When Do Learners Rewatch Videos in MOOCs?

Max Bothe
Hasso Plattner Institute
Potsdam, Germany
max.bothe@hpi.de

Christoph Meinel
Hasso Plattner Institute
Potsdam, Germany
christoph.meinel@hpi.de

***Abstract*—Mobile applications for MOOCs (Massive Open Online Courses) offer the possibility to download learning material to enable network independent learning sessions. The management of downloaded content on mobile devices is a manual process for the learner, which has the potential for automation. This includes the deletion of learning material that is likely to be no longer consumed. In this paper, a metric was defined to quantify learners' references to previous videos based on the order in which the videos were viewed. In an observational study involving three MOOCs in the field of computer science and IT systems engineering, learners referred to previous video content only a single time on average. Outliers made use of earlier content up to 44 times during a course. Referenced videos belonged in most cases to the current or previous course section. The learners referred more often to previous videos during the course period compared to when participating in self-paced mode, while learners who earned a record of achievement referred to previous videos significantly more frequently than those who did not.**

***Index Terms*—MOOCs, Mobile Seamless Learning, Mobile Learning, E-Learning**

I. Introduction

With the popularization of MOOCs (Massive Open Online Courses), a novel approach to learn on the Internet was formed. Characteristically for MOOCs, new content is published every week, which can be consumed by the learners in the preferred pace and learning environment [1]. The provided learning material often consists primarily of videos, self-tests, and additional reading material. Depending on the type of content, the learning material can be more or less data-intensive. With the increasing acceptance and usability of mobile devices and the mobile web for universal purpose tasks, learners started accessing MOOC platforms on mobile devices [2], [3]. MOOC platform providers adapted to this shift by optimizing their web applications for smaller screens [4]. But as the mobile web is designed to provide a generic user experience, which is not optimized to the learners' needs, platform providers started to offer mobile applications to make use of the deeper system capabilities of the mobile devices (e.g., notifications) [5]. One of these system capabilities is to download and manage content for later usage independent of an Internet connection. While nowadays access to the Internet is considered omnipresent in bigger cities, there are still remote locations (e.g., in rural areas but also during flights) that do not offer an adequate Internet connection [6]. Data-intensive network tasks like streaming videos become a burden — especially when relying on a costly mobile data plan [7]. To overcome this, mobile applications for MOOC platforms gained the ability to download learning material and storing it for later network-independent usage [4]. In this way, the downloaded content is embedded in the context of the mobile application — creating a more comprehensive seamless learning experience [8], [9]. This approach does not reduce usability compared to manually downloading video files and storing them outside of the learning environment [10].

Learners are required to plan their next learning session by downloading upcoming learning material manually. They select the content to be downloaded based on their needs and the available time, as well as on the amount of available downloadable content. For the learners' convenience, this manual process should be automated. Mobile applications can make use of the reoccurring availability of WiFi connections. Unlike web applications, they are capable of running in the background in periodic intervals and without user interactions. When the device is connected to a WiFi network, these background activities can be used to automatically download upcoming content. However, storage capabilities on mobile devices are limited. Therefore, downloaded learning material should be deleted as soon as it will no longer be referenced by the learner. As of now, this is also a manual process for learners. When automating learning material downloads, it will be required to delete content automatically. Here it's important to avoid deleting content that has already been consumed by the learner but is likely to be consumed again. The reason for this is that learners are not required to consume MOOC content linearly or might refer to previous learning material for verification or reassurance [11]. This paper examines the learner's references to previous learning to quantify the linearity in the learner's activities. As videos are the most used medium for information transfer in MOOCs, as well as the most data-intensive content, only the consumption of video items is considered in this work. The following research questions were defined to measure how often learners refer to previous videos and the distance between sub-sequential video consumptions, as well as to explore correlation with the course outcome independent of the learners' motives and intentions:

RQ1 How linear is the video consumption in MOOCs?
- **RQ1a** How often do learners refer to previous videos?
- **RQ1b** To which section belongs the referred content?

RQ2 Does a linearity metric differ between in course and out of course learning activities?

RQ3 Is there a correlation between linear video consumption and the course outcome?

Accepted at: 2020 IEEE Learning With MOOCS (LWMOOCS) DOI: 10.1109/LWMOOCS50143.2020.9234368

II. Defining Backward References

Video learning analytics in general sets the focus on analyzing the learners' actions on a single video [12]. Here, the learning experience is improved through insights from timestamps a video was paused, as well as at which segments the video playback was repeated. By these means, important or ambiguous content can be identified. Instead of measuring the quality of single video items, this work aims to create a better understanding of the interconnections of multiple videos. The order of the learning material in the linear MOOC structure is usually chosen carefully according to different insights from the field of learning design. But even a well-designed course structure is no guarantee for linear content consumption. Especially when learners experience difficulties, they tend to refer to earlier content [13].

