



The Road Not Taken: Preempting Dropout in MOOCs

Lele Sha¹(✉), Ed Fincham¹, Lixiang Yan¹, Tongguang Li¹, Dragan Gašević^{1,2},
Kobi Gal^{2,3}, and Guanliang Chen¹(✉)

¹ Centre for Learning Analytics, Faculty of Information Technology, Monash University,
Clayton, Australia

{lele.sha, lixiang.Yan, tongguang.Li, dragan.gasevic,
guanliang.chen}@monash.edu, ed.fincham@proton.me

² School of Informatics, University of Edinburgh, Edinburgh, UK

kgal@exseed.ed.ac.uk

³ Department of Software and Information Systems, Ben-Gurion University, Be'er Sheva, Israel

Abstract. Massive Open Online Courses (MOOCs) are often plagued by a low level of student engagement and retention, with many students dropping out before completing the course. In an effort to improve student retention, educational researchers are increasingly turning to the latest Machine Learning (ML) models to predict student learning outcomes, based on which instructors can provide timely support to at-risk students as the progression of a course. Though achieving a high prediction accuracy, these models are often “black-box” models, making it difficult to gain instructional insights from their results, and accordingly, designing meaningful and actionable interventions remains to be challenging in the context of MOOCs. To tackle this problem, we present an innovative approach based on Hidden Markov Model (HMM). We devoted our efforts to model students’ temporal interaction patterns in MOOCs in a transparent and interpretable manner, with the aim of empowering instructors to gain insights about actionable interventions in students’ next-step learning activities. Through extensive evaluation on two large-scale MOOC datasets, we demonstrated that, by gaining a temporally grounded understanding of students’ learning processes using HMM, both the students’ current engagement state and potential future state transitions could be learned, and based on which, an actionable next-step intervention tailored to the student current engagement state could be formulated to recommend to students. These findings have strong implications for real-world adoption of HMM for promoting student engagement and preempting dropouts.

Keywords: Hidden Markov Models · MOOCs Dropout · Student Engagement

1 Introduction

Over the last decade, the proliferation of MOOCs has provided millions of students with unprecedented access to open educational resources. However, from their conception, MOOCs have attracted widespread criticism due to students’ limited interactions with the course materials and high dropout rates [17]. Driven by this, an extensive body

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

N. Wang et al. (Eds.): AIED 2023, LNAI 13916, pp. 164–175, 2023.

https://doi.org/10.1007/978-3-031-36272-9_14

of research has developed accurate predictive models of student dropout [3,9]. Despite achieving a high accuracy (around 90% AUC [17]) with increasingly more advanced Machine Learning (ML) models, these models provide a limited understanding of the underlying phenomenon, i.e., how a student disengages from the course which ultimately leads to a dropout. In providing course instructors with relevant insights, on the one hand, educational researchers have attempted to represent students' interactions as a set of "prototypical", richly descriptive patterns of student activity [12,15], from which a high-level description of students' engagement with the course material may emerge (e.g., "on-track", "behind" or "disengaged" [12]). However, studies of this strand of research often failed to provide actionable insights to instructors/students, i.e., how to persist student in an on-track state, or how to prevent students from being disengaged. On the other hand, researchers have foregone the understanding of student engagement and have instead aimed to directly predict a future learning path towards successful completion of a course [16,25]. Though achieving a high prediction accuracy (up-to 80%), these approaches relied on "black-box" models, making it difficult for instructors/students to embrace such suggestions as actionable. In an effort to bridge these two disparate fields (i.e., understanding student engagement and providing actionable learning suggestions), and develop models that offer both interpretable insights as well as actionable suggestions about improving student retention in the future, we posit that it is essential to consider students' engagement patterns as a *non-static* phenomenon which may change with time, e.g., an "on-track" student may be demotivated and become less engaged with the course materials [2]. By taking such *temporal dynamics* of engagement (i.e., how students' engagement evolves as the course progresses) into account, meaningful future actions could be inferred directly based on the current engagement state for improving student retention.

