An Approach to Build *in situ* Models for the Prediction of the Decrease of Academic Engagement Indicators in Massive Open Online Courses

Miguel L. Bote-Lorenzo
(Universidad de Valladolid, Spain
migbot@tel.uva.es)

Eduardo Gómez-Sánchez
(Universidad de Valladolid, Spain
edugom@tel.uva.es)

Abstract: The early detection of learners who are expected to disengage with typical MOOC tasks such as watching lecture videos or submitting assignments is necessary to enable timely interventions aimed at preventing it. This can be done by predicting the decrease of academic engagement indicators that can be derived for different MOOC tasks and computed for each learner. *A posteriori* prediction models can yield a good performance but cannot be built using the information that is available in an ongoing course at the moment the predictions are required. This paper proposes an approach to build *in situ* prediction models using such information. Models were derived following both approaches and employed to predict the decrease of three indicators that quantify the engagement of learners with the main tasks typically proposed in a MOOC: watching lectures, solving finger exercises, and submitting assignments. The results show that *in situ* models yielded a good performance for the prediction of all engagement indicators, thus showing the feasibility of the proposed approach. This performance was very similar to that of *a posteriori* models, which have the clear disadvantage that they cannot be used to make predictions in an ongoing course based on its data.
Key Words: MOOC, engagement, supervised machine learning
Category: L.2.0, L.3.5

1 Introduction

Student engagement has been defined as "the student's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or crafts that academic work is intended to promote" [Newmann, 1992]. It is considered to be highly related with academic achievement [Finn and Zimmer, 2012], which is the reason why this concept has attracted a great interest from the scientific community as a possible antidote to academic failure [Fredricks et al., 2004].

There are four components that can be identified in recent models of student engagement [Finn and Zimmer, 2012]. *Academic engagement* is related to the observable behaviors directly related to the learning process that are exhibited by students participating in course work, for instance, attentiveness and assignment completion. *Social engagement* determines to which extent a student adheres to written and unwritten rules of behavior such as, for example, interacting appropriately with teachers and peers. *Cognitive engagement* refers to the intellectual effort made to comprehend complex ideas. *Affective engagement* is associated with the feelings of involvement in the learning community. Interestingly, these components tend to be highly intercorrelated [Finn and Zimmer, 2012]. In this way, students are usually engaged or disengaged on multiple dimensions.

Learner engagement within the context of Massive Open Online Courses (MOOCs) is currently being actively researched with a special focus on academic engagement [Sinclair and Kalvala, 2016], possibly because it is easier to quantify it than other types of engagement based on the learner activity traces that are typically stored in MOOC platforms. Various indicators that describe different observable behaviors of MOOC learners have been proposed with the aim of measuring their academic engagement. Examples of such indicators include the number of lecture videos watched, the number of submitted assignments or the number of posts created by each learner. The evolution of academic engagement indicators along the course has been analyzed revealing the existence of different patterns of engagement among MOOC learners [Ferguson and Clow, 2015, Ferguson et al., 2015, Anderson et al., 2014, Kizilcec et al., 2013, Ramesh et al., 2013, Milligan et al., 2013]. For instance, the learners that follow the so called "completing" or "all-rounders" pattern feature high values in assignment engagement indicators along the course, while those belonging to the "auditing" or "viewers" group show low values in the indicators for assignment engagement while reaching higher ones for video engagement indicators.

Interestingly, it has been consistently reported in the literature [Alario-Hoyos et al., 2014, Ferguson and Clow, 2015, Ferguson et al., 2015, Kizilcec et al., 2013, Ramesh et al., 2013, Ramesh et al., 2014] that the values of academic engagement indicators of many MOOC learners decay over time and that, quite often, this leads to dropping out. The decrease of engagement limits the educational impact of the MOOCs in learners [Ramesh et al., 2013], even if they do not eventually drop out. This is the reason why it is important to maintain and cultivate learner engagement in MOOCs [Ramesh et al., 2013].

In an attempt to tackle this problem, some works such as [Halawa et al., 2014, Vitiello et al., 2017, Xing and Du, 2018] propose predicting whether a learner will eventually drop out a MOOC in order to enable an early intervention aimed at avoiding it. In this way, drop out prediction cannot be employed to detect those learners whose academic engagement decreases but do not stop participating in the course, which precludes the possibility of making adequate interventions in their case. Moreover, in most proposals the predictions models cannot be built until learners have already dropped out [Gardner and Brooks, 2018]. These *a posteriori* prediction models are valuable to carry out a *post hoc* analysis of the relationships between academic engagement indicators and dropout. However, they cannot be used to make predictions that provide actionable information for the very same run of the course from which they originate. In addition, it was shown in [Boyer and Veeramachaneni, 2015, Whitehill et al., 2017] that these models have a limited performance when employed to make predictions in a subsequent run of the same course.

This paper proposes an approach to build models that predict the decrease of academic engagement indicators of MOOC learners using data from an ongoing course that are already available at the moment in which the predictions must be made. These *in situ* prediction models can be employed to identify learners whose academic engagement indicators in a MOOC are expected to decrease in the near future, thus providing actionable information that allows making suitable interventions aimed at maintaining the engagement of learners, regardless of whether they are expected to drop out or not. For example, the suggestion of an interesting lecture video, or a hint to solve an assignment could be provided to a learner if a decrease in her video or assignment engagement indicator has been predicted, respectively. The paper also reports the experiments that were carried out in order to compare the performance of *in situ* models with *a posteriori* models and a simple baseline for the prediction of the decrease of three different engagement indicators within the context of a MOOC delivered in the edX platform.

The rest of this paper is organized as follows. Section 2 discusses the related work that can be found in the literature. Next, section 3 provides a more detailed description of the problem of predicting the decrease of engagement indicators, explains how *a posteriori* predictions models can be built and proposes an approach to build *in situ* prediction models. Section 4 introduces the datasets that were employed in the experiments that were carried out to verify the feasibility of the proposed approach. The results of these experiments are presented and discussed in Sections 5 and 6, respectively. Finally, section 7 includes the main conclusions of the paper and presents future work.