In the context of smart automatic management of learning resources, MOOC platform providers need to know the order of the video consumption, as well as the total amount of references to previous videos by the course participants to provide the required functionality. To quantify the learning behavior, a metric was defined that only considers references to previous videos and neglects transitions to future video items. Watching the same video multiple times in a row does not affect the metric. Under these presumptions, the metric for backward video references has been defined as follows for a single user u in a course c:

$$dist_i(c,u) = pos(c, vp_i(c,u)) - pos(c, vp_{i+1}(c,u)) \quad (1)$$

$$adjustedDist_i(c,u) = \begin{cases} 0 & \text{if } dist_i(c,u) \leq 0. \\ 1 & \text{if } dist_i(c,u) > 0. \end{cases} \quad (2)$$

$$metric(c,u) = \sum_{i \in WV} adjustedDist_i(c,u) \quad (3)$$

Equation 1 calculates the distance of the two adjacent video play events. For this, the videos played (i and $i+1$) by a user u in course c are determined (vp). Afterward, the position of the respective videos within the course structure is determined. Equation 2 adjusts the distance of Equation 1 by applying the Heaviside step function. In this way, forward references are disregarded, while the distance of backward references is unified. Equation 3 accumulates all adjusted distances for the set of all watched videos (WV) for the user u in course c.

III. Evaluation

This section covers multiple applications of the previously defined metric. While each application field has to potential to be analyzed separately in more detail, this first study aims to create an initial overview of possible directions.

A. Methodology

The defined metric was tested in an observational study with three courses from the HPI MOOC Platform. These courses ran for six weeks in 2018, followed the traditional MOOC structure — primarily consisting of videos mixed with self-tests and reading material — and were held over a predefined course period with a start and end date. Afterward, the courses are accessible in a self-paced mode. Choosing courses from 2018 ensured a sufficient number of learners who enrolled after the course end date. In doing so, this study can analyze the learning behavior during the course period and in self-paced mode. The analyzed data sets were collected on Jan 31, 2020.

TABLE I: General Metric Analysis Among Active Learners

		Metric		
Course	N	Mean	Std.Dev.	Max
bigdata2017	4430	1.03	2.63	39
intsec2018	3446	0.65	2.11	44
semanticweb2017	1469	1.09	2.76	30

TABLE II: References to Previous Videos by Section

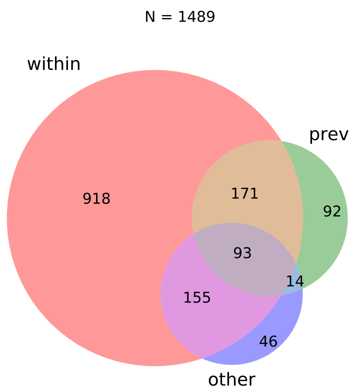
		Referred Section		
Course	N	Same	Previous	Other
bigdata2017	4560	79.9%	10.7%	9.4%
intsec2018	2242	76.9%	11.8%	11.3%
semanticweb2017	1606	77.4%	14.4%	8.2%

B. General Analysis

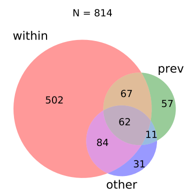
Table I displays the basic information of the defined metric in the studied courses, next to the count of active learners in the course period. On average, learners referred only a single time to the previous content (with a standard deviation of approx. 2.5). This can be an indicator of well-structured course design, given that the supervising professors already provided a good amount of knowledge in teaching students. However, in each course, some learners extensively referred to previous content (30–44 times). For those outliers, it would be practical to keep downloaded learning material on the device to ensure content availability and to avoid additional network activities.

Well-structured MOOCs divide the learning material into different sections, among other things, to keep the learners' activities focused on a specific topic. Ideally, these sections are self-contained to avoid learners cross-referencing between sections. For verification, the defined metric was modified to only consider references to a video of previous sections instead of simple previous video items. Table II shows the results of this adjusted calculation. In the course period, the majority of backward referenced video items belonged to the same section (77.4–79.9%). Further, 10.7–14.4% of the videos were part of the previous section. Based on these results, it can be inferred that learning material is most likely to be backward referenced while it belongs to the last published section. Downloaded resources can remain on the mobile devices for the learner convenience before being deleted automatically.

