Learners' Interactions in Massive Open Online Courses: Analysis and Interpretation

Anna A. Shurtina
Department of Open Education
ITMO University
Saint Petersburg, Russia
a.shurtina@cde.ifmo.ru

Andrey V. Lyamin
Department of Open Education
ITMO University
Saint Petersburg, Russia
lyamin@mail.ifmo.ru

Elena N. Cherepovskaya
Department of Open Education
ITMO University
Saint Petersburg, Russia
cherepovskaya@cde.ifmo.ru

Abstract—**The research is devoted to the study of data on learners' interactions in massive open online courses. Based on the logs of online-learning platforms, the following research was made: a comparison of the behaviour of motivated and unmotivated learners regarding of video lectures, identification of the most valuable for the successful completion of the course activities of learners, creating a model of going through time-limited assignments and identification of cheating approach based on this model.**

The following conclusions were made: motivated and unmotivated learners watch video lectures in different ways, motivated learners appeared to be 14 times more active, the most interesting and most viewable videos were revealed. When identifying the most valuable theoretical materials influencing the successful completion of the course, the following results were obtained: some of the videos have a strong influence on the successful completion of the final assignment. Some of the videos appeared to have weak effect, they can be interpreted as non-obligatory. Ungraded tests have a positive but moderate effect on learners' success, while communication via discussion forum has no effect at all. In addition, a model of going through time-limited assignments was built using the average passing time of reliable learners, the approach for identifying cheating with examples is presented in the study.

Keywords— learning analytics, MOOC.

I. INTRODUCTION

The rapid development of information technology has an impact on all spheres of human life, education is no exception. Already for a long time, traditional classroom lessons have been accompanied by the use of electronic educational resources and distance learning technologies and ultimately have gone online.

The technology of massive open online course allows hundreds of thousands of learners to study for free from the best teachers at leading universities in the world, regardless of time or location, provided they have a device with Internet access [1]. Modern online-learning platforms have powerful tools for presenting educational content [2], allowing students not only to gain theoretical knowledge, but also to immediately work it out in practice, providing a full range of knowledge and skills in the chosen field.

According to the Class Central portal (https://www.class-central.com/), which regularly provides MOOC statistics, as of January 2019, there are more than 900 universities-providers of MOOCs all over the world, and the number of MOOCs exceeds 11 thousand. The number of learners studying online is more than 101 million people.

Despite the high popularity of MOOCs and their lightning-fast development, students and developers of online courses encounter a number of issues. First, due to the high rate of development of the MOOCs, there are gaps in terms of the pedagogical design of the courses and the methodological and technical quality of the materials presented. Secondly, students are faced with a lack of self-organization and motivation to complete the learning: the world statistics says that on average less than 13% of enrolled learners successfully complete the MOOC [2, 3].

Learning analytics - a set of approaches for collecting, storing and processing online learning data - appears to the rescue of these problems [4]. Millions of learner watch videos, solve problems, do practical exercises, communicate via discussion forums of MOOCs and so on. Data on these activities is stored in the so-called logs of online-learning platforms — all events caused by learners' clicks on the various components of online courses are recorded. Such data fits into the concept of the so-called "Big data" - huge amounts of information, the study of which is aimed at finding latent patterns and regularities [5]. By now, none of the online-learning platform provides detailed statistics based on data from logs, so a lot of research has been done in order to prove that learning analytics could be helpful and needs to be embedded and used [6]. Different approaches of learning analytics are aimed to support students' motivation to complete learning [7, 8], personalize and individualize learning [9], monitor and improve the quality of the content provided and create feedback needed for both learners and course developers [10]. Several of approaches applied are presented in the following study.

II. MATERIALS AND METHODS

Data from online-learning platform logs of tree online courses on technical topics was chosen for this study. These courses are selected as they were mentioned in the list of recommended online courses of All-Russian Olympiads, which characterizes them as recognized by the academic community and qualitative in terms of teaching materials and their technical presentation. Logs contain data on all the events caused by the learners' clicks on MOOC components with a mention of a timestamp, a link to the course component and other meta-data of each event.

