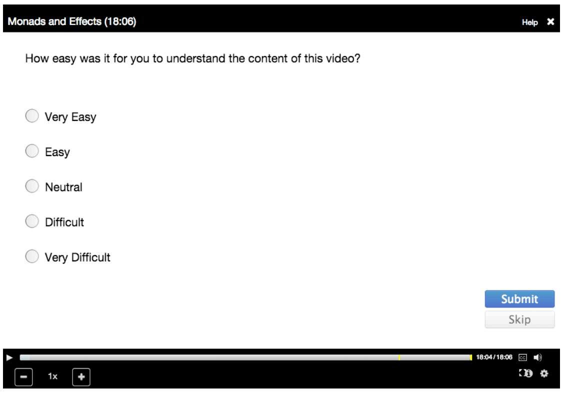

Figure 1. End-video surveys about perceived video difficulty.

On the MOOC platform, the students are allowed to watch videos as many times as they wanted, but the number of views per video varies. Table 1 lists a descriptive overview of the video sessions of the two courses separated by the video visiting time (either first-time visit or revisit). As students who find videos more difficult are more likely to revisit them, the average difficulty in the revisiting sessions is higher. In this study, we do not investigate video revisiting behaviors and will only focus on the first-time visiting video sessions.

Table 1: Overview of the two Coursera coursers in our dataset.

Course	Videos	Active learners	Visit	Sessions	Response rates	Difficulty
RP	36	22,794	**First**	**265,493**	**49.1%**	**2.699**
			Revisiting	205,501	23.7%	2.837
DSP	58	9,086	**First**	**58,349**	**32.8%**	**2.478**
			Revisiting	59,610	12.7%	2.593

Many students left in the middle of the videos, leading to the so-called in-video dropouts. During the enactment of the two courses, Coursera did not always successfully log the time when in-video dropouts occurred. We filtered out the video sessions that did not contain any video interactions in the last 10 percent lengths of the videos. This eliminates the sessions with early in-video dropouts. The remaining dataset is our target, which contains only the video sessions where the students almost reached to the end. We also removed data entries containing inconsistent timestamps or event types. Finally, we keep 188,138 video sessions with 79.0% survey response rate for the RP course, and 28,994 sessions with 60.8% response rate for the DSP course.

Video Interaction Profiles

Coursera supports four different types of video controls, namely, play/pause (toggle), seek forward, seek backward and adjust video speed. Multiple types of interactions can coexist in the same video sessions, and the effects may interact with one another. For example, a student may pause a few times and increase the play speed in the same video. Then the rated video difficulty may relate to the effects of both interactions. Any of the four interactions can coexist, resulting in complex interactions. As an initial exploration, this paper aims at understanding each individual type of interactions, so analysis of complex combinations are left for future analyses.

We divide the dataset into subsets of interaction profiles. Each of the four video controls is associated with a simple video interaction profile, which we name as *pausing, skipping, replaying* and *speeding* respectively. Video sessions with a combination of interactions are called *mixed-interacting*. In addition, the video player in Coursera is found to maintain cross-video video speed consistency. That is, when a user changes the speed for a video, the new speed is kept as default for her subsequent video sessions. In other words, video sessions without speed changing events may be found to start with higher or lower speed. This phenomenon introduces a new *speeding* profile called *implicit-speeding*. The sessions with explicit speed changing events are referred to as *explicit-speeding*.

By naming the sessions without any video interactions as *silent*, we finally obtained 6 simple and 1 complex interaction profiles (*mixed-interacting*) as summarized in Table 2. The three rows (up to down) describe the number of video sessions, the proportion of videos and the average difficulty for each profile.

Table 2: Descriptive Statistics of Interaction Profiles.