As such, we investigated the following **Research Questions (RQs)**: **RQ 1** – To what extent can *temporal engagement patterns* be identified that are reflective of student retention in MOOCs? **RQ 2** – How can these patterns be used for *promoting engagement* and *preempting dropout*? To address RQ 1, we took temporal dynamics into account by utilising a fully transparent and interpretable model for analyzing sequential data (e.g., students' interaction with different course materials over time), namely, a Hidden Markov Model (HMM). The HMM consists of a set of latent states, which represent discrete points in time, along with a transition matrix that governs the switching dynamics between them. So we may capture the underlying latent states behind students' interactions, and the potential future transitions between different states. To address RQ 2, we further explored the use of HMM-identified patterns and transitions as a basis to suggest next-step actionable learning activities to students (i.e., the learning material to be accessed at the next step) based on their current engagement states. Through extensive evaluation on the two MOOCs datasets, we demonstrated that HMM may capture not only richly descriptive patterns about student interactions (termed as engagement patterns) in the hidden states, but also salient information regarding the future engagement trend (e.g., whether a student may continue to engage with the course). This indicates that HMM may be a valuable tool for researchers and practitioners to understand the temporal dynamics of student engagement. Besides, in a simulated dropout study, the HMM-suggested next-step learning activity is the only one that can

achieve simultaneously a high contribution to student retention and a high probability of being performed by a student given their current engagement state, compared to the two other baselines, thereby providing strong motivation for future real-world evaluation and adoption.

2 Related Work

Interpreting Interactions; Illuminating Engagement. While students' engagement may be opaque, the engagement state (i.e., the underlying state which drives how a student interacts with course material) generally manifests itself as a mediating role in students' subsequent behaviour [3, 15]. This behaviour is observable and may be directly modeled [3, 12, 18, 21]. For instance, Herskovits et al. [8] proposed that diverse interaction patterns reflect high levels of engagement. To quantify this diversity, the authors used principal component analysis at regular time intervals to project student's interactions along the top three principal components. These interactions were then clustered into common trajectories. In a similar vein, Kizilcec et al. [12] adopted a simple but effective k-means clustering approach and showed that high-level prototypical engagement trajectories consistently emerged from three different computer science MOOCs. Then, the later work of Coleman et al. [1] rejected the curation of rigidly defined feature sets and instead used Latent Dirichlet Allocation (LDA) to discover behavioural patterns, or "topics", directly from students' interactions (in an unsupervised manner). By representing students as a mixture of these latent types, students may be characterised by a distribution over the set of course resources (termed as "use-cases"), e.g., a shopping use case (where students only access the free components of the course) vs. a disengaging use case (where students interaction gradually attenuated as the course progresses). The shortcoming of these prior approaches is that they treat student engagement patterns mostly as a static phenomenon, where user activity is represented as a "bag-of-interactions", which ignores the temporal dependency that is inherent to the data – a student may be, for example, actively engaged for the first week of a course, before transitioning into a state of disengagement in the following weeks. We argue that such a temporal factor could bring valuable insights about how and when a student first started to disengage, which subsequently may inform relevant engagement strategies to preempt their dropout.

Promoting Engagement; Preempting Dropout. While the combination of richly descriptive models of students' interactions with the improvement of learning outcomes (e.g., dropout) has been largely overlooked by the literature, a number of fields come close. For instance, the literature surrounding learning strategies (defined as "thoughts, behaviours, beliefs, or emotions that facilitate the acquisition, understanding, or later transfer of new knowledge and skills" [8]), identify common interaction patterns which are associated with learning outcomes [4, 11, 13, 14, 20]. While these studies generally found that students' learning strategies/trajectories were associated with their learning outcome, they did not explicitly model this relationship, so were unable to evaluate the extent to which different learning strategies are contributing to a dropout/non-dropout of a student. In a different vein, researchers have forgone the notion of student

engagement and instead, attempted to generate a learning path (i.e., the learning materials/resources a student accesses at the next step in order to achieve a certain learning outcome) directly from the latest black-box ML models. For instance, Mu et al. [16] attempted to suggest learning resources based on a student’s past learning experience using a XGBoost coupled with an explainer model. So the past activity with the highest feature importance (based on model explanation) and an undesirable learning outcome were diagnosed as important actions to be performed by students. Similarly, Zhou et al. [25] recommended learning resources based on students’ expected future performance, which are predicted using a deep neural network-based model. Although achieving a high accuracy (up-to 80%) at recommending the correct resources, these approaches were limited from the perspective of a learner’s dynamic engagement state, as the learning paths are separately modelled from how a student may actually interact with the course material given their current engagement state (e.g., recommending highly effective strategies to a “shopping” student who merely intended to browse the free component of the course is considered pointless). Besides, black-box models lack transparency and interpretability, which can make it challenging to gain educational insights from the results of these models and subsequently limit actual adoption in real-world educational scenarios. Driven by this, we adopted a fully transparent HMM model to analyse student engagement dynamics, which we then use for informing an engaging next-step learning activities based on a student’s current engagement state. To our knowledge, our approach is the first to suggest actionable activities based on HMM.

