Behavior-Based Latent Variable Model for Learner Engagement

Andrew S. Lan[1], Christopher G. Brinton[2], Tsung-Yen Yang[3], Mung Chiang[1]
[1]Princeton University, [2]Zoomi Inc., [3]National Chiao Tung University
andrew.lan@princeton.edu, christopher.brinton@zoomiinc.com, tsungyenyang.eecs02@nctu.edu.tw, chiangm@princeton.edu

ABSTRACT

We propose a new model for learning that relates video-watching behavior and engagement to quiz performance. In our model, a learner's knowledge gain from watching a lecture video is treated as proportional to their latent engagement level, and the learner's engagement is in turn dictated by a set of behavioral features we propose that quantify the learner's interaction with the lecture video. A learner's latent concept knowledge is assumed to dictate their observed performance on in-video quiz questions. One of the advantages of our method for determining engagement is that it can be done entirely within standard online learning platforms, serving as a more universal and less invasive alternative to existing measures of engagement that require the use of external devices. We evaluate our method on a real-world massive open online course (MOOC) dataset, from which we find that it achieves high quality in terms of predicting unobserved first-attempt quiz responses, outperforming two state-of-the-art baseline algorithms on all metrics and dataset partitions tested. We also find that our model enables the identification of key behavioral features (e.g., larger numbers of pauses and rewinds, and smaller numbers of fast forwards) that are correlated with higher learner engagement.

Keywords

Behavioral data, engagement, latent variable model, learning analytics, MOOC, performance prediction

1. INTRODUCTION

The recent and rapid development of online learning platforms, coupled with advancements in machine learning, has created an opportunity to revamp the traditional "one-size-fits-all" approach to education. This opportunity is facilitated by the ability of many learning platforms, such as massive open online course (MOOC) platforms, to collect several different types of data on learners, including their assessment responses as well as their learning behavior [9]. The focus of this work is on using different forms of data to model the learning process, which can lead to effective learning analytics and potentially improve learning efficacy.

1.1 Behavior-based learning analytics

Current approaches to learning analytics are focused mainly on providing feedback to learners about their knowledge states – or the level to which they have mastered given concepts/topics/knowledge components – through analysis of their responses to assessment questions [10, 24]. There are other cognitive (e.g., engagement [17, 31], confusion [37], and emotion [11]) as well as non-cognitive (e.g., fatigue, motivation, and level of financial support [14]) factors beyond assessment performance that are crucial to the learning process as well. Accounting for them thus has the potential to yield more effective learning analytics and feedback.

To date, it has been difficult to measure these factors of the learning process. Contemporary online learning platforms, however, have the capability to collect *behavioral data* that can provide some indicators of them. This data commonly includes learners' usage patterns of different types of learning resources [12, 15], their interactions with others via social learning networks [7, 28], their clickstream and keystroke activity logs [2, 8, 30], and sometimes other metadata including facial expressions [35] and gaze location [6].

Recent research has attempted to use behavioral data to augment learning analytics. [5] proposed a latent response model to classify whether a learner is gaming an intelligent tutoring system, for example. Several of these works have sought to demonstrate the relationship between behavior and performance of learners in different scenarios. In the context of MOOCs, [22] concluded that working on more assignments lead to better knowledge transfer than only watching videos, [12] extracted probabilistic use cases of different types of learning resources and showed they are predictive of certification, [32] used discussion forum activity and topic analysis to predict test performance, and [26] discovered that submission activities can be used to predict final exam scores. In other educational domains, [2] discovered that learner keystroke activity in essay-writing sessions is indicative of essay quality, [29] identified behavior as one of the factors predicting math test achievement, and [25] found that behavior is predictive of whether learners can provide elegant solutions to mathematical questions.

In this work, we are interested in how behavioral data can be used to model a learner's *engagement*.

1.2 Learner engagement

Monitoring and fostering engagement is crucial to education, yet defining it concretely remains elusive. Research has sought to identify factors in online learning that may drive engagement; for example, [17] showed that certain production styles of lecture videos promote it. [20] defined disengagement as dropping out in the middle of a video and studied the relationship between disengagement and video content, while [31] considered the relationship between engagement and the

semantic features of mathematical questions that learners respond to. [33] studied the relationship between learners' self-reported engagement levels in a learning session and their facial expressions immediately following in-session quizzes, and [34] considered how engagement is related to linguistic features of discussion forum posts.