Course	Non-Interactive		Interactive				
	Silent	Implicit-Speeding	Explicit-Speeding	Pausing	Skipping	Replaying	Mixed-Interacting
RP	35,873	9,896	6,140	27,792	2,856	5,485	71,102
	22.54%	6.22%	3.86%	17.46%	1.79%	3.45%	44.68%
	2.61	2.64	2.34	2.72	2.52	2.73	2.78
DSP	5,479	885	527	5,545	1,084	905	14,569
	18.90%	3.05%	1.82%	19.12%	3.74%	3.12%	50.25%
	2.51	2.41	2.30	2.43	2.64	2.60	2.51

Note that we only consider whether or not specific types of video interactions occur in the sessions, e.g. the video sessions with *replaying* profile contain only backward seeks but nothing else, regardless of how much content is replayed. For the pausing profile, we ignore pauses shorter than 2 seconds (probably accidental pauses) or longer than 10 minutes (long breaks). Automatic pauses generated by in-video quizzes are also not considered. In Table 2, three interaction profiles stand out. In both courses, nearly half of the video sessions contain more than a single type of video interactions (*mixed-interacting*); one fifth of the video sessions (*silent*) contain no interactions at all; pause (*pausing*) is the most frequently used video interactions. Table 2 also shows that the interaction profiles reflect different perceived video difficulty: the *explicit-speeding* profile indicates the least perceived difficulty, while *pausing*, *replaying* as well as *mixed-interacting* inform that the students may have more difficulty.

In the upcoming sections, we will extract video interaction features for the simple profiles and build regression models to deeply investigate the relationship between video interactions and perceived video difficulty. Note that the perceived difficulty may depend on user-specific characteristics. Having several observations per user in our dataset allows us to adopt a mixed model, where *user* is modeled as a random effect. Mixed models are known to be robust to missing values and unbalanced groups. In addition, least-square means (hereafter referred as LS means) mimic the main-effect means but are adjusted for group imbalance. These methods are used throughout our analysis. We will only report the analysis of the RP course due to its larger size, however the results for the DSP course are analogous.

Implicit-Speeding Profile

Video sessions of *implicit-speeding* do not contain video interaction events, but are started with an initial speed other than 1.0. As discussed before, the initial speeds are inherited from previous sessions. It is arguable that the choice of video speed may depend on the students' language skills or personal preferences. However, in the video sessions with very high or low speed, we find the voices are significantly distorted. We then argue the speed may also relate to other factors, such as the perceived video difficulty.

In this section we attempt to model the effect of initial speed. Coursera video player offers 7 levels of speed from 0.75 to 2 with a stepwise change of 0.25. We compute the LS means for the video sessions with different initial speeds and show the means with confidence intervals in Figure 2. The two numbers separated by a slash ("/") under each category are respectively the number of survey responses and the total number of video sessions in the corresponding category.

Finding 1: Implicit-speeding shows a negative linear effect on the perceived video difficulty.
Figure 2 shows a linear relationship. Considering the levels are numeric, statistically we assess the effects with a mixed linear model, which shows significant negative effects ($\beta = -0.08$, 95% CI = [-0.10, -0.05], $p < .0001$). That is, an increase of 0.25 video speed results in an average decrease of perceived difficulty by 0.08.