3 Method

3.1 Dataset

To ensure a robust analysis on student engagement in MOOCs, we adopted two MOOCs datasets in the evaluation. The first dataset **Big data in Education** (denote as **Big data**) was a course offered by Columbia University in 2013 and delivered on the Coursera platform. In this course, students learned a variety of educational data-mining methods and applied them to research questions related to the design and improvement of interventions within educational software and systems. The course material consisted of lecture videos, formative in-video quizzes, and 8 weekly assignments. The second dataset **Code Yourself** was offered by the University of Edinburgh in 2015 on the Coursera platform. The course was designed to introduce teenagers to computer programming, while also covering basic topics in software engineering and computational thinking. The course materials consisted of lectures, videos, formative in-video quizzes, peer-reviewed programming projects, and 5 weekly assignments. Although the two courses had initial enrolments of over 45,000 and 59,900, respectively (as shown in the statistics Table 1), a large proportion of these students did not actively participate, and only 18,222 and 26,514 accessed at least a single item in the course; we restrict our study to these students. Of these active students, only 3,600 and 10,183 was active the entire course, and only about 2% of all students have successfully completed the course. This low completion rate can be attributed to a high rate of student attrition – common across MOOC environments – which, although significantly attenuated after the first weekly assignment, remained substantial throughout the course.

Table 1. The descriptive statistics of the two MOOCs datasets used in this study. The fractions within brackets indicate the percentage of No. enrolled.

Dataset	Duration	No. enrolled	No. dropout	No. active learners		
				At least once	Entire course	Completed
Big data	8 weeks	45,000	41,400 (92.00%)	18,222 (40.49%)	3,600 (8.00%)	638 (1.42%)
Code yourself	5 weeks	59,900	49,717 (83.00%)	26,514 (44.26%)	10,183 (17.00%)	1,591 (2.66%)

3.2 Modeling Student Engagement by HMM

Hidden Markov Model (HMM) is a well-established unsupervised machine learning approach that helps to make inferences about the latent variables (i.e., hidden states) through analyzing the observable actions [6, 15]. Importantly, we opted for a variant known as sticky HMM [6]. The “sticky” assumption in the Markov model is that once a student has adopted a particular state, they persist in that state for as long as possible until a new state is required to describe their actions. Not only does this assumption represent many scenarios in real-world data, where states persist through time [6], but it also helps combat the unrealistically rapid switching dynamics that are present in models without this state-persistence bias [5]. The sticky HMM requires us to specify the number of states. However, to mitigate the impact this has on our model, we place a non-parametric prior over the state space [6]. The implementation uses a weak-limit approximation to the hierarchical Dirichlet process [24], which approximates the unbounded state space by a truncated representation with τ states, where we specify $\tau = 10$ [15]. The prior places a diminishing probability mass on infrequent states and thus focuses the majority of the probability mass on a relatively small number of major states in the model. For brevity, we refer to the sticky HMM as an HMM in this paper.

RQ 1: Understanding student engagement. In our engagement analysis, we used student interaction in MOOC as observable actions to model the student engagement state as hidden states. As we were interested in the temporal relation of student interaction, we represented students’ interactions on a *weekly* basis. Specifically, we created a 10 element vector. The first three elements of the vector represented whether the student took actions for each type (i.e., assignment a , lecture l , quiz q) from the past week (denote as a_p, l_p and q_p respectively); the next three elements represented actions from the current week; the next three elements represented actions from the future week; and finally, the last element indicated whether the student interacted with any resources at all during the current week (denote as o). Therefore, during each week, a student G_i is represented by the 10 binary-element vector: $G_i = [a_p, l_p, q_p, a_c, l_c, q_c, a_f, l_f, q_f, o]$. By doing so, we can model temporal engagement and understand how a student engagement from the perspective of their interaction with the past, current, and future weeks.