2 Related Work

Since MOOCs gained popularity among learners and academic institutions, prediction models have been explored in order to gain insight that help alleviate some of MOOCs most relevant limitations. Among them, probably the most often cited concern is the low completion rates that are typically observed in these type of courses, being below 13% in most cases [Jordan, 2015]. In this context, there are works that have addressed the certification prediction problem. These works aim at identifying learners that will not complete the course successfully, thus not obtaining a certificate of accomplishment, so that remedial interventions can be made eventually. For example, [Joksimović et al., 2016] revealed that Social Network Analysis can be very useful for certification prediction. [Xu and Yang, 2016] proposed a method based on learner motivation to make the predictions. N-gram features in clickstream data were employed in [Li et al., 2017] to predict achievement.

Most of the learners that do not complete the course stop participating before it ends [Jordan, 2015]. This is the reason why many works in the literature try to identify which learners will stop participating in a course before it ends so that, once again, adequate interventions can be eventually triggered. For instance, [Halawa et al., 2014] introduced a simple prediction approach based on the comparison of features that describe the student behavior with thresholds. A unified model that allows for the early prediction of dropout users across different systems was presented in [Vitiello et al., 2017]. [Xing and Du, 2018] proposed the use of deep learning algorithms to build dropout prediction models. It is noteworthy that most works dealing with dropout prediction use *a posteriori* models that cannot be implemented in an active course since they require information that is not known until the course completes [Gardner and Brooks, 2018]. Interestingly, [Boyer and Veeramachaneni, 2015, Whitehill et al., 2017] showed that *in situ* models outperformed the use of transfer learning techniques in the dropout prediction problem.

The ultimate goals of both the certification and dropout prediction problem are thus different to that of the problem of predicting the decrease of engagement indicators, which is maintaining the engagement of learners with the different activities that can be carried out in MOOCs. Even if dropout prediction can be considered to implicitly predict the disengagement of drop out learners, it should be noticed that it cannot predict it in the case of learners that continue participating in the course. Furthermore, it cannot be employed to detect the decrease of engagement with specific MOOC activities (e.g. video watching, assignment submissions), which hinders the possibility of making interventions targeted to the activity in which the disengagement is identified. In this way, predicting the decrease of engagement indicators can help not only to have more learners that pass the course or that do not drop out, but also to avoid the decrease of the educational impact that the course might have on learners even if they are not expected to abandon the course or to fail to obtain a certificate.

Disengagement prediction is another related problem that has been tackled in non-MOOC contexts. [Beck, 2005] used item response theory to predict learner disengagement in a computer tutor based on the time taken to answer questions. In [Cocea and Weibelzahl, 2011], different supervised learning algorithms were evaluated for the detection of disengaged learners in a Learning Management System. [Mills et al., 2014] also studied the prediction of disengagement in a tutoring system based on reading patterns. In these cases, the proposed solutions aim at predicting the engagement states in which the learners will be in the near feature. However, this paper proposes an approach to predict the decrease of the individual indicators that can be used to define such engagement states.

Finally, it is worth mentioning the existence of many works that research different aspects of academic engagement within the context of MOOCs but do not deal with prediction. [Goldberg et al., 2015, Zheng et al., 2015] studied how different factors affect academic engagement. Works such as [Kizilcec et al., 2013, Ferguson and Clow, 2015] focus on understanding he different ways in which participants engage with a MOOC. [Chang and Wei, 2016, Rizzardini and Amado-Salvatierra, 2017] propose different approaches to improve the engagement of MOOC learners.

3 Predicting the Decrease of Engagement Indicators

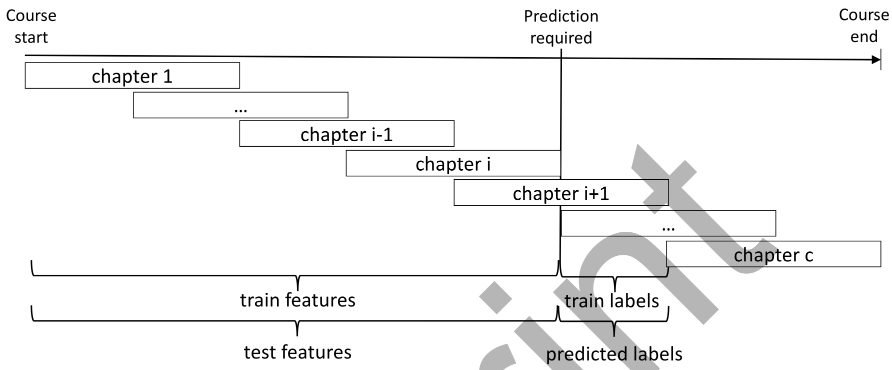
Instructor-paced MOOCs are typically structured as a sequence of chapters. In many cases the learning materials corresponding to each chapter are released along the course on a regular basis. The date in which such release takes place can be considered the chapter start. Chapters usually include assignments that must be submitted before a deadline that determines the chapter end. Here, it should be taken into account that a chapter might start before the previous one has ended.

Learners are normally expected to perform different types of tasks such as watching video or submitting assignments using the learning materials of every MOOC chapter. It is possible to define an indicator that quantifies the engagement of a learner with a given type of task from the beginning of the course until the moment in which the indicator is computed. For example, an assignment engagement indicator could be obtained by averaging the percentage of assignments submitted in each chapter that has already ended. Clearly, the value of an engagement indicator defined in this way can increase or decrease at the end of every new chapter, depending on work performed by the learner in the tasks of the previous one.