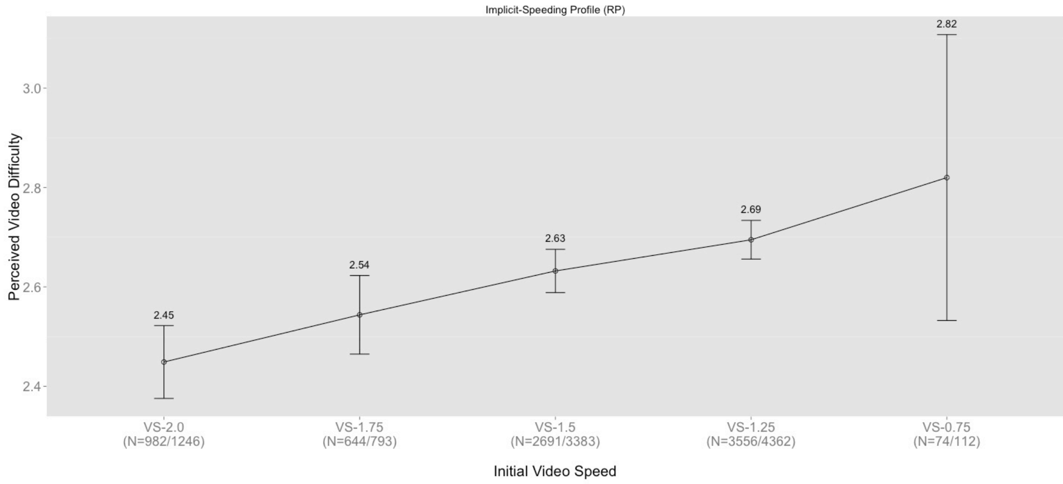

Figure 2. LS Means of Video Session with Different Initial Speed

Explicit-Speeding Profile

Compared to the *implicit-speeding* sessions, students explicitly change the video speed in the *explicit-speeding* video sessions. We hypothesize that the following features may relate to the perceived difficulty:
(1) Frequency of effective speed ups/downs. *When we count the frequency of speed change events, we group the speed ups/downs that happen within 10 second as a single event, because the learner may simply try out different speeds during this period. We do it for both up and down events, and call these new events effective speed ups/downs.*
(2) Amount of average speed change. *In order to obtain the average speed change, we first compute the average speed for the video session, which is the weighted arithmetic mean of the video speeds at all video seconds. Then, we subtract the initial speed from this average to obtain the amount of average speed change.*

Considering the initial video speed may influence the analysis of these features, we analyze the video sessions that are started with 1.0, which are the most common observations (36%) for this profile. Both features are continuous, and are distributed empirically lognormal. The average speed changes are of ratio values ranging between -0.25 and 1.0, while the 99% frequencies are integers less than 8. The relationships between the features and the perceived video difficulty are not necessarily linear, so we build Generalized Additive Mixed Models (GAMM) for the analysis. Compared to Generalized Linear Models (GLM), GAMM fits the data points with a spline smoother, which is able to capture non-linear relationship. Our reported statistics will include the estimated degrees of freedom (edf) together with the p-value of an F-test that test whether the smoothed function significantly reduced model deviance. This GAMM modeling technique will be used primarily throughout this paper for analyzing continuous, widespread features.

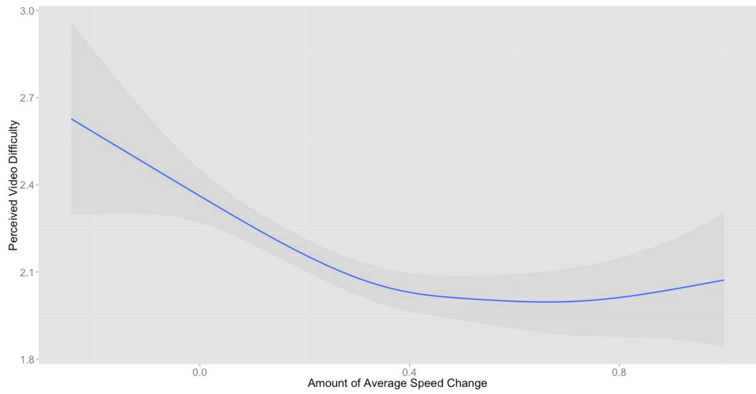

Figure 3. GAMM Fit for Amount of Average Speed Change with Confidence Interval Band

Finding 2: Speed-down frequency has a positive linear effect, while the amount of average speed increase has a monotonically negative effect till saturation point 0.4

Our multiple GAMM regression model with the frequency of speed ups/downs and average speed change as explanatory variables shows that the frequency of the speed-ups does not significantly correlate with perceived difficulty (p=0.73), but the other two have significant effects. The effect of speed-down frequency is positively linear (β = 0.06, 95% CI = [0.02, 0.09], p < .005), while the effect of average amount of speed change is non-linear with (edf = 2.683, p < .0001) as depicted in Figure 3.