RQ 2: Recommending learning path. Based on the HMM-identified student engagement states and their transitions, we further explored the use of these states and transitions to suggest important actions to be performed in future weeks in order for students to transition/persist in the engaging state. Specifically, at a given week w , based on a student’s performed actions μ_w at w (e.g., accessing lecture slides from $w - 1$), we identified a student’s current engagement state S_w by finding the state with the highest

probability of emitting μ . Next, we utilised the state transition probability matrix T to calculate a favorable state at next week S_{w+1} , where S_{w+1} was the top-3 most likely state to be transitioned from S_w and had the highest probability to persist in engaging states (determined via a descriptive analysis shown in Sect. 4) in the following weeks. To illustrate this, we present an example in Fig. 1: suppose **Learner** started a 5-week course. After finishing week 1, based on their interactions with course material, we identify their current engagement state ① (i.e., the state with the highest probability of emitting μ). The aim is to decide which state should **Learner** proceed to next in order to maximize their chance of staying engaged with the course. Therefore, we identify the top-3 most probable transitions from ① based on T : ⑤ (which is on Path **a**), ③ (Path **b**) and ② (Path **c**) respectively. Since ⑤ is not an engaging state, we move on to Path **b** (on the left side of Fig. 1). To calculate the total probability of staying engaged on Path **b** (denote as E_b), we multiply all transition probabilities between engaging state pairs step-wise on the path, e.g., the first step being ③ \rightarrow ① is determined by $T(\textcircled{3}, \textcircled{1})$. Then $E_b = T(\textcircled{3}, \textcircled{1}) \times T(\textcircled{1}, \textcircled{2}) \times T(\textcircled{2}, \textcircled{3})$, similarly, the engaging probability of $E_c = T(\textcircled{2}, \textcircled{3}) \times T(\textcircled{3}, \textcircled{1}) \times T(\textcircled{1}, \textcircled{2})$ (Path **c** shown on the right side of Fig. 1). Lastly, the higher probability between the E_b and E_c is selected as the engaging path, and learning activities are recommended based on activities with the highest emission probability from the state of first step e.g., if E_b has a higher engaging probability, then activities with the highest emission probability of state ③ are selected from the activity vector as instruction to students. Since the student may still not behave as instructed, we recalculate the path at each week to ensure the engaging path is up-to-date with the student current engagement state.

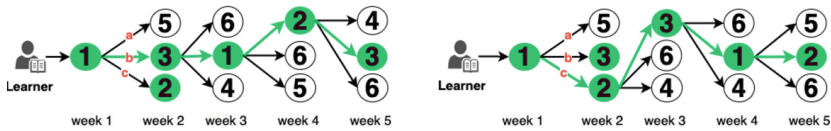


Fig. 1. An example illustration of a recommended learning path.

Evaluation Approach. We assessed the effectiveness of HMM-recommended next-step learning actions from the perspective of student dropout [3, 9], i.e., by taking the learning activity in the recommended path, how likely would a student dropout in the following week. In line with previous studies [3, 7], the *dropout* was a binary label and is set to 1 if a student takes no further actions in the following weeks, and set to 0 otherwise [3, 17]. Given that the latest Deep Learning (DL) dropout predictor can achieve a high prediction accuracy (around 90% AUC [17]), we utilise the predictive power by building a LSTM model – one of the representative DL approaches widely adopted to predict dropout in MOOCs [16, 17, 25]. We input students’ temporal activity representation (as detailed in Sect. 3.2) and their dropout labels to the LSTM model, which output whether a student would dropout on a weekly basis. Importantly, for each week’s prediction, we changed the last week’s activity to the HMM-recommended activity (i.e., assuming the student followed the recommendation) to assess the impact on a dropout prediction. For instance, suppose we were assessing the impact of HMM-recommended activity in week 3, the input to the LSTM model consisted of $[\mu_1, \mu_2, \mu_3]$, we replaced

μ_3 with HMM-recommended activity resulting in $[\mu_1, \mu_2, \mu'_3]$, and performed subsequent dropout prediction. Although we could simply compare the predicted dropout label between the original $[\mu_1, \mu_2, \mu_3]$ and HMM-suggested $[\mu_1, \mu_2, \mu'_3]$, to understand specific impact of μ'_3 , we adopted a widely-used black-box explainer model LIME [19] to examine the importance of μ'_3 towards a dropout/non-dropout prediction. Given that LIME is more efficient than other black-box explainers (e.g., SHAP [23]), LIME can be used to explain a large number of student instances in MOOCs. LIME produce a feature importance score in the interval $[-100, 100]$, where 100 indicates a high importance to a dropout prediction, and -100 indicates high importance to a non-dropout prediction. To demonstrate the effectiveness of our approach, we included in total three baseline approaches, inspired from previous literature:

- Random samples from the learning activities of those who successfully completed the course randomly as the recommended learning action to a learner.
- Diagnose Past Performance (DPP) [16] utilised a XGBoost model to recommend missing learning activities from previous weeks that has the highest contribution (measured in SHAPLEY value) to a dropout label.
- Future Performance Prediction (FPP) [25] utilised a LSTM model to predict the optimal learning activities to be performed, which would result in the lowest dropout probability based on a learner’s past activity.