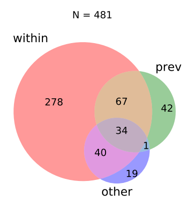
To measure the relevance of the previously gained results, the references to other sections were grouped by the respective learner groups (*Same*, *Previous*, and *Other*) formed by the various uses of backward references. By doing so, it is possible



(a) bigdata2017



(b) intsec2018



(c) semanticweb2017

Fig. 1: Learners Referring to Content of Different Sections

to identify the number of users that would have been fully supported by offline available learning material. Figure 1 contains a Venn diagram for each of the studied courses showing the resulting learning groups and their intersections. By holding on to learning material from the current section, 57.8 to 61.7% of learners would have always referred to videos still available on the mobile device. When keeping the downloaded material of the current and previous section, 76.9 to 80.5% of learners would have accessed only downloaded learning material. These results indicate that by storing the learning material of the last course section, most learners would already have benefited from smart automatic download management.

C. Enrollment Date

During the course period, learning activities are bundled to enable social learning [1]. Course participants in self-paced mode, and therefore applying self-regulated learning, reveal different learning patterns [14]. Hence, the defined metric for activities within the course period should be compared to the learning activities in self-paced mode. Learners were assigned to either group based on their enrollment date. Learners who enrolled after course end were assigned to the group *After Course End* while all other users were assigned to the group *During Course*. Both groups were filtered to only contain learners who received a confirmation of participation[1]. By these means, only learners who intended to finish the course were considered in the analysis. Especially after the course end, learners might interact more selectively with the course material [14]. Table III shows the group sizes as well as the mean and the standard deviation of the defined metric for each group and each of the examined courses. During the course period, learners more often referred to previous videos than learners who participated in the self-paced mode. A Mann-Whitney-U test showed significant differences between the two groups with medium effect sizes. Based on the available data, it was not possible to reliably assign the learners to the two groups. Learners could have enrolled during the course period but finished the course after the course end.

D. Course Outcome

Depending on the learners' objectives for a course, they show different engagement levels during the course period [14]. The behavior might also influence the number of references to previous video items. To examine this aspect, the defined metric was compared for learners who earned a record of achievement[2] to those who only received a confirmation of participation[1] during the course period. Table IV contains data of learners who received either document. In all of the studied courses, learners who received a record of achievement referred more often (1.71 to 3.43) to previous videos item than learners who got a confirmation of participation (0.94 to 1.48). The applied Mann-Whitney-U test revealed high significant differences. The effect sizes determined using Cohen's d ranged from a medium to a large scale.

IV. Conclusion and Future Work

With these findings of the linearity of video consumption in MOOCs, some insights into the applicability can be gained. On average, learners referred on a single time to previous video content (RQ1a). However, some highly active learners made use of earlier content up to 44 times in a course. When learners refer to previous videos, they most likely belong to the current course section or in extension to the previous section (RQ1b). Therefore, storing the latest learning material on the mobile devices would decrease networking activities for the majority of learners, while at the same time makes another step towards network independent learning activities. Learners who achieved a confirmation of participation in the course period referred more often to previous videos than learners who earned the same confirmation in self-paced mode (RQ2). It can be concluded that during the course downloaded learning material should be stored more liberally on the device while it can be extensively deleted after the course end given learners have agreed to such an automatic behavior of the mobile application. Learners who earned a record of achievement in the course period were significantly more likely to refer to previous videos than those who only received a confirmation of participation (RQ3).

While automatic management for downloaded resources on mobile devices might not be relevant for all course participants at the same level, it is worth noting the practical effect on a variety of smaller target groups. Besides learners with limited access to the Internet, this includes learners who actively browse through the course material, as well as learners who want to earn a record of achievement. All these groups would receive an improved learning experience and less planning effort through an automatic download management system.

[1] A confirmation of participation is granted to all learners who visit at least 50% of the provided learning material.

[2] A record of achievement was granted to all learners who earned at least 50% of the available points in homework and the final exam.