A. Unmotivated and motivated video-interaction patterns

For the video-interaction research the course with 1354 learners was selected, who all together initialized more than 54,000 events. It is known that some learners are enrolled in a course with the condition that successful completion guarantees them a positive assessment in the relevant discipline at the university. The logins of these learners on the platform are known; in this study, the data of learners' interaction is separated into two groups called motivated and unmotivated.

The tendency of launches and completed views of video lectures by motivated and unmotivated learners is being observed. 48 video lectures with theoretical materials of the course with different duration from 88 to 1333 seconds were taken into account. All of the video lectures contain a screencast of presentation slides and a voiceover. Some of the lectures contain scenes with the course author speaking. Downloading and launching a video lecture from its beginning on the platform triggers a "load" event with a unique identifier of the learner and the video lecture. A completed view of a video lecture, when the slider on the roller duration scale reaches the extreme position, triggers the "stop" event with the same parameters. Those events, when the video lecture was stopped at the initiative of the learner within 5 seconds before the end of the video, are also taken to account. This is necessary because some videos finish with a non-informative screensaver, or the informative part of the lecture just stops a few seconds before the end, therefore, the learner watched the content part completely, but was not patient enough to initiate the "stop" event. When a learner clicks on a pause, the platform triggers the "pause" event, which contains information about the learner's unique identifier, video lecture, and pause timestamp. In this case, we will take into account those "pause" events that have a timestamp different from the total duration of the video by no more than 5 seconds.

All the events described above are to be processed for both motivated and unmotivated students in order to compare their behavior patterns and determine the most popular and the most unwatched videos in the course.

B. Identification of the most important for the successful completion of the course activities of learners

As each MOOC is aimed to provide learners with useful resources to gain knowledge and skills, descriptive analytics such as counting times when learners referred to course components can be not enough. The more informative data describes if the provided material really does help learners to study or maybe it does not influence their performance at all. Using the logs of the online-learning platform on which the course was placed, data about the interaction of learners with various types of components of the course was obtained, such as watching video lectures, taking ungraded quizzes, mostly multiple-choice and checkbox problems with a detailed solution provided after passing, interacting on forums and, finally, performing weekly test.

To determine the most important components for successful completion of the course, data of activities within the first week of the course was used. 63 students initiated more than 2100 log events that were taken into account. All of the above activities were calculated until spending the maximum number of attempts when solving the graded weekly test. Before counting, each activity was measured by three states: when a learner did not initiate any events connected with the course component, when a learner started to interact but didn't complete (e.g. loaded the video but did not watch it till the end, took quiz but did not succeed, etc.), and when a learner fully completed the activity.

In order to assess the importance of interaction with different components, the Pearson correlation coefficients between the measured activities of learners and their performance in the weekly test are to be found. The maximum degree of correlation is taken as 1. There are the following types of correlation: strong ($\pm 0.7 \ldots \pm 1.0$), moderate ($\pm 0.5 \ldots \pm 0.7$) and weak ($\pm 0.3 \ldots \pm 0.5$). If coefficient is $\leq \pm 0.3$, the connection is practically absent. The minus or plus sign of the correlation coefficient indicates the direction of the connection — the plus sign means that the connection between the activity and performance is direct (positive), the minus sign is the reverse connection (negative). These coefficients are to show how interacting with different types of components influence learners' performance.

C. Creating a model of going through time-limited assignments

Online-learning platforms today have various automated tools for assessing learners' knowledge gained. One of the common tools is the usage of time-limited sets of problems, randomly distinguished from libraries. Using data about the time spent on solving problems, a model of going through such assignments can be built. This method of research is possible in the case when the difficulty of the task correlates with its average solving time.