There are many types of engagement [3], with the type of interest depending on the specific learning scenario. Several approaches have been proposed for measuring and quantifying different types. These approaches can be roughly divided into two categories: device-based and activity-based. Device-based approaches measure learner engagement using devices external to the learning platform, such as cameras to record facial expressions [35], eye-tracking devices to detect mind wandering while reading text documents [6], and pupil dilation measurements, which are claimed to be highly correlated with engagement [16]. Activity-based approaches, on the other hand, measure engagement using heuristic features constructed from learners' activity logs; prior work includes using replies/upvote counts and topic analysis of discussions [28], and manually defining different engagement levels based on activity types found in MOOCs [4, 21].

Both of these types have their drawbacks. Device-based approaches are far from universal in standard learning platforms because they require integration with external devices. They are also naturally invasive and carry potential privacy risks. Activity-based approaches, on the other hand, are not built on the same granularity of data, and tend to be defined from heuristics that have no guarantee of correlating with learning outcomes. It is therefore desirable to develop a statistically principled, activity-based approach to inferring a learner's engagement.

1.3 Our approach and contributions

In this paper, we propose a probabilistic model for inferring a learner's engagement level by treating it as a latent variable that drives the learner's performance and is in turn driven by the learner's behavior. We apply our framework to a real-world MOOC dataset consisting of clickstream actions generated as learners watch lecture videos, and question responses from learners answering in-video quiz questions.

We first formalize a method for quantifying a learner's behavior while watching a video as a set of nine *behavioral features* that summarize the clickstream data generated (Section 2). These features are intuitive quantities such as the fraction of video played, the number of pauses made, and the average playback rate, some of which have been associated with performance previously [8]. Then, we present our statistical model of learning (Section 3) as two main components: a *learning model* and a *response model*. The learning model treats a learner's gain in concept knowledge as proportional to their latent engagement level while watching a lecture video. Concept knowledge is treated as multidimensional, on a set of latent concepts underlying the course, and videos are associated with varying levels to different concepts. The response model treats a learner's performance on in-video quiz questions, in turn, as proportional to their knowledge on the concepts that this particular question relates to.

By defining engagement to correlate directly with performance, we are able to learn which behavioral features lead to high engagement through a single model. This differs from prior works that first define heuristic notions of engagement and subsequently correlate engagement with performance, in separate procedures. Moreover, our formulation of latent engagement can be made from entirely within standard learning platforms, serving as a more universally applicable and less invasive alternative to device-based approaches.

Finally, we evaluate two different aspects of our model (Section 4): its ability to predict unobserved, first-attempt quiz question responses, and its ability to provide meaningful analytics on engagement. We find that our model predicts with high quality, achieving AUCs of up to 0.76, and outperforming two state-of-the-art baselines on all metrics and dataset partitions tested. One of the partitions tested corresponds to the beginning of the course, underscoring the ability of our model to provide early detection of struggling or advanced students. In terms of analytics, we find that our model enables us to identify behavioral features (e.g., large numbers of pauses and rewinds, and small numbers of fast forwards) that indicate high learner engagement, and to track learners' engagement patterns throughout the course. More generally, these findings can enable an online learning platform to detect learner disengagement and perform appropriate interventions in a fully automated manner.

2. BEHAVIORAL DATA

In this section, we start by detailing the setup of lecture videos and quizzes in MOOCs. We then specify video-watching clickstream data and our method for summarizing it into behavioral features.

2.1 Course setup and data capture

We are interested in modeling learner engagement while watching lecture videos to predict their performance on in-video quiz questions. For this purpose, we can view an instructor's course delivery as the sequence of videos that learners will watch interspersed with the quiz questions they will answer. Let $Q = (q_1, q_2, \ldots)$ be the sequence of questions asked through the course. A video could have any number of questions generally, including none; to enforce a 1:1 correspondence between video content and questions, we will consider the "video" for question q_n to be all video content that appears between q_{n-1} and q_n. Based on this, we will explain the formats of video-watching and quiz response data we work with in this section.

Our dataset. The dataset we will use is from the fall 2012 offering of the course *Networks: Friends, Money, and Bytes* (FMB) on Coursera [1]. This course has 92 videos distributed among 20 lectures, and exactly one question per video.