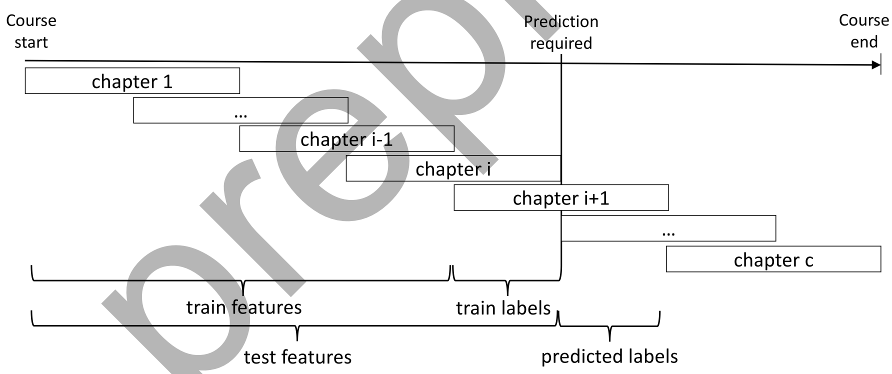
Considering a course with c chapters, the prediction of the decrease of an academic engagement indicator tries to determine whether its value for a given learner at the end of chapter $i+1$ will be lower or not than its value for the same learner at the end of chapter i based on the learner's activity in the MOOC until the end of chapter i. The following subsection describes how *a posteriori* models can be built to make such prediction. Next, an approach to build *in situ* prediction models is proposed.

3.1 *A Posteriori* Prediction Model

An *a posteriori* prediction model can be built using a training set consisting of input vectors with features that describe the activity of a set of learners until the end of chapter i and output labels stating whether the corresponding values of the engagement indicator decreased at the end of chapter $i+1$ with respect to the end of chapter i or not. Vectors with features describing the activity of a different set of learners until the end of chapter i can then be employed as inputs for the model in order to predict if the engagement indicator decreases at the end of chapter $i+1$ with respect to the end of chapter i. Figure 1(a) depicts the time span of input vectors and labels required to train and test an *a posteriori* prediction model.



(a) *A posteriori* prediction model.



(b) *In situ* prediction model.

Figure 1: Time span of features and labels employed to train and test prediction models.

The output labels that are required to train the prediction model in this way can only be computed using information that is not available until the end of chapter $i+1$. This implies that the *a posteriori* prediction model can only be built after chapter $i+1$ has been completed. Therefore, while this model can be employed to understand how the decrease of engagement can be explained from past activity, it cannot be used to predict what will happen in the future during the actual run of a MOOC, and thus it does not provide actionable information that may help to overcome the situations that lead to a lower engagement. The models could also be used in a future run of the same MOOC by means of transfer learning techniques such as the ones described in [Boyer and Veeramachaneni, 2015]. However, it must be noticed that even small changes on the structure, contents or activities of the MOOC with respect to the previous run may affect their performance significantly (e.g. if a model partially relies on an activity that is removed in the new run of the MOOC).

The performance of *a posteriori* prediction models is usually assessed using k-fold cross validation. Such performance is expected to be an optimistic estimate of what can be achieved with *in situ* prediction models, as it is in the case of dropout prediction [Boyer and Veeramachaneni, 2015].

3.2 *In Situ* Prediction Model

An *in situ* prediction model can be trained employing input vectors with features that describe the activity of all learners until the end of chapter $i-1$ along with the corresponding output labels that state if the value of the engagement indicator decreased at the end of chapter i with respect to the end of chapter $i-1$. Vectors with features that describe the activity of learners until the end of chapter i can then be provided as inputs for the model in order to obtain predictions about the decrease of the indicator at the end of chapter $i+1$ with respect to the end of i. In other words, the model is trained to make predictions for the end of chapter i but it is actually employed to generate predictions for the end of chapter $i+1$. The time span of input vectors and labels required to train and test an *in situ* prediction model is shown in Figure 1(b).

It can be noticed that this approach assumes that the same set of features that describe the learners' activity can be computed for both chapters $i-1$ and i. Furthermore, it requires each feature to be in the same range in both chapters $i-1$ and i so that features are comparable across chapters. For example, the variable percentage of submitted assignments can be compared in different chapters, but not the number of submitted assignments since the total number of requested assignments may vary from one chapter to another.

The computation of output labels that are employed to train the prediction model thus requires information that is already available at the end of chapter i provided that $i \geq 2$. In this way, it is possible to build an *in situ* prediction model at the end of a given chapter during the actual run of a MOOC to make predictions for the end of the next chapter. This allows making interventions in real time aimed at preventing the disengagement in the type of task used to define the indicator.

4 Datasets

4.1 Course Description

The datasets employed in the experiments reported in this paper were obtained from the MOOC "6.002x Circuits and Electronics" that was offered on edX in the spring of 2013 [Seaton et al., 2014]. The course was structured in 14 chapters and included a midterm and a final exam.

The main contents of every chapter are explained in two sequences of lecture videos interspersed with short and simple comprehension questions called finger exercises. Chapters 1 to 12 also comprised two types of assignments: homework problems that included numerical and formula responses, and lab exercises based on an interactive circuit simulator. In addition, most chapters provided optional tutorial videos that helped reinforcing concepts by showing how to solve circuit problems or illustrating interesting principles. The chapter learning materials were supplemented by on-line sections of the course textbook, a forum where learners and staff could engage in discussions, a staff-learner editable wiki, and ungraded access to the interactive circuit simulator.

The course schedule comprised 15 calendar weeks. A chapter was released every Monday except for week 1, when it was made on Wednesday, and week 8, in which the midterm exam took place and there was no chapter release. The final exam was made in the last week of the course. The deadline for the submission of the assignments included in each of the first 12 chapters was set for the second Sunday after the corresponding chapter release, except in the case of chapter 7, in which it was the third Sunday also due to the occurrence of the midterm exam. This implies that the deadline for the assignments of a given chapter always took place nearly one week after the release date of the next chapter.

Final course grades were based on homework sets (15%), online laboratories (15%), a midterm (30%) and a final exam (40%). Each chapter had a homework grade and a lab grade. The homework and laboratories grades used in the final course grades were obtained by adding the highest 10 out of 12 chapter grades. Learners with a final grade of 50% or greater received a certificate of accomplishment.