As part of the study, the events of the solution of 329 tasks by 578 learners was processed. At the beginning of the study, it is necessary to determine all the events in which the leaners solve specific problems, taking into account the time they spent on it. In this study, the course on a technical topic is being observed, it means that the solution of a problem implies the execution of several calculated actions, i.e. it is impossible to solve

the problem in one step just by reading its condition. In addition, information on the final grade of the learners on the course will be required in order to determine the group of students, which results can be considered as reliable: learners who received good and excellent final marks during the course and completed all the activities of the course and study all the theoretical materials.

Accordingly, for each set of problems with the above data a solving model could be build using the upper and lower expected time limits counted with the 95% confidence interval of reliable learners' solving time. If the course instructor has to validate some learners' results of solving a set of problems, the actual time of solving of this learner can be compared with the constructed model to check whether the actual time correlates with the expected time and falls into a certain confidence interval or not.

III.   RESULTS AND DISCUSSION

A.   *Unmotivated and motivated video-interaction patterns*

The number of motivated students of the chosen course is 109 people, i.e. 8% of the total. The number of events caused by the actions of learners in relation to video content was counted. In total, 54,303 events were recorded, of which 24,348 (or 45%) were caused by the actions of unmotivated learners, 29955 (55%) were caused by the actions of motivated learners. Consequently, one motivated learner interacts with video content approximately 14 times more active than the unmotivated one.

Figure 1 shows the graphs of the distribution of launches (black) and completed views (grey) of video lectures by unmotivated learners over the weeks of the online course.

Unmotivated learners show a high interest to watching videos only in the first week of the online course, by the second week the interest is reduced three times, on the subsequent weeks the number of launches is on average 20 times less than in the first week of the online course. This confirms the hypothesis that the overwhelming majority of learners drop out at the very beginning of the course due to lack of motivation.
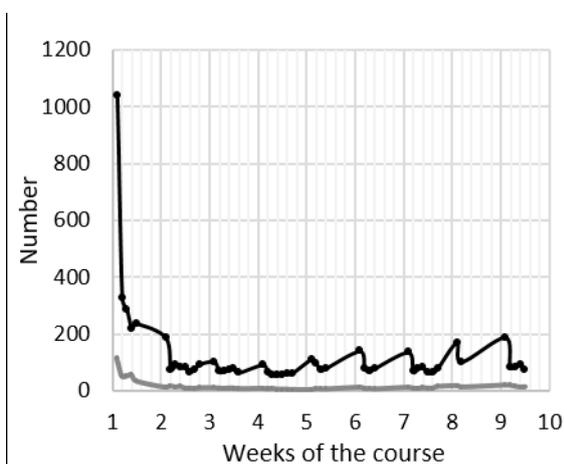
Fig. 1.   Launches (black) and completed views (grey) of videos by unmotivated learners throug the course

In addition, it was revealed that at weeks 1–3 and 7–10 the percentage ratio of completed video lectures to launches is on average 11%. This means that 11% of started views were completed. At 4, 5 and 6 weeks of the online course this value is 6%, therefore, it can be assumed that at these weeks of the online course the video content was too complicated, too simple or of poor quality from a technical or methodological point of view, it is necessary to work to identify the causes of low interest of learners.

Figure 2 shows the graphs of the distribution of launches (black) and completed views (grey) of video lectures by motivated learners over the weeks of the online course.
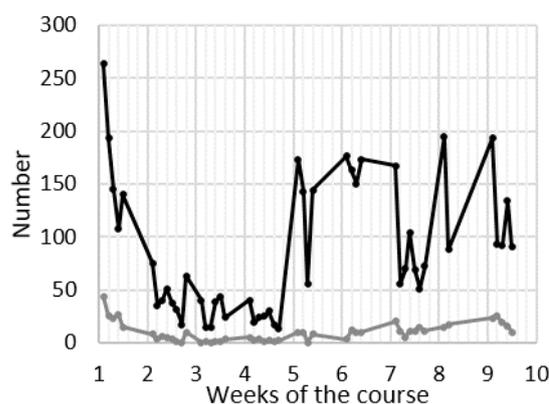
Fig. 2.   Launches (black) and completed views (grey) of videos by motivated learners throug the course

Unlike the group of unmotivated learners, there is a different trend. The number of launches and completed views after the first week is reduced by half. In the period from the second to the fourth week of the online course, there is also a decrease in the activity of learners to launch video lectures, however, interest grows again in the fifth week and remains at a high level until the end of the course, except for materials of the seventh week. However, in relative terms, the seventh week of the course is one of the most popular in terms of completed views: 16% of the started views were completed, when for a fifth week, which is the most successful one in terms of the number of video loadings, this value is only 5%.