2.1.1 Video-watching clickstreams

When a learner watches a video on a MOOC, their behavior is typically recorded as a sequence of clickstream actions. In particular, each time a learner makes an action – `play`, `pause`, `seek`, `ratechange`, `open`, or `close` – on the video player, a clickstream event is generated. Formally, the ith event created for the course will be in the format

$$E_i =< u_i, v_i, e_i, p_i^{'}, p_i, x_i, s_i, r_i >$$

Here, u_i and v_i are the IDs of the specific learner (user) and video, respectively, and e_i is the type of action that u_i made on v_i. p_i is the position of the video player (in seconds) immediately after e_i is made, p'_i is the position immediately before,[1] x_i is the UNIX timestamp (in seconds) at which e_i was fired, s_i is the binary state of the video player – either `playing` or `paused` – once this action is made, and r_i is the playback rate of the video player once this action is made. Our FMB dataset has 314,632 learner-generated clickstreams from 3,976 learners.[2]

The set $E_{u,v} = \{E_i | u_i = u, v_i = v\}$ of clickstreams for learner u recorded on video v can be used to reconstruct the behavior u exhibits on v. In Section 2.2 we will explain the features computed from $E_{u,v}$ to summarize this behavior.

2.1.2 *Quiz responses*

When a learner submits a response to an in-video quiz question, an event is generated in the format

$$A_m =< u_m, v_m, x_m, a_m, y_m >$$

Again, u_m and v_m are the learner and video IDs (*i.e.,* the quiz corresponding to the video). x_m is the UNIX timestamp of the submission, a_m is the specific response, and y_m is the number of points awarded for the response. The questions in our dataset are multiple choice with a single response, so y_m is binary-valued.

In this work, we are interested in whether quiz responses were correct on first attempt (CFA) or not. As a result, with $A_{u,v} = \{A_m | u_m = u, v_m = v\}$, we consider the event $A'_{u,v}$ in this set with the earliest timestamp $x'_{u,v}$. We also only consider the set of clickstreams $E'_{u,v} \subseteq E_{u,v}$ that occur before $x'_{u,v}$, as the ones after would be anti-causal to CFA.

2.2 Behavioral features and CFA score

With the data $E'_{u,v}$ and $A'_{u,v}$, we construct two sets of information for each learner u on each video v, *i.e.,* each learner-video pair. First is a set of nine behavioral features that summarize u's video-watching behavior on v [8]:

(1) Fraction spent. The fraction of time the learner spent on the video, relative to the playback length of the video. Formally, this quantity is $e_{u,v}/l_v$, where

$$e_{u,v} = \sum_{i \in \mathcal{S}} \min(x_{i+1} - x_i, l_v)$$

is the elapsed time on v obtained by finding the total UNIX time for u on v, and l_v is the length of the video (in seconds). Here, $\mathcal{S} = \{i \in A'_{u,v} : a_{i+1} \neq \texttt{open}\}$. l_v is included as an upper bound for excessively long intervals of time.

(2) Fraction completed. The fraction of the video that the learner completed, between 0 (none) and 1 (all). Formally, it is $c_{u,v}/l_v$, where $c_{u,v}$ is the number of unique 1 second segments of the video that the learner visited.

[1] p_i and p'_i will only differ when i is a skip event.

[2] This number excludes invalid `stall`, `null`, and `error` events, as well as `open` and `close` events which are generated automatically.

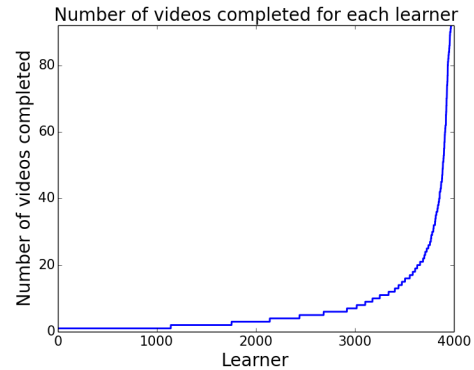


Figure 1: Distribution of the number of videos that each each learner completed in FMB. More than 85% of learners completed less than 20 videos.

(3) Fraction played. The fraction of the video that the learner played relative to the length. Formally, it is calculated as $g_{u,v}/l_v$, where

$$g_{u,v} = \sum_{i \in \mathcal{S}} \min(p'_{i+1} - p_i, l_v)$$

is the total length of video that was played (while in the playing state). Here, $\mathcal{S} = \{i \in A'_{u,v} : a_{i+1} \neq \texttt{open} \wedge s_i = \texttt{playing}\}$.