3.3 Study Setup

Pre-processing. Given the goal of our study was to analyse student engagement with the MOOCs, we restricted our study to students who accessed the course material at least once, i.e., the *At least once* column in Table 1. In line with [15], the training and testing results of HMM and LSTM models are reported based on the 5-fold cross-validation. For all experiments, HMM and LSTM were trained using the same training set, and we ensured that the testing set was not used in the training phase. Given the high class imbalance ratio of dropout vs. non-dropout labels inherently in a MOOC course, which may cause the LSTM to over-classify the majority class (i.e., dropouts), we applied random under-sampling to ensure the training set of the LSTM model has an equal distribution of dropout and non-dropout samples.

Model Implementation. We implement HMM by Python package `bayesian-hmm`¹. To train HMM, we used the Markov Chain Monte Carlo (MCMC) sampling estimation with an iteration of 2000 steps. The LSTM model was implemented using the Python package `keras`. In line with the prior dropout prediction literature [17], the model was composed by two LSTM layers of size 32 and 64 and a Dense layer (with sigmoid activation) with a hidden size of 1. During training, we used a batch size of 32 with an Adam optimiser and set the learning rate and training epoch parameters to 0.001 and 20, respectively. Lastly, LIME was implemented using the Python `LIME` package². In response to the call for increased reproducibility in educational research [7], which benefit practitioners, we open-source code used in this study³

¹ <https://pypi.org/project/bayesian-hmm/>.

² <https://github.com/marcotcr/lime>.

³ <https://github.com/lsha49/hmm-engagement>.

Evaluation Metrics. Given that the feature importance of HMM-suggested activity (to a non-dropout label) depends on the accuracy performance of the LSTM dropout prediction, we adopted two commonly adopted accuracy metrics that are robust against class imbalance: AUC and F1 scores.

4 Results

To investigate student engagement, we generated plots illustrating their parameters: state and state transition. For the sake of brevity, these plots are only displayed for the larger course under study (Big Data in Education). Although we observed similar results across both MOOCs datasets.

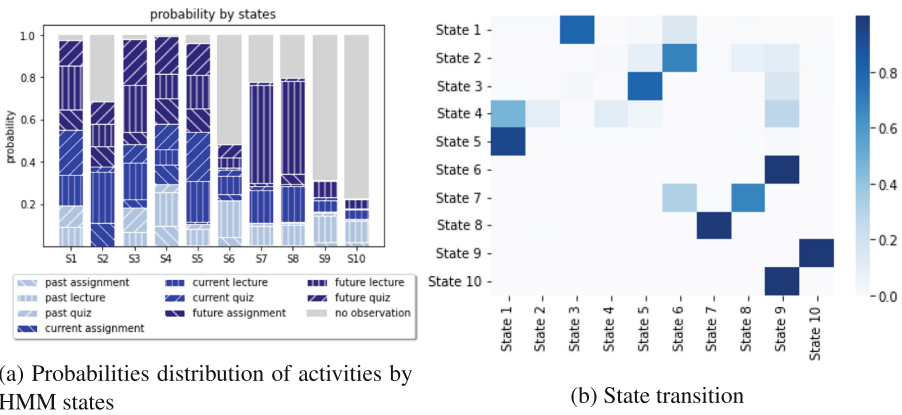


Fig. 2. The HMM states and transition illustration of Big data in education.

RQ1: Understanding student engagement. To answer our first research question, we generated plots illustrating the hidden states (Fig. 2a) and transitions between different states (Fig. 2b) in Fig. 2. Among all states, students belonged to *state 1–5* generally tended to interact with the current week (denoted in blue) and the next week’s (in dark blue) learning materials. As such, these states could also be given semantic labels in keeping with those found in the literature – “on-track” [12]. By contrast, *state 6–10* were characterised mostly by inactivity. For instance, *state 9–10* were mostly non-active (i.e., no observation as denoted in grey), while *state 7–8* were mostly only viewing lectures, thus may be categorised as “disengaged”. Lastly, we noted that *state 6* was more active than *state 9–10*, and performed all the action types (as opposed to just viewing lectures). However, students in *state 6* tended to focus on past activities (denoted in light blue), compared to activities in the current or future weeks, so we assign it a “behind” label. It is important to note that, while some states appeared to have similar observable patterns, e.g., between the on-track *state 4* and *state 5* or between the disengaged *state 7* and *state 8*, their transitions may have been different, as detailed in Fig. 2b. For instance,