As for the finding of the speed changing frequency, it should be noted that the analysis is based on a subset of video sessions that are started with speed 1.0. There are more higher speed options than lower in this condition. However, we find out that the effect of the speed-up frequency is not significant, while the frequency of change in the other direction is significant. In fact, frequenter speed-downs are only possible if the video speed had already been increased before. Keeping the amount of speed change constant, it is interesting that the events that revert the video speed are stronger indicator of higher video difficulty.

As expected, the amount of speed change negatively correlates with the perceived video difficulty. This effect is only prominent when the amount is less than 0.4, after which the effect saturates. Further increases do not significantly reflect the changes in perceived difficulty.

Pausing Profile

Similar to the previous modeling process, for the *pausing* profile we hypothesize that the frequency and duration of pauses may relate to the perceived video difficulty.

(1) Frequency of pauses. *As discussed previously, we take pauses that last between 2 seconds and 10 minutes into account only. In fact, numerous pauses shorter than 2 seconds or across several days are observed in the dataset. The extremely short pauses do not make much sense in terms of cognitive processing, while the long ones may indicate the learners are taking breaks, when they are out of the cognitive processes for video comprehension. The choices of 2 seconds / 10 minutes as thresholds are arguably arbitrary, i.e. it is difficult to argue why 3 second or 11 minutes are not chosen, but we have tried slightly different values and achieve results that are robust to the choices of thresholds.*

(2) Median duration of pauses. *The durations of pauses distribute exponentially with long tail. We then use the median of pause duration to gauge the time dimension of pauses. This statistic is more robust compared to mean or sum under the given data distribution.*

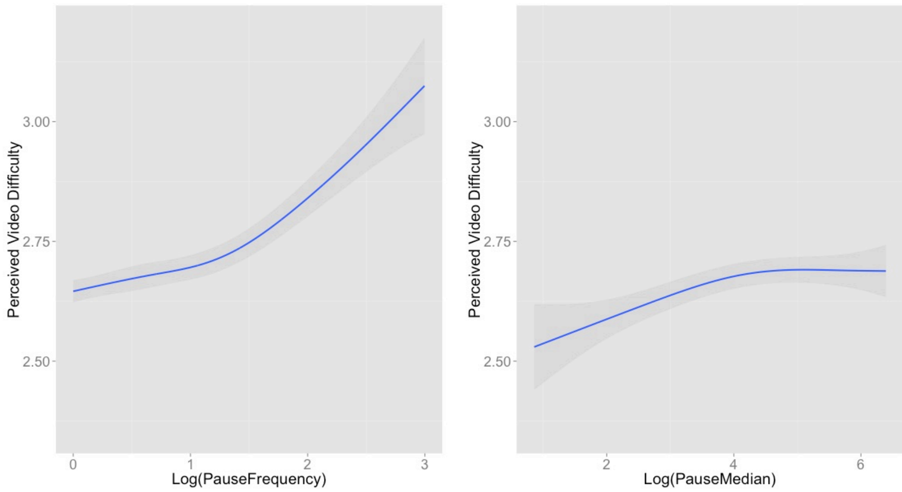

Figure 4. GAMM Fit for Pausing Profile with Confidence Interval Band

In fact, the data distributions of both features are highly skewed with long tail, so logarithm transformations (natural base) are used before fitting the GAMM model.

Finding 3: Pause Frequency matters more than duration

The pause frequency (edf = 3.14, p < .0001) and the pause median duration (edf = 2.439, p < .0001) both show significant non-linear effects on perceived video difficulty when fitted in multiple GAMM regression as covariates. The GAMM fit is illustrated in Figure 4. We can see that the effect of pause frequency has visually larger slope over the pause median duration. Lots of video sessions are found to contain high pause frequency (e.g. more than 10), where the students may constantly encounter problems in the videos. Note that the curve for median pause duration stabilized and maximized at around 4.1, which corresponds to 60 seconds. This indicates that the pauses longer than 1 minute reflect on average a similar level of perceived video difficulty.