There were 26,947 learners enrolled in the course by the deadline established for the submission of the final exam. Out of these, only 6,752 watched at least a lecture video or answered at least a finger exercise or submitted at least an assignment in one of the first 12 chapters before the corresponding deadline. A certificate of accomplishment was granted to 1,099 learners.

4.2 Engagement Indicators

As in many other MOOCs, watching lecture videos, answering finger exercises and submitting assignments were the three main tasks that learners were expected to carry out in the course introduced in the previous subsection. This is the reason why a different academic engagement indicator was defined for each of these tasks as described next.

The *video engagement indicator* can be obtained by averaging the percentages of lecture videos that were totally or partially watched by a learner in every chapter before reaching its end. The *exercise engagement indicator* can be computed by averaging the percentages of finger exercises answered by a learner in every chapter before its end. The *assignment engagement indicator* can be calculated by averaging the percentages of assignments submitted in each chapter that has already ended.

4.3 Labels and Features

The values of the three engagement indicators were computed for every learner at the end of each of the first 12 chapters. They were not calculated for chapters 13 and 14 since they did not include any assessment. Then for each chapter $i \in \{1, \ldots, 11\}$, a label is derived for every learner and each engagement indicator stating whether the value of the indicator was lower by the end of chapter $i + 1$ than it was at the end of chapter i. These labels will be used as the outputs to train and test the models, as explained in section 3.

The values of 16 features were also computed for every learner at the end of each of the first 11 chapters. As it can be seen in Table 1, these features describe the activity of the corresponding learner in the course from its beginning until the end of the chapter for which they are computed. More specifically, 4 features characterized video watching activity ($v1$ to $v4$), 6 features described the activity regarding finger exercises ($e1$ to $e6$), and other 6 features represented the activity regarding assignments ($a1$ to $a6$).

It can be noticed that every feature was defined so that its range of values is the same across chapters. It is also worth mentioning that the existence of features related to chapter $i + 1$ calculated at the end of chapter i is possible since, as explained before, the end of every chapters 1 to 12 took place nearly one week after the start of the next chapter. Interestingly, a value of 0 or a negative value in $v4$, $e4$, or $a4$ implies that the corresponding engagement indicator will not decrease by the end of chapter $i+1$ since the learner has already done enough work by the end chapter i as to assure it.

Id.	Description
$v1$	Percentage of lecture videos totally or partially watched in chapter i
$v2$	Value of video engagement indicator at the end of chapter i
$v3$	Percentage of lecture videos totally or partially watched in chapter $i+1$
$v4$	Difference between value of video engagement indicator and percentage of lecture videos totally or partially watched in chapter $i+1$ ($v2-v3$)
$e1$	Percentage of finger exercises answered in chapter i
$e2$	Value of exercise engagement indicator at the end of chapter i
$e3$	Percentage of finger exercises answered in chapter $i+1$
$e4$	Difference between value of exercise engagement indicator and percentage of finger exercises answered in chapter $i+1$ ($e2-e3$)
$e5$	Normalized grade of finger exercises in chapter i
$e6$	Normalized total grade of finger exercises in chapters 1 to i
$a1$	Percentage of assignments submitted in chapter i
$a2$	Value of assignment engagement indicator at the end of chapter i
$a3$	Percentage of assignments submitted in chapter $i+1$
$a4$	Difference between value of assignment engagement indicator and percentage of assignments answered in chapter $i+1$ ($a2-a3$)
$a5$	Normalized grade of assignments in chapter i
$a6$	Normalized total grade of assignments in chapters 1 to i

Table 1: Features derived at the end of chapter i to describe the activity of a learner regarding video watching, finger exercises and assignments.

4.4 Number of Samples

A dataset was created for each engagement indicator and chapters 1 to 11. Each dataset contains the learners' feature vectors computed at the end of a given chapter along with the corresponding stating whether the engagement indicator decreased or not at the end of the next chapter after filtering out two types of samples. First, samples in which it is already known at the end of a given chapter that the target engagement will not decrease at the end of the next chapter, since there is no need to make any prediction. Second, samples from learners who have not watched any lecture video, answered any finger exercise of submitted any assignment in the last three chapters since it is assumed that they have dropped out. Table 2 shows the number of samples included in each dataset.

It must be noticed that the fact that chapters 13 and 14 did not include any assignment implies not only that the assignment engagement indicator cannot be computed, but also that a chapter end cannot be defined according to our problem formulation. This also entails that the video and exercise engagement

	Video engagement	**Exercise engagement**	**Assignment engagement**
Chapter 1	4,637	2,863	3,456
Chapter 2	5,021	3,218	3,895
Chapter 3	5,374	3,395	4,056
Chapter 4	3,206	2,465	3,125
Chapter 5	2,501	1,974	2,295
Chapter 6	2,110	1,709	1,924
Chapter 7	1,759	1,451	1,659
Chapter 8	1,536	1,281	1,417
Chapter 9	1,349	1,173	1,265
Chapter 10	1,186	1,042	1,158
Chapter 11	1,069	954	1,120

Table 2: Number of samples in each dataset.

indicators can neither be computed for such chapters. As a consequence, datasets with features gathered at the end of chapters 12 and 13 and indicators computed at the end of the same chapters 13 and 14 could not be generated.

5 Experiments

5.1 Feature Selection

All datasets underwent a feature selection process prior to building the prediction models with the aim of reducing the potential overfitting as well as of increasing their accuracy. More specifically, the Correlation based Feature Selection (CFS) method [Hall, 1999] was employed to identify the most relevant features for the prediction of the decrease of engagement in every dataset corresponding to each engagement indicator and chapter. CFS was chosen because, unlike other methods, it aims not only at finding subsets of features that have high individual prediction ability, but also a low degree of redundancy among them. Furthermore, CFS is not a computationally intensive method. Figure 2 shows the number of chapters in which each feature was selected.