The main result of the observation of these data is that motivated and unmotivated learners do behave differently in interacting with content of a MOOC, so course instructors need to determine a reliable group of learners whose interaction data can be considered as informative feedback.

B.   *Identification of the most important for the successful completion of the course activities of learners*

Results of calculating Pearson correlation coefficients between measured learners' activities and their performance on the first week of a course are represented in the Table I.

TABLE I. Correlation Coefficients Between Learners Activities and Peformance

	Weekly Test Performance
Watching Video 1	0,32
Watching Video 2	0,76
Taking Quiz 1	0,61
Watching Video 3	0,71
Taking Quiz 2	0,63
Watching Video 4	0,48
Taking Quiz 3	0,61
Watching Video 5	0,49
Taking Quiz 4	0,52
Discussing on Forum	0,02

The results can be interpreted as follows: important for successful completion of the first weekly test are watching video 2 and watching video 3. All the quizzes of the first week of the course have a moderate impact on the success of the final test. The first, fourth and fifth videos have a weak effect. Communication on the course forum does not correlate with the success in the weekly test at all.

As recommendations for the course developers, the following can be singled out: video 1 is to be placed in a separate section, since it seems to be introductory and has a weak effect on the successful completion of the first weekly test, the difficulty of quizzes in the first week of the course is to be increased by changing the problem type, e.g. from close to open response form, to make them more challenging.

As recommendations for learners, the following can be singled out: videos 2 and 3 must be viewed till the end and all surveys are recommended to be resolved, since these activities will most likely help them successfully complete the weekly test.

C. Creating a model of going through time-limited assignments

The correlation coefficient between solving time and problem difficulty for a group of students who successfully completed the course appeared to be 0.7034, which indicates a strong connection and the possibility of building a model for problem sets based on this data. For each problem, a 95% confidence interval of the solving time by reliable group of students was calculated. After that, every learners' solution can be placed on the graph with a created model.

For example, Figure 3 shows the model of passing the set of 13 problems, the green dashed line indicates the maximum expected solution time, the blue dotted line indicates the minimum expected solution time, the red solid line indicates the actual time of problem solving by the learner.
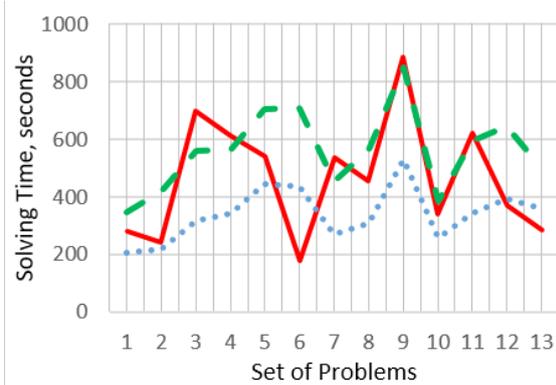
Fig. 3. Maximum expected time (green), minimum expected time (blue) and actual solving time (red) of a specific learner

The graph on Figure 3 shows that the solution of some problems falls within the confidence interval, the solution of some problems takes more or less time than expected. However, the graph clearly shows the dynamics of the time spent on the decision, which may indicate that the learner solved each task independently.