(4) Fraction paused. The fraction of time the learner stayed paused on the video relative to the length. It is calculated as $h_{u,v}/l_v$, where

$$h_{u,v} = \sum_{i \in \mathcal{S}} \min(t_{i+1} - t_i, l_v)$$

is the total time the learner stayed in the `paused` state on this video. Here, $\mathcal{S} = \{i \in A'_{u,v} : a_{i+1} \neq \texttt{open} \wedge s_i = \texttt{paused}\}$.

(5) Number of pauses. The number of times the learner paused the video, or

$$\sum_{i \in A'_{u,v}} \mathbb{1}\{a_i = \texttt{pause}\}$$

where $\mathbb{1}\{\}$ is the indicator function.

(6) Number of rewinds. The number of times the learner skipped backwards in the video, or

$$\sum_{i \in A'_{u,v}} \mathbb{1}\{a_i = \texttt{skip} \wedge p'_i < p_i\}$$

(7) Number of fast forwards. The number of times the learner skipped forward in the video, *i.e.,* with $p'_i > p_i$ in the previous equation.

(8) Average playback rate. The time-average of the learner's playback rate on the video. Formally, it is calculated as

$$\bar{r}_{u,v} = \frac{\sum_{i \in \mathcal{S}} r_i \cdot \min(x_{i+1} - x_i, l_v)}{\sum_{i \in \mathcal{S}} \min(x_{i+1} - x_i, l_v)}$$

where $\mathcal{S} = \{i \in A'_{u,v} : a_{i+1} \neq \texttt{open} \wedge s_i = \texttt{playing}\}$.

(9) Standard deviation of playback rate. The standard deviation of the learner's playback rate. It is calculated as

$$\sqrt{\frac{\sum_{i\in\mathcal{S}}(r_i-\bar{r}_{u,v})^2\cdot\min(x_{i+1}-x_i,l_v)}{\sum_{i\in\mathcal{S}}\min(x_{i+1}-x_i,l_v)}}$$

with the same $\mathcal{S}$ as the average playback rate.

The second piece of information for each learner-video pair is u's CFA score $y_{u,v}\in\{0,1\}$ on the quiz question for v.

2.3 Dataset subsets

We will consider different groups of learner-video pairs when evaluating our model in Section 4. Our motivation for doing so is the heterogeneity of learner motivation and high dropoff rates in MOOCs [9]: many will quit the course after watching just a few lectures. Modeling in a small subset of data, particularly those at the beginning of the course, is desirable because it can lead to "early detection" of those who may drop out [8].

Figure 1 shows the dropoff for our dataset in terms of the number of videos each learner completed: more than 85% of learners completed just 20% of the course. "Completed" is defined here as having watched some of the video and responded to the corresponding question. Let T_u be the number of videos learner u completed and $\gamma(v)$ be the index of video v in the course, we define $\Omega^{u_0,v_0}=\{(u,v):T_u\geq u_0\ \wedge\ \gamma(v)\leq v_0\}$ to be the subset of learner-video pairs such that u completed at least u_0 videos and v is within the first v_0 videos. The full dataset is $\Omega^{1,92}$, and we will also consider $\Omega^{20,92}$ as the subset of 346 active learners over the full course and $\Omega^{1,20}$ as the subset of all learners over the first two weeks[3] in our evaluation.

3. STATISTICAL MODEL OF LEARNING WITH LATENT ENGAGEMENT

In this section, we propose our statistical model. Let U denote the number of learners (indexed by u) and V the number of videos (indexed by v). Further, we use T_u to denote the number of time instances registered by learner u (indexed by t); we take a time instance to be a learner completing a video, i.e., watching a video and answering the corresponding quiz question. For simplicity, we use a discrete notion of time, i.e., each learner-video pair will correspond to one time instance for one learner.

Our model considers learners' responses to quiz questions as measurements of their underlying knowledge on a set of concepts; let K denote the number of such concepts. Further, our model considers the action of watching lecture videos as part of learning that changes learners' latent knowledge states over time. These different aspects of the model are visualized in Figure 2: there are two main components, a response model and a learning model.

3.1 Response Model

Our statistical model of learner responses is given by

$$p(y_u^{(t)}=1|\mathbf{c}_u^{(t)})=\sigma(\mathbf{w}_{v(u,t)}^T\mathbf{c}_u^{(t)}-\mu_{v(u,t)}+a_u),\qquad(1)$$

[3]In FMB, the first two weeks of lectures is the first 20 videos.