It can be observed that 2 video features ($v1$ and $v2$), 1 exercise feature ($e3$) and 2 assignment features ($a4$ and $a5$) were considered useful for prediction in most datasets related to video engagement. In the case case of exercise engagement, 1 video feature ($v1$), 5 exercise features ($e1$, $e2$, $e3$, $e5$ and $e6$) and 2 assignment features ($a4$ and $a5$) were selected for prediction in most datasets. Furthermore, 1 exercise feature ($e5$) and 4 assignment features ($a1$, $a2$, $a5$ and $a6$) were kept in most datasets related to assignment engagement. This implies

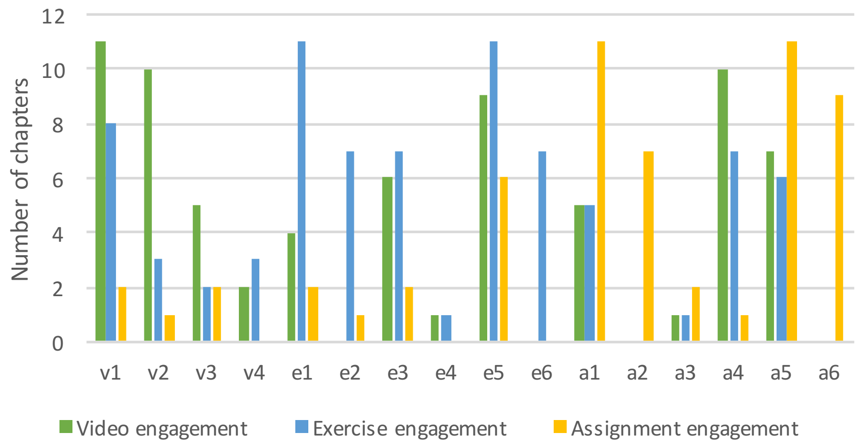


Figure 2: Number of chapters in which each feature was selected for each engagement indicator using CFS.

that the prediction of the decrease of any of the three indicators benefits not only from features that describe the activity related to the predicted indicator, but also from the features that describe the activity related to other indicators.

It is also interesting to see that two features related with the grades of exercises and assignments ($e5$ and $a5$) were selected in most datasets for all engagement indicators. This fact suggests the importance of grades for prediction regardless of the type of engagement indicator to be predicted.

It is worth noting that features measuring the accumulated variation of the video, exercise and assignment engagement indicators half way through the chapter ($v4$, $e4$ and $a4$) are not very relevant to predict their decrease at the end of the chapter, i.e. the decrease of the indicator cannot be anticipated by simply looking at its current value. This is due to the fact that, as noted before, predictions are not generated for learners that at mid-chapter have already reached an indicator value higher than in previous chapter. Curiously, $a4$ is very informative to predict the video and exercise indicators but not the assignment indicator. A visual inspection of $a2$ revealed that learners with a high assignment engagement in previous chapter are more likely not to decrease their engagement with videos, exercises or assignments by the end of the chapter. It was also observed in $a3$ that learners with a high assignment engagement in the great majority of cases decrease their engagement with video or finger exercises by the end of the chapter, but in most cases do not to decrease their engagement with assignments. As a consequence, the subtraction of $a3$ from $a2$ makes $a4$ more separable in the case of video and exercise engagement but not in the case of assignment engagement.

5.2 Prediction

Logistic regression was employed to build the models for the prediction of the three engagement indicators following both the *a posteriori* and the *in-situ* approaches. Logistic regression was selected because several works, including [Kizilcec and Halawa, 2015, Boyer and Veeramachaneni, 2015], have already reported its good performance in the related problem of dropout prediction.

More specifically, *a posteriori* prediction models were built and tested using 10-fold cross validation with each dataset. Both model training and testing were made the features selected by CFS for the corresponding dataset. *In situ* prediction models were built using the datasets of chapters 1 to 10 and then tested employing the datasets of chapters 2 to 11, respectively (keep in mind that the dataset of chapter i consists of the features from the activity until the end of chapter i, as input, while the output is a label reflecting whether the engagement indicator at the end of chapter $i+1$ decreases or not with respect to its value at the end of chapter i). The trainings were carried out using the features selected by CFS for each dataset. Obviously, the features employed in the training phase of each model were also used in the corresponding test phase.

A simple baseline predictor was also employed to make predictions in order to better assess the performance of the prediction models built using both approaches. The baseline simply predicts that the value of a given engagement indicator will decrease at the end of chapter of chapter $i+1$ with respect to the end of chapter i if it also decreased at the end of chapter i with respect of the end of chapter $i-1$. Otherwise, the baseline predicts that the value of the indicator will not decrease. Note that this baseline is to be applied *in situ* and can thus only be evaluated in the datasets of chapters 2 to 11.

The performance of all predictors was measured using area under the curve (AUC). It must be noticed that AUC has been preferred to other performance metrics such as accuracy or Cohen's kappa because, unlike them, AUC is not affected by imbalanced distributions of data [Jeni et al., 2013]. AUC informs of the probability that a model correctly predicts the decrease of an engagement indicator for a randomly selected learner. In this way, a useless predictor will have an AUC of 0.5 while a perfect predictor will feature AUC 1. Following [Swets, 1988], model performance can be categorized in general as excellent for values of AUC greater than 0.9, good between 0.8 and 0.9, fair between 0.7 and 0.8, and poor below 0.7.

The results of the prediction tests carried out with each dataset are reported in Figure 3. It can be noticed that results are not provided for chapter 1 using *in* *situ* models and the baseline since, as explained before, they both make predictions for a given chapter based on the information obtained from the previous chapter. However, they are provided for *a posteriori* models since they are built using information from the chapter for which they generate predictions. Table 3 shows the weighted average AUC across chapters 2 to 11 achieved by each approach employed for the prediction of the decrease of the different engagement indicators. It is important to remember that predictions were not made for learners that did not show any activity in three chapters in a row, thus avoiding the artificial improvement of the results with predictions that would have been very easy to make.