Skipping Profile

For the video sessions of *skipping* profile, we evaluate the following two features:

(1) Frequency of forward seeks. *A seek event is created when the user scrubs the playhead to a new position or click a new position on the time indicator. In the case of scrubbing, the system automatically generates a number of intermediate seeking events. In this analysis we analyze the number of raw forward seek events.*

(2) Skipped video length. *The skipped video length refers to the amount of video seconds skipped by forward seeks. Some video sessions contain less than 10% unwatched content due to slightly earlier video closing, but they are not considered when we compute this feature.*

These two features are also highly skewed with long tail, so we apply natural logarithm before fitting the model.

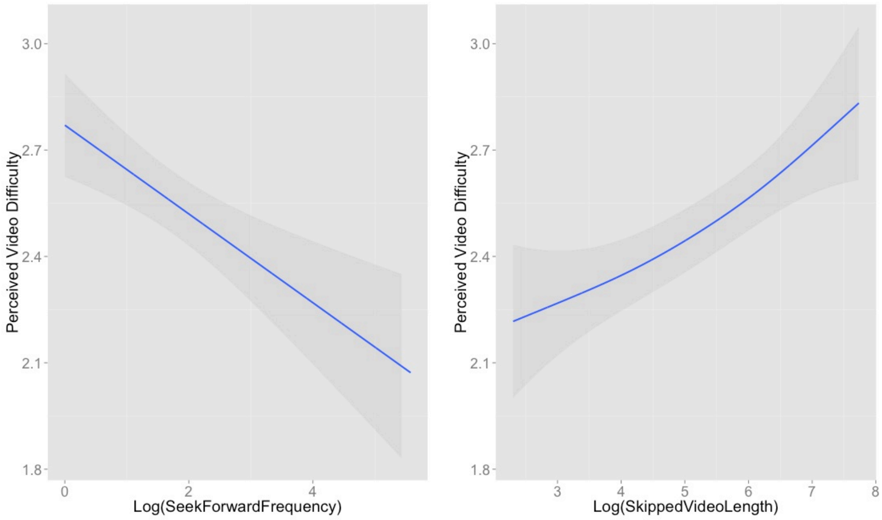

Figure 5. Model Fit for Skipping Profile with Confidence Interval Band

Finding 4: Infrequent or large skip suggests higher perceived video difficulty

The forward seeking frequency shows a negative linear effect (β = -0.13, 95% CI = [-0.19, -0.06], $p < .0005$). This is not surprising since we anticipate the students to "jump" forward more often when they think the videos are easier. In practice, frequently seeking forward leads to video skimming, which can be seen as an alternative way for speeding up the video. Recall the results in the analysis of the speeding profile: largely decreasing the video speed has a similar negative correlation with perceived video difficulty. The students who interact in this way may find skimming through the content sufficient for understanding the video.

On the other hand, when we keep the seek frequency constant, we find the skipped video length exerts a positive non-linear effect (edf = 1.56, $p < .0005$). The estimated degree of freedom is quite close to 1, so the latter effect is close to a negative linear result, as depicted in Figure 5. This finding contradicts our expectation that more skipped content may indicate the videos are boring and easy. In fact, this may be an indication of higher difficulty. So, if highly frequent forward seeking is for quickly grasping the gist of the video, then large skipping may be "giving up" the video.

Replaying Profile

The video sessions of *replaying* profile are analyzed in a similar way as we did for the *skipping* profile. The following two features are analyzed:

(1) Frequency of backward seeks.

(2) Replayed video length. *The replayed video length refers to the video seconds that are re-watched. Note that the same parts of video can be watched several times. This measure accumulatively sums the replayed video length.*

Similar to our analysis of the *skipping* profile, these two features are highly skewed. We take the natural log of the original features and model them with multiple GAMM regression.