Figure 4 shows a graph of the solving time of another set of problems by another learner.
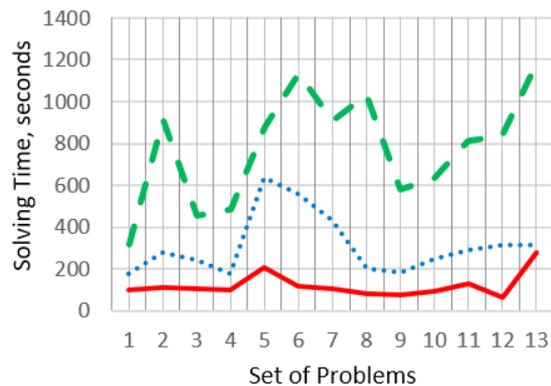
Fig. 4. Maximum expected time (green), minimum expected time (blue) and actual solving time (red) of another specific learner

The result from the Figure 4 can be interpreted unequivocally - the learner spends about the same amount of time on each problem, none of the problems' solving times do not fall into the expected interval. In case if all the problems were solved successfully, it can be concluded that the student was cheating.

Two of the most illustrative instances of applying the approach are shown in this study. Such identification can be required when a high score of solving a set of problems can be put by the course instructor into doubt and verified by comparing the actual solving time with the expected pattern.

IV. CONCLUSIONS

The implementation of the learning analytics methods presented in the study allows determining the activity and involvement of learners, finding a correlation between the activity of learners and their performance, forming recommendations for course developers and learners, as well as building models for passing problem sets and identifying cheating. Using the considered methods in e-learning systems can help authors of online courses assess the quality indicators of published content, and learners to maintain motivation and successfully complete learning on the course.

REFERENCES

[1] Baturay, M. (2015) An Overview of the World of MOOCs. Procedia - Social and Behavioral Sciences, 174, Ankara: Elsevier, pp. 427-433

[2] Davis, D., Jivet, I., Kizilcec, R.F., Chen G., Hauff, C., Houben, G.J.: Follow the successful crowd: raising MOOC completion rates through social comparison at scale. In: Proceedings of LAK 2017, pp. 454–463. ACM (2017)

[3] Gomez-Zermeno, M. G., & Garza, L. A. (2016). Research analysis on MOOC course dropout and retention rates. Turkish Online Journal of Distance Education, 17(2), 3–14.

[4] Song, Donggil. (2018). Learning Analytics as an Educational Research Approach. International Journal of Multiple Research Approaches. 10. 102-111. 10.29034/ijmra.v10n1a6.

[5] D. Gibson, "Big Data in Higher Education: Research Methods and Analytics Supporting the Learning Journey," Technol. Knowl. Learn., vol. 22, no. 3, pp. 237–241, Oct. 2017.

[6] Banihashem, Seyyed Kazem & Aliabadi, Khadijeh & Ardakani, Saeid & Delavar, Ali & Nili Ahmadabadi, Mohammadreza. (2018). Learning Analytics: A Critical Literature Review. 9. 10.5812/ijvlms.63024.

[7] Alario-Hoyos, Carlos & Estévez-Ayres, Iria & Pérez-Sanagustín, Mar & Delgado-Kloos, Carlos & Fernández Panadero, Carmen. (2017). Understanding Learners' Motivation and Learning Strategies in MOOCs. The International Review of Research in Open and Distributed Learning. 18. 10.19173/irrodl.v18i3.2996.

[8] Onah, Daniel & Sinclair, Jane & Boyatt, R. (2014). Dropout Rates of Massive Open Online Courses: Behavioural Patterns. 10.13140/RG.2.1.2402.0009.

[9] Maseleno, Andino & Sabani, Noraisikin & Huda, Miftachul & Ahmad, Roslee & Jasmi, Kamarul Azmi & Basiron, Bushrah. (2018). Demystifying Learning Analytics in Personalised Learning. International Journal of Engineering and Technology. 7. 10.14419/ijet.v7i3.9789.

[10] Michos, K., Hernández-Leo, D., Jiménez, M., (2017) How educators value data analytics about their MOOCs, CEUR Proceedings of Work in Progress Papers of the Experience and Research Tracks and Position Papers of the Policy Track at EMOOCs 2017 co-located with the EMOOCs 2017 Conference (Vol-1841), Madrid, Spain, 77-82