while *state 5* almost exclusively transitioned to another on-track *state 1*, *state 4* may have transitioned to a quite different on-track *state 2* and a disengaged *state 9*, which highlighted the importance of closely monitoring the state transitions beyond the static patterns. Interestingly, students who were in a disengaged or behind state almost exclusively kept transitioning to disengaged states. This indicated that, once students started to disengage from the course material, they were unlikely to revert back to an engaging state and complete the course, which corroborated with previous findings [2, 22]. In comparison, though the “on-track” *state 1–5* in general, tended to transition to another “on-track” state, they may have also transitioned to a disengaged state – especially *state 2* and *state 4* (except for *state 5*), indicating the importance of persisting students in an engaging state despite that they are already engaging.

Table 2. The LIME feature importance score (denoted as L_{score}) and transition probability (denoted as P_{path}) of the HMM-suggested path. The results in row `Actual` were calculated by using students’ original activity. The signs \uparrow and \downarrow are used to indicate whether a higher (\uparrow) or lower (\downarrow) value was more preferred in a metric.

Dataset	Path	week 2		week 3		week 4		week 5		week 6		week 7	
		$\downarrow L_{score}$	$\uparrow P_{path}$	$\downarrow L_{score}$	$\uparrow P_{path}$	$\downarrow L_{score}$	$\uparrow P_{path}$	$\downarrow L_{score}$	$\uparrow P_{path}$	$\downarrow L_{score}$	$\uparrow P_{path}$	$\downarrow L_{score}$	$\uparrow P_{path}$
Big data	Actual	0.04	0.72%	0.11	0.82%	0.12	1.24%	-0.03	1.33%	-0.01	0.61%	0.01	1.18%
	Random	-0.93	0.01%	-0.54	0.75%	-2.76	1.69%	-5.76	1.56%	-0.03	1.23%	0.02	0.65%
	DPP	1.02	0.02%	-1.31	0.55%	-2.24	0.46%	1.35	0.28%	0.01	0.42%	-0.05	0.39%
	FPP	5.89	0.06%	-6.51	0.57%	0.74	0.47%	-0.12	0.28%	-0.31	0.45%	0.14	0.39%
	HMM	-1.25	1.18%	-1.6	0.76%	-4.59	2.22%	-6.40	2.40%	-0.07	1.88%	0.02	0.65%
Code yourself	Actual	-1.61	9.79%	-2.30	2.29%	6.03	3.82%	-	-	-	-	-	-
	Random	4.66	4.51%	-4.43	0.91%	-2.36	0.81%	-	-	-	-	-	-
	DPP	-5.43	10.17%	3.64	2.29%	3.65	3.82%	-	-	-	-	-	-
	FPP	-10.28	2.68%	-15.50	2.30%	-2.22	3.82%	-	-	-	-	-	-
	HMM	2.85	6.25%	-3.89	2.71%	-2.64	2.29%	-	-	-	-	-	-

RQ2: Promoting students onto engaging paths. Given the above findings, we further explored whether the proposed HMM-based approach could be used to promote student engagement and preempt dropout. To this end, we utilised HMM to generate engaging paths and test the feature importance score of the path via a dropout predictor (as detailed in Sect. 3.2). Given that the validity of this approach relied on using an accurate dropout predictor, we first evaluated the accuracy of the LSTM model (measured in F1 score and AUC). Overall, we observed that the model achieved a high F1 (0.85–0.91) and AUC (0.82–0.92) in weeks 3–7 for Big data and a high F1 and AUC (0.80) in week 3 for Code yourself, and thus could serve as a strong basis to evaluate the generated path. (We have included a complete report of F1 and AUC results in digital appendix⁴) Given that the rest of the weeks achieved lower F1 and AUC (below 0.80), we restricted our analysis of the feature importance score in weeks 3–7 for Big data and week 3 for Code yourself. We summarise the LIME importance score towards a dropout/non-dropout prediction in Table 2. To take into account the scenario where the model recommended a path consisting of highly engaging learning activities, but were difficult to follow (e.g., recommending students to perform all activities), we included a

⁴ https://docs.google.com/document/d/1hntPaZ1NfSJF4t2I9x83CKtAt4C0jRB9Ma8OfdYj2_Q/edit?usp=sharing.