Finding 5: Less frequent or large amount of re-watching indicates higher video difficulty

The replayed video length shows a positive effect on the perceived difficulty (edf = 2.20, $p < .0001$) as depicted in Figure 6 (Right). We can see that the curve has a monotonically sharp increasing trend until the value on the x-axis reaches around 6, which translates to replayed video length of 300-400 seconds. Afterwards, the curve bends down a little bit and the confidence interval band starts expanding. This shows that the more a student replays the video, the more she may perceive the video as difficult. The effect is shown stronger when the replayed length is less than 5 minutes. The finding generally coincides with our expectations.

However, keeping the replayed video length constant, the frequency of backward seeks surprisingly shows a significant, though visually small negative effect on the perceived video difficulty (edf = 1.36, $p < .0005$). We have similar findings in the DSP dataset as well. This is interesting, since it suggests that higher average replayed length per seek reflects higher video difficulty. In addition, we found in the sessions containing high number of frequent backward seeks, the events typically occurred within very short intervals, indicating that the students were deliberately looking for specific video frames. In this case, the behavior can be seen as more of "frame-seeking" than "re-watching".

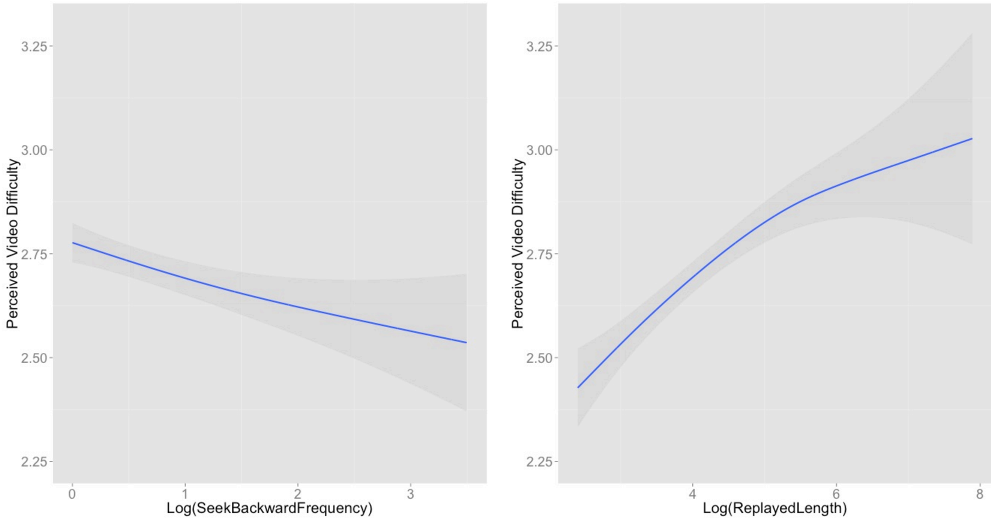

Figure 6. GAMM Model Fit for Replaying Profile with Confidence Interval Band

Discussion and Conclusion

So far we have employed statistical methods to reveal the relationships between video interactions and perceived video difficulty. We identified simple video interaction features that indicate students' perceived video difficulty. With the mixed model analysis, we can infer the changes of subjective video difficulty of a student from video to video, based on the changes of the aforementioned features. To summarize, video speed decreases, frequent and long pauses, infrequent seeks with high amount of skipping or re-watching suggest higher video difficulty. These findings have answered the main research questions of this paper.

Limitations

Although the results presented in this paper are statistically significant, the magnitudes of the effects are small, i.e. we have not seen the average perceived difficulty changes drastically within the variation range of the presented video features. There are many possible reasons. First, students study MOOCs with various motives, educational background, personal characteristics, habits and learning strategy, and all of these factors may also explain much of the variance in the perceived video difficulty. Second, the students can externalize the perceived difficulty in alternative ways, e.g. instead of adopting a video interaction style, they may choose to tackle the problem in the forum or search in the internet after watching the videos etc. Third, in this paper we only analyze video behaviors of the first-time watching sessions, video-revisiting behaviors would probably also influence the perceived difficulty of the current or next couple of videos. Last but not least, the perceived video difficulty was posteriorly measured per video, however the video interactions occurred at different parts of different videos. In this paper, we pursue to generalize the effects of the video interaction features rather than video content features, so we did not analyze the video content, which definitely also exert effects.