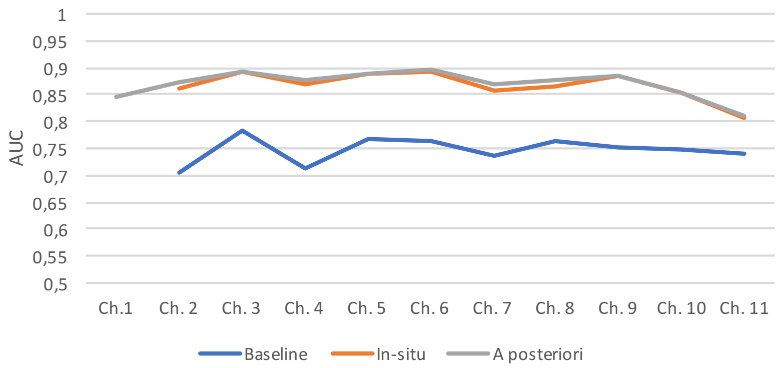
	Decrease of video engagement	**Decrease of exercise engagement**	**Decrease of assignment engagement**
A posteriori	0.878	0.893	0.874
In situ	0.872	0.882	0.864
Baseline	0.744	0.781	0.799

Table 3: Comparison of weighted average AUC achieved across chapters 2 to 11.

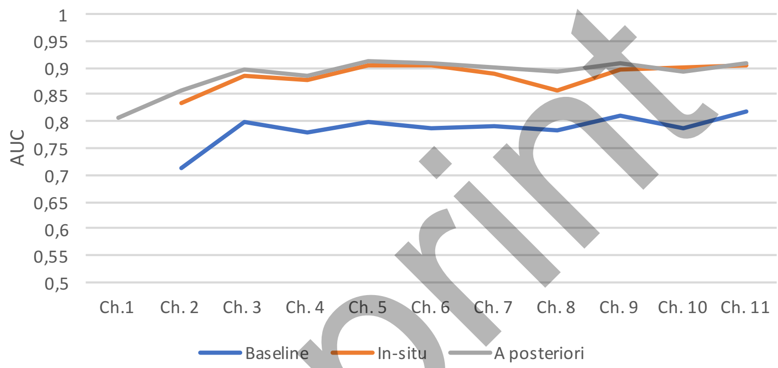
As expected, the *a posteriori* models yielded the best results. On average, they showed a good performance for the prediction of the decrease of all engagement indicators. However, as explained above, the predictions made with these models cannot be used to trigger interventions in an ongoing course based on its data.

Remarkably, the performance of *in situ* models was very similar to that of *a posteriori* models in all cases except for the prediction of the decrease of both exercise and assignment indicator in chapter 8. In spite of this, the average performance of *in situ* models was good for all indicators showing weighted average AUC values that, in the worst case, was just 0.011 below that of the *a posteriori* models. Nevertheless, the use of *in situ* models do enable the possibility of making interventions in an ongoing course using the data available at the moment in which the predictions are required.

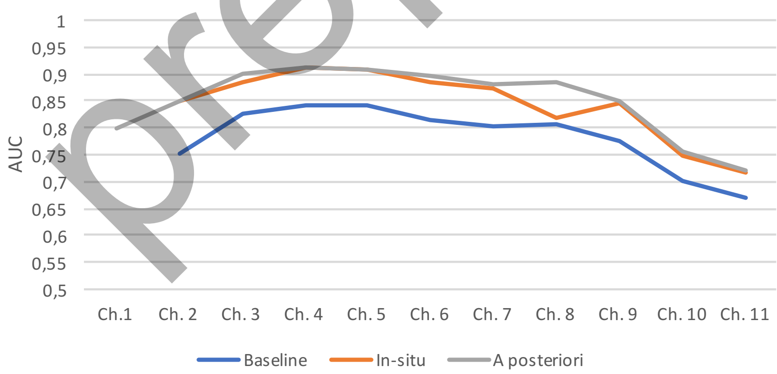
The baseline showed on average a fair performance for the prediction of the decrease of all indicators. Even so, the *in situ* prediction models performed clearly better than the baseline in all cases except for the prediction of the assignment engagement indicator at the end of chapter 8. In fact, the weighted average AUC values of *in situ* models were between 0.075 and 0.134 higher than those of the baseline. A Wilcoxson signed rank test also showed that the difference of performance between the *in situ* models and the baseline was statistically sig-



(a) Decrease of video engagement.



(b) Decrease of exercise engagement.



(c) Decrease of assignment engagement.

Figure 3: Comparison of AUC achieved in each chapter.

nificant ($p < 10^{-5}$). This supports the idea that the *in situ* models can harness the selected features describing the activity of learners to bring an important performance improvement with respect to the baseline.

It can be observed in Figure 3 that there is a noticeable increase in the AUC values obtained by all prediction approaches in chapter 3 with respect of the previous chapters in the case of exercise and assignment indicators and, to a lesser extent, of video indicator. This increase can be attributed, at least in part, to the fact that many learners abandoned the course before the end of chapter 3 even though they had been participating very actively, just as many others that did not abandon. This is the reason why it is difficult to predict the decrease of engagement indicators of these participants, known as "strong starters" [Ferguson and Clow, 2015], in the first course chapters.

It can also be noticed in Figure 3(c) that the AUC values obtained in the prediction of the assignment engagement indicator for chapters 10 and 11 are clearly lower than for the rest of the datasets. This could be due to the fact that the lowest two assignment grades obtained in any chapter were not taken into account to compute the final grade. This idea is supported by the fact that an inspection of the datasets revealed the existence of a relevant number of learners that maintained a high level of assignment engagement and had obtained high grades in all previous chapters that significantly reduced their activity in the last two chapters.