measure of the probability of students performing the recommended activity on a path (i.e., P_{path}), based on the HMM transition matrix (detailed in Sect. 3.2). Overall, we observed that, for both datasets, FPP achieved the best performance in terms of LIME score (L_{score}) that contributed to a non-dropout (3 out of 6 in weeks 3–7 for Big data and week 3 for Code yourself), closely followed by the proposed HMM approach (2 out of 6). Though, HMM achieved the highest path transition probability P_{path} (4 out of 6), even higher than the Actual activity (i.e., student’s original action), while none of the other baseline approaches managed to surpass Actual. In particular, while FPP generated activity performed the best in terms of LIME score, it had the lowest or the second lowest path probability in 5 out of 6 instances (i.e., week 3–7 in Big data). This indicated that, FPP tended to recommend highly engaging learning activities that were unlikely to be performed by the learners. In comparison, HMM-recommended activities had a more balanced performance between the LIME score (top-2) and transition probability (top-1), indicating that the recommended engaging paths were more likely to be adopted by learners (compared to other path-generating strategies) and as such are more likely to preempt students’ dropout.

5 Discussion and Conclusion

This paper investigated the effectiveness of a HMM-based approach to analyse student temporal engagement patterns. Compared to previous approaches, HMM was shown to be able to not only capture students’ static engagement patterns (represented as “state”), but also the transitions between different states. So, we were able to uncover insights about the temporal dynamics about student engagement. In particular, we found that while disengaged students were unlikely to transition to an engaging state, the engaging students may either stayed engaged or became disengaged in future weeks. This highlighted the importance of persisting students’ engagement state in order to preempt their dropout. Driven by this, we further explored the use of the student engagement state and their transitions to inform a learning path that can promote students’ engagement. Through extensive experimentation, we demonstrated that, not only our HMM-suggested learning activities were effective at preempting student dropout, but these activities were also more likely to be performed by students given their current engagement state (compared to other baseline approaches). These results provide a strong motivation for adoption of the proposed HMM approach in a real-world MOOC setting.

Implications. Firstly, our evaluations demonstrated the effectiveness of HMM as a fully transparent and interpretable approach for modeling student-temporal interactions and identifying hidden patterns and their transitions. All this provides useful insights about students’ current engagement states (e.g., whether they are on-track for completion) and future engagement trends (e.g., whether they are likely to stay engaged in the future with the course material), making it a valuable tool for educators and researchers potentially beyond a MOOC setting (e.g., to model student interactions in a hybrid learning environment). Secondly, given that students who have already been disengaged are almost never going to transition back to an engaging state in a MOOC setting, we posit that providing guidance/intervention at the time of dropout or immediately before dropout

(by utilising predictive models) may be futile and laborious. Instead, educators and practitioners may repurpose such efforts onto persisting the engagement of those “on-track” students and preempt their dropout. Thirdly, we showed that HMM-generated learning path is the only approach that can achieve simultaneously a high contribution to non-dropout prediction and a high probability of being performed by a student given their current engagement state. This implies that students may be inclined to follow the suggested activity and stay engaged with minimal human intervention. Given the low teacher-student ratio in MOOCs settings [10], the HMM-based approach may be implemented to persist student engagement and support teaching at scale.

Limitations. We acknowledge the following limitations of our study. First, the analysis involved only two MOOCs datasets, which enabled us to model student engagement using three different student interaction types (i.e., assignment, lecture, and quiz). To further increase the generalizability of our findings, we plan on adopting additional datasets, potentially beyond MOOC settings, and with more interactions (e.g., forum activity) so as to further explore HMM’s capability in modeling student engagement. Second, we have not conducted a human evaluation on the real-world impact of HMM-suggested engagement paths. Given the promising findings, for future work, we plan on implementing the proposed approach in a real-world MOOC setting and evaluate the effectiveness of preempting student dropouts.