Impact of the Results

The majority of existing MOOC research focus on predicting students' dropouts or performance, whose relationships to video behaviors may not be causal. Factors such as learning motives and online learning experiences may confound in between. Since lecture videos play a central role in MOOC learning, how students perceive the videos is an important measure of learning experiences. Despite the limitations presented before, the greatest impact of this paper is the empirical revelation of the variation trends of the perceived video difficulty from the perspective of video interactions. There are usually dozens of lecture videos for each MOOC. The findings in this paper have the potential to help detect how the students may perceive the video difficulty differently among the lecture videos. MOOC practitioners may use the insights to capture the students' potential change of difficulty perception from their video behaviors, so that proper interventions (e.g. supporting materials) can be introduced for help

Future Work

Within the scope of this paper, we only analyzed simple video profiles. Future work may include analyses of the complex video profile, which contains multiple types of video interactions. Clustering the data with video features into groups of similar patterns is a tentative method for analysis. In addition, more comprehensive insights about how the students learn require combing video interaction analysis with analyses of other MOOC interactions, such as those happened in the form and quizzes. Video-revisiting behaviors can also be examined.

References

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014, April). Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web* (pp. 687-698). International World Wide Web Conferences Steering Committee.

Branch, P., Egan, G., & Tonkin, B. (1999). Modeling interactive behaviour of a video based multimedia system. In *Communications, 1999. ICC'99. 1999 IEEE International Conference on* (Vol. 2, pp. 978-982). IEEE.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F. M. F. (2013). Learning about social learning in moocs: From statistical analysis to generative model. *arXiv preprint arXiv:1312.2159*.

Dey-Sircar, J. K., Salehi, J. D., Kurose, J. F., & Towsley, D. (1994). Providing VCR capabilities in large-scale video servers. In *Proceedings of the second ACM international conference on Multimedia* (pp. 25-32). ACM.

Gopalakrishnan, V., Jana, R., Ramakrishnan, K. K., Swayne, D. F., & Vaishampayan, V. A. (2011). Understanding couch potatoes: measurement and modeling of interactive usage of IPTV at large scale. In*Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference* (pp. 225-242). ACM.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50). ACM.

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 21-30). ACM.

Halawa, S., Greene, D., & Mitchell, J. (2014) Dropout Prediction in MOOCs using Learner Activity Features.

Jiang, S., Warschauer, M., Williams, A. E., O'Dowd, D., & Schenke, K. (2014). Predicting MOOC Performance with Week 1 Behavior. In *Proceedings of the 7th International Conference on Educational Data Mining*.

Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014). Understanding in-video dropouts and interaction peaks inonline lecture videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 31-40). ACM.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170-179). ACM.

Li, V. O. K., Liao, W., Qiu, X., & Wong, E. W. M. (1996). Performance model of interactive video-on-demand systems. *Selected Areas in Communications, IEEE Journal on*, *14*(6), 1099-1109.

Sinha, T., Li, N., Jermann, P., & Dillenbourg, P. (2014). Capturing" attrition intensifying" structural traits from didactic interaction sequences of MOOC learners. *arXiv preprint arXiv:1409.5887*.

Shenoy, P. J., & Vin, H. M. (1995). Efficient support for scan operations in video servers. In *Proceedings of the third ACM international conference on Multimedia* (pp. 131-140). ACM.

Syeda-Mahmood, T., & Ponceleon, D. (2001). Learning video browsing behavior and its application in the generation of video previews. In*Proceedings of the ninth ACM international conference on Multimedia* (pp. 119-128). ACM.