6 Discussion

The results of the experiments presented in previous section show that the *in situ* predictions models built using a selection of features that have a low degree of redundancy among them yielded a good performance in the prediction of the decrease of engagement indicators derived for the three main tasks that were carried out by learners in a MOOC. As noted before, predictions were made at the end of chapters 2 to 11 using only information that was available at those moments. It is noteworthy that during the experiments, predictions were not made for obvious cases, such as learners that at mid-chapter have already shown higher engagement than in the previous chapter or learners that can be considered to have dropped out. Adding these learners would easily improve the prediction performance metrics, but would not provide any useful information in order to intervene. The *in situ* predictions models built following the approach proposed in this paper would thus have been useful to identify many learners that could have benefited from an intervention aimed at preventing their disengagement.

One limitation of the proposed approach is that the first predictions cannot be made before the end of chapter 2. Predictions cannot be made at the end of chapter 1 since there are obviously no data from previous chapters that can be employed to build the predictive model. This problem would not exist in a different approach that could use data from previous editions of the course along with transfer learning techniques to build predictors. However, such approach would have the important limitation that it could not be applied to make predictions in courses that do not have a previous edition. If there were data from a previous run of the course, or for courses with a similar start, a hybrid prediction scheme could be derived in which transfer learning is employed to build predictors at the end of chapter 1, and then the *in situ* approach is followed for the rest of the chapters, using data from the ongoing course.

Another limitation of the proposed approach that can be observed in the results of the experiments is that predictions regarding the decrease of the video and exercise engagement indicators could not be made for chapters 13 and 14. Tackling this limitation could require not only changing the definition of end of chapter so that it is only based on the deadlines for submissions. Possibly, it will also entail defining a method to build the predictors taking into account that the behavior of learners in chapters that do not have assignments might be very different with respect to chapters in which they have them. Again, transfer learning techniques could be useful to build these predictors.

Features derived from other types of data (e.g. forum events, time between learner's consecutive events) and the information about the course structure and requirements (the learning design) could be further exploited to improve predictive performance. So far, only the temporal information about chapter start and submission deadlines is considered to generate features that describe the learners' behavior. Introducing in the model the knowledge about how the final grade is determined (only the 10 highest scores among 12 submissions will be averaged) could have improved the prediction in the last two chapters. This could be tailored by defining *ad hoc* features using the knowledge about the design of each particular course.

It is noteworthy that the results of the *in situ* prediction models reported in this paper are comparable to those reported in [Whitehill et al., 2017] for the dropout prediction problem. More specifically, AUC values around 0.9 were obtained using *a posteriori* models and values ranging from 0.85 to 0.9 using *in situ* models. Interestingly, the performance of the *in situ* models for dropout prediction also followed closely that of the *a posteriori* models.

The performance of the proposed approach thus opens the possibility of making a wide range of interventions aimed at maintaining the engagement of learners with the specific activity for which a disengagement is predicted. Examples include simple strategies such as sending mails reminding pending activities or providing learners with hints to solve an assignment, or more elaborate approaches such as using a recommendation system to make suggestions of videos or assignments that are expected to be engaging for the target learner.

7 Conclusions and Future Work

Predicting the decrease of learners' academic engagement indicators of is key to avoid the disengagement before it happens. *A posteriori* models cannot be built using the data that is available in a MOOC at the moment the predictions are required, thus precluding the possibility of making interventions in the course in real time. This paper proposes an approach to build *in situ* models using such information so that their predictions can be employed to trigger timely interventions in an ongoing course.

Experiments were carried out to compare the performance of models built following both approaches as well as of a simple baseline for the prediction of three engagement indicators derived for the main tasks that are typically proposed in a MOOC: watching lectures, solving finger exercises, and submitting assignments. Prediction models were built following *in situ* and *a posteriori* approaches using CFS method for feature selection and logistic regression for classification. The *in situ* models exhibited a good prediction performance which was very similar to that of *a posteriori* models and significantly higher that that of the baseline. This supports the idea that the *in situ* models would have been useful to detect disengaging learners in that MOOC and suggests that *in situ* models could be useful in other MOOCs too.

The next steps in this research will be to evaluate the proposed approach in other MOOCs and to study its applicability to engagement indicators derived from other MOOC activities, including those of a more social nature such as posting in forums or participating in peer reviews. The possibility of improving predictions using features that describe demographic, including year of birth or level of education, or aspects of learner participation in MOOCs other than behavior, such as motivation, will be explored. Besides, the limitation regarding the lack of predictions at the end of the first chapter or for chapters that do not include assignments will be addressed. Future work also includes the design of intervention mechanisms that could be triggered when the decrease of an engagement indicator is predicted. In addition, there are plans to run a MOOC in which these intervention mechanisms will be used along with *in situ* prediction models with the aim of assessing their value for both instructors and learners.

Acknowledgments

Access to the data employed in this paper was kindly granted by MIT's Institutional Research Office. This work has been partially funded by projects TIN2014-53199-C3-2-R (AEI, FEDER), TIN2017-85179-C3-2-R (AEI, FEDER), VA277U14 (Junta de Castilla y León, FEDER), and 588438-EPP-1-2017-1-EL-EPPKA2-KA (European Commission).

References

[Alario-Hoyos et al., 2014] Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Parada, H. A., and Muñoz-Organero, M. (2014). Delving into participants' profiles and use of social tools in MOOCs. *IEEE Transactions on Learning Technologies*, 7(3):260–266.

[Anderson et al., 2014] Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2014). Engaging with Massive Online Courses. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 687–698.

[Beck, 2005] Beck, J. E. (2005). Engagement tracing: using response times to model student disengagement. In *Proceedings of the 2005 Conference on Artificial Intelligence in Education*, pages 88–95, Amsterdam, The Netherlands.

[Boyer and Veeramachaneni, 2015] Boyer, S. and Veeramachaneni, K. (2015). Transfer learning for predictive models in Massive Open Online Courses. In *Proceedings of the 17th Conference on Artificial Intelligence in Education, AIED 2015*, pages 54–63.

[Chang and Wei, 2016] Chang, J. and Wei, H. (2016). Exploring engaging gamification mechanics in Massive Online Open Courses. *Educational Technology & Society*, 19(2):177–203.

[Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S. (2011). Disengagement detection in online learning: Validation studies and perspectives. *IEEE Transactions on Learning Technologies*, 4(2):114–124.

[Ferguson and Clow, 2015] Ferguson, R. and Clow, D. (2015). Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs). In *Proceedings of the 5th Conference on Learning Analytics And Knowledge*, LAK '15, pages 51–58.

[Ferguson et al., 2015] Ferguson, R., Clow, D., Beale, R., Cooper, A. J., Morris, N., Bayne, S., and Woodgate, A. (2015). Moving through MOOCS: Pedagogy, learning design and patterns of engagement. In *Proceedings of the 10th European Conference on Technology Enhanced Learning, EC-TEL 2015*, pages 70–84.

[Finn and Zimmer, 2012] Finn, J. D. and Zimmer, K. S. (2012). *Student engagement: what is it? Why does it matter?*, pages 97–131. Springer US, Boston, MA.

[Fredricks et al., 2004] Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1):59–109.

[Gardner and Brooks, 2018] Gardner, J. and Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, 28(2):127–203.

[Goldberg et al., 2015] Goldberg, L. R., Bell, E., King, C., O'Mara, C., McInerney, F., Robinson, A., and Vickers, J. (2015). Relationship between participants' level of education and engagement in their completion of the Understanding Dementia Massive Open Online Course. *BMC Medical Education*, 15(1):60.

[Halawa et al., 2014] Halawa, S., Greene, D., and Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. In *Proceedings of the 2nd European MOOCs Stakeholders Summit*, pages 58–65, Laussane, Switzerland.

[Hall, 1999] Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. PhD thesis, University of Waikato, Hamilton, New Zealand.

[Jeni et al., 2013] Jeni, L. A., Cohn, J. F., and De La Torre, F. (2013). Facing imbalanced data–recommendations for the use of performance metrics. In *Proceedings of the 2013 Conference on Affective Computing and Intelligent Interaction*, ACII '13, pages 245–251.

[Joksimović et al., 2016] Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., and de Kereki, I. F. (2016). Translating network position into performance: Importance of centrality in different network configurations. In *Proceedings of the 6th Conference on Learning Analytics & Knowledge*, LAK '16, pages 314–323.

[Jordan, 2015] Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning*, 16(3).

[Kizilcec and Halawa, 2015] Kizilcec, R. F. and Halawa, S. (2015). Attrition and achievement gaps in online learning. In *Proceedings of the 2nd ACM Conference on Learning @ Scale*, L@S '15, pages 57–66, New York, NY, USA. ACM.

[Kizilcec et al., 2013] Kizilcec, R. F., Piech, C., and Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the 3rd Conference on Learning Analytics & Knowledge*, LAK '13, pages 170–179.

[Li et al., 2017] Li, X., Wang, T., and Wang, H. (2017). Exploring n-gram features in clickstream data for MOOC learning achievement prediction. In *Database Systems for Advanced Applications*, pages 328–339.

[Milligan et al., 2013] Milligan, C., Littlejohn, A., and Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *Online Learning and Teaching*, 9(2):149–159.

[Mills et al., 2014] Mills, C., Bosch, N., Graesser, A., and D'Mello, S. (2014). To quit or not to quit: Predicting future behavioral disengagement from reading patterns. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems, ITS 2014 Conference on Intelligent Tutoring Systems*, ITS 2014, pages 19–28.

[Newmann, 1992] Newmann, F. M. (1992). *Student engagement and achievement in American secondary schools*. Teachers College Press, New York.

[Ramesh et al., 2013] Ramesh, A., Goldwasser, D., and Huang, B. (2013). Modeling learner engagement in MOOCs using probabilistic soft logic. In *Proceedings of NIPS Workshop on Data Driven Education*.

[Ramesh et al., 2014] Ramesh, A., Goldwasser, D., Huang, B., Daume, III, H., and Getoor, L. (2014). Uncovering hidden engagement patterns for predicting learner performance in MOOCs. In *Proceedings of the 1st ACM Conference on Learning @ Scale Conference*, L@S '14, pages 157–158.

[Rizzardini and Amado-Salvatierra, 2017] Rizzardini, R. H. and Amado-Salvatierra, H. R. (2017). Full engagement educational framework: A practical experience for a micromaster. In *Proceedings of the 4th ACM Conference on Learning @ Scale*, L@S '17, pages 145–146.

[Seaton et al., 2014] Seaton, D. T., Reich, J., Nesterko, S. O., Mullaney, T., Waldo, J., Ho, A. D., and Chuang, I. (2014). 6.002x Circuits and Electronics MITx on edX course report - 2013 spring. Technical Report 8, MIT Office of Digital Learning, and the HarvardX Research Committee.

[Sinclair and Kalvala, 2016] Sinclair, J. and Kalvala, S. (2016). Student engagement in Massive Open Online Courses. *Int. J. Learn. Technol.*, 11(3):218–237.

[Swets, 1988] Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857):1285–1293.

[Vitiello et al., 2017] Vitiello, M., Walk, S., Chang, V., Hernandez, R., Helic, D., and Guetl, C. (2017). MOOC dropouts: A multi-system classifier. In *Data Driven Approaches in Digital Education*, pages 300–314.

[Whitehill et al., 2017] Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., and Tingley, D. (2017). MOOC dropout prediction: How to measure accuracy? In *Proceedings of the 4th ACM Conference on Learning @ Scale*, L@S '17, pages 161–164.

[Xing and Du, 2018] Xing, W. and Du, D. (2018). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*. In press.

[Xu and Yang, 2016] Xu, B. and Yang, D. (2016). Motivation classification and grade prediction for MOOCs learners. *Comp. Int. and Neurosc.*, 2016:2174613:1–2174613:7.

[Zheng et al., 2015] Zheng, Z., Vogelsang, T., and Pinkwart, N. (2015). The impact of small learning group composition on student engagement and success in a MOOC. In *Proceedings of the 8th International Conference on Educational Data Mining, EDM 2015*, pages 500–503, Madrid, Spain.