TABLE III: Backward Video References During the Course Period and After Course End

	During Course			After Course End			Mann–Whitney U		
Course	N	Mean	Std.Dev.	N	Mean	Std.Dev.	U	p-value	Cohen's d
bigdata2017	1705	2.23	3.84	284	1.45	2.34	222333.5	0.011	0.278
intsec2018	1331	1.45	3.14	64	0.69	1.25	37264.0	0.031	0.340
semanticweb2017	498	2.60	4.16	51	1.63	2.99	10562.5	0.020	0.324

TABLE IV: Backward Video References when Earning a Record of Achievement (RoA) or Confirmation of Participation (CoP)

	RoA achieved			Only Cop achieved			Mann–Whitney U		
Course	N	Mean	Std.Dev.	N	Mean	Std.Dev.	U	p-value	Cohen's d
bigdata2017	829	3.15	4.71	876	1.35	2.48	254020.0	<0.001	0.475
intsec2018	872	1.71	3.54	459	0.94	2.08	172321.0	<0.001	0.284
semanticweb2017	287	3.43	4.76	211	1.48	2.79	20577.0	<0.001	0.516

In this study, only MOOCs with a six-week course period were examined. Courses spanning over a shorter period might differ in the learning behavior from the described findings. Courses that are not following the traditional MOOC structure can also be subject to different results. As the presented study was of a smaller scale, the gathered insights should be reviewed in future studies with larger data sets and with less technical courses. Further, the usefulness and learners' acceptance of smart automatic management for downloaded resources has to be evaluated as a whole, as automatic deletions of downloaded learning material are only desirable for the learner in addition to automatic downloads. This study provides first insights into such a system while outlining appropriate default settings of the download management. Nevertheless, learners should always be able to override such default settings to fine-tune the learning experience to their needs and preferences. Suitable options could be: Deleting downloaded videos immediately after playback, removing all used learning material from the mobile device at the end of the day, or generally disabling automatic deletion.

Although initially formulated for the use case of smart automatic download management of learning resources, the defined metric can be considered as a general measured of linearity in the learning process. By including all available learning resources, a universal metric could be created. However, such a metric would consequently consider multiple other incentives for backward references within a course, like retaking previous self-tests as preparation for graded assignments. Hence a universal metric has not been considered in this work.

References

[1] L. Guàrdia, M. Maina, and A. Sangrà, "Mooc design principles: A pedagogical approach from the learner's perspective," *elearning papers*, no. 33, 2013.

[2] I. deWaard, A. Koutropoulos, N. Keskin, S. C. Abajian, R. Hogue, C. O. Rodriguez, and M. S. Gallagher, "Exploring the mooc format as a pedagogical approach for mlearning," in *Proceedings of 10th World Conference on Mobile and Contextual Learning*, 2011, pp. 138–145.

[3] M. Sharples, C. Delgado Kloos, Y. Dimitriadis, S. Garlatti, and M. Specht, "Mobile and accessible learning for moocs," *Journal of interactive media in education*, vol. 1, no. 4, pp. 1–8, 2015.

[4] J. Renz, T. Staubitz, and C. Meinel, "Mooc to go.," *International Association for Development of the Information Society*, 2014.

[5] A. Charland and B. LeRoux, "Mobile application development: Web vs. native," *Queue*, vol. 9, no. 4, pp. 20–28, Apr. 2011, ISSN: 1542-7730. DOI: 10.1145/1966989.1968203. [Online]. Available: https://doi.org/10.1145/1966989.1968203.

[6] S. Strover, "Rural internet connectivity," *Telecommunications policy*, vol. 25, no. 5, pp. 331–347, 2001.

[7] I. T. Union, *ICT Prices 2017*, 2017. [Online]. Available: https://www.itu.int/itu-d/apis/clients/publications/res/bdt/BDT-REPORT-ICTPRICES-2017-PDF-E.pdf.

[8] L.-H. Wong, "A brief history of mobile seamless learning," in *Seamless learning in the age of mobile connectivity*, Springer, 2015, pp. 3–40.

[9] M. Bothe and C. Meinel, "Applied mobile-assisted seamless learning techniques in moocs," in *European MOOCs Stakeholders Summit*, Springer, 2019, pp. 21–30.

[10] F. Martin and D. U. Bolliger, "Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment.," *Online Learning*, vol. 22, no. 1, pp. 205–222, 2018.

[11] D. Davis, G. Chen, C. Hauff, and G.-J. Houben, "Gauging mooc learners' adherence to the designed learning path.," *International Educational Data Mining Society*, 2016.

[12] N. Li, Ł. Kidziński, P. Jermann, and P. Dillenbourg, "Mooc video interaction patterns: What do they tell us?" In *Design for teaching and learning in a networked world*, Springer, 2015, pp. 197–210.

[13] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, "Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses," *Computers & education*, vol. 104, pp. 18–33, 2017.

[14] T. Rohloff, D. Sauer, and C. Meinel, "On the acceptance and usefulness of personalized learning objectives in moocs," in *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, ser. L@S '19, Chicago, IL, USA: Association for Computing Machinery, 2019, ISBN: 9781450368049. DOI: 10.1145/3330430.3333624. [Online]. Available: https://doi.org/10.1145/3330430.3333624.