References

1. Coleman, C.A., Seaton, D.T., Chuang, I.: Probabilistic use cases: discovering behavioral patterns for predicting certification. In: Proceedings of the Second ACM Conference on Learning@ Scale, pp. 141–148 (2015)
2. Deng, R., Benckendorff, P., Gannaway, D.: Learner engagement in MOOCs: scale development and validation. *Br. J. Edu. Technol.* **51**(1), 245–262 (2020)
3. Fei, M., Yeung, D.Y.: Temporal models for predicting student dropout in massive open online courses. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 256–263. IEEE (2015)
4. Fincham, E., Gašević, D., Jovanović, J., Pardo, A.: From study tactics to learning strategies: an analytical method for extracting interpretable representations. *IEEE Trans. Learn. Technol.* **12**(1), 59–72 (2018)
5. Fox, E.B., Sudderth, E.B., Jordan, M.I., Willsky, A.S.: An HDP-HMM for systems with state persistence. In: Proceedings of the 25th International Conference on Machine Learning, pp. 312–319 (2008)
6. Fox, E.B., Sudderth, E.B., Jordan, M.I., Willsky, A.S.: A sticky HDP-HMM with application to speaker diarization. *Ann. Appl. Stat.* 1020–1056 (2011)
7. Gardner, J., Yang, Y., Baker, R.S., Brooks, C.: Modeling and experimental design for MOOC dropout prediction: a replication perspective. *Int. Educ. Data Min. Soc.* (2019)
8. Hershcovits, H., Vilenchik, D., Gal, K.: Modeling engagement in self-directed learning systems using principal component analysis. *IEEE Trans. Learn. Technol.* **13**(1), 164–171 (2019)
9. Jiang, S., Williams, A., Schenke, K., Warschauer, M., O’ Dowd, D.: Predicting MOOC performance with week 1 behavior. In: Educational Data Mining 2014 (2014)
10. Joksimović, S., et al.: How do we model learning at scale? A systematic review of research on MOOCs. *Rev. Educ. Res.* **88**(1), 43–86 (2018)

11. Jovanović, J., Gašević, D., Dawson, S., Pardo, A., Mirriahi, N., et al.: Learning analytics to unveil learning strategies in a flipped classroom. *Internet High. Educ.* **33**(4), 74–85 (2017)
12. Kizilcec, R.F., Piech, C., Schneider, E.: Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 170–179 (2013)
13. Lin, J., Lang, D., Xie, H., Gašević, D., Chen, G.: Investigating the role of politeness in human-human online tutoring. In: Bittencourt, I.I., Cukurova, M., Muldner, K., Luckin, R., Millán, E. (eds.) *AIED 2020. LNCS (LNAI)*, vol. 12164, pp. 174–179. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52240-7_32
14. Lin, J., et al.: Is it a good move? Mining effective tutoring strategies from human-human tutorial dialogues. *Futur. Gener. Comput. Syst.* **127**, 194–207 (2022)
15. Mogavi, R.H., Ma, X., Hui, P.: Characterizing student engagement moods for dropout prediction in question pool websites. arXiv preprint: [arXiv:2102.00423](https://arxiv.org/abs/2102.00423) (2021)
16. Mu, T., Jetten, A., Brunskill, E.: Towards suggesting actionable interventions for wheel-spinning students. *Int. Educ. Data Min. Soc.* (2020)
17. Prenkaj, B., Velardi, P., Stilo, G., Distanto, D., Faralli, S.: A survey of machine learning approaches for student dropout prediction in online courses. *ACM Comput. Surv. (CSUR)* **53**(3), 1–34 (2020)
18. Raković, M., et al.: Towards the automated evaluation of legal casenote essays. In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds.) *AIED 2022. LNCS*, vol. 13355, pp. 167–179. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-11644-5_14
19. Ribeiro, M.T., Singh, S., Guestrin, C.: “Why should I trust you?” explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144 (2016)
20. Saqr, M., López-Pernas, S., Jovanović, J., Gašević, D.: Intense, turbulent, or wallowing in the mire: a longitudinal study of cross-course online tactics, strategies, and trajectories. *Internet High. Educ.* 100902 (2022)
21. Sha, L., et al.: Assessing algorithmic fairness in automatic classifiers of educational forum posts. In: Roll, I., McNamara, D., Sosnovsky, S., Luckin, R., Dimitrova, V. (eds.) *AIED 2021. LNCS (LNAI)*, vol. 12748, pp. 381–394. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78292-4_31
22. Sinclair, J., Kalvala, S.: Student engagement in massive open online courses. *Int. J. Learn. Technol.* **11**(3), 218–237 (2016)
23. Štrumbelj, E., Kononenko, I.: Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **41**(3), 647–665 (2014)
24. Teh, Y., Jordan, M., Beal, M., Blei, D.: Sharing clusters among related groups: hierarchical Dirichlet processes. In: *NeurIPS*, vol. 17 (2004)
25. Zhou, Y., Huang, C., Hu, Q., Zhu, J., Tang, Y.: Personalized learning full-path recommendation model based on LSTM neural networks. *Inf. Sci.* **444**, 135–152 (2018)