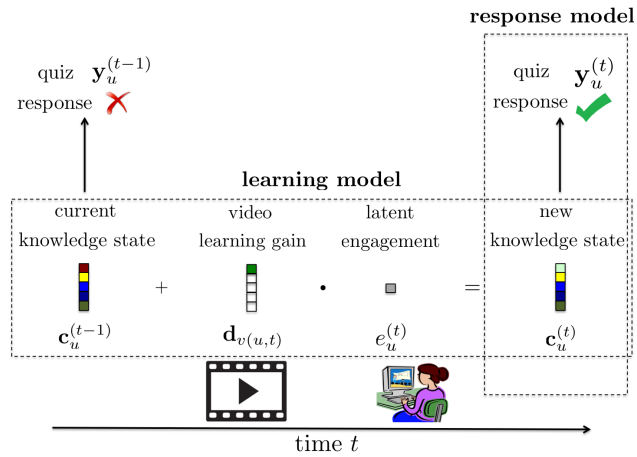


Figure 2: Our proposed statistical model of learning consists of two main parts, a response model and a learning model.

where $v(u,t):\Omega\subseteq\{1,\ldots,U\}\times\{1,\ldots,\max_u T_u\}\rightarrow\{1,\ldots,V\}$ denotes a mapping from a learner index-time index pair to the index of the video v that u was watching at t. $y_u^{(t)}\in\{0,1\}$ is the binary-valued CFA score of learner u on the quiz question corresponding to the video they watch at time t, with 1 denoting a correct response (CFA) and 0 denoting an incorrect response (non-CFA).

The variable $\mathbf{w}_v\in\mathbb{R}_+^K$ denotes the non-negative, K-dimensional quiz question–concept association vector that characterizes how the quiz question corresponding to video v tests learners' knowledge on each concept, and the variable μ_v is a scalar characterizing the intrinsic difficulty of the quiz question. $\mathbf{c}_u^{(t)}$ is the K-dimensional concept knowledge vector of learner u at time t, characterizing the knowledge level of the learner on each concept at the time, and a_u denotes the static, intrinsic ability of learner u. Finally, $\sigma(x)=\frac{1}{1+e^{-x}}$ is the sigmoid function.

We restrict the question–concept association vector $\mathbf{w}_v$ to be non-negative in order to make the parameters interpretable [24]. Under this restriction, the values of concept knowledge vector $\mathbf{c}_u^{(t)}$ can be understood as follows: large, positive values lead to higher chances of answering a question correctly, thus corresponding to high knowledge, while small, negative values lead to lower chances of answering a question correctly, thus corresponding to low knowledge.

3.2 Learning Model

Our model of learning considers transitions in learners' knowledge states as induced by watching lecture videos. It is given by

$$\mathbf{c}_u^{(t)}=\mathbf{c}_u^{(t-1)}+e_u^{(t)}\mathbf{d}_{v(u,t)},\quad t=1,\ldots,T_u,\qquad(2)$$

where the variable $\mathbf{d}_v\in\mathbb{R}_+^K$ denotes the non-negative, K-dimensional learning gain vector for video v; each entry characterizes the degree to which the video improves learners' knowledge level on each concept. The assumption of non-negativity on $\mathbf{d}_v$ implies that videos will not negatively affect learners' knowledge, as in [23]. $\mathbf{c}_u^{(0)}$ is the initial knowledge state of learner u at time $t=0$, i.e., before starting the

	$\Omega^{20,92}$		$\Omega^{1,20}$		$\Omega^{1,92}$	
	ACC	AUC	ACC	AUC	ACC	AUC
Proposed model	**0.7293±0.0070**	**0.7608±0.0094**	**0.7096±0.0057**	**0.7045±0.0066**	**0.7058±0.0054**	**0.7216±0.0054**
SPARFA	0.7209±0.0070	0.7532±0.0098	0.7061±0.0069	0.7020±0.0070	0.6975±0.0048	0.7124±0.0050
BKT	0.7038±0.0084	0.7218±0.0126	0.6825±0.0058	0.6662±0.0065	0.6803±0.0055	0.6830±0.0059

Table 1: Quality comparison of the different algorithms on predicting unobserved quiz question responses. The obtained ACC and AUC metrics on different subsets of the FMB dataset are given. Our proposed model obtains higher quality than the SPARFA and BKT baselines in each case.

course and watching any video.

The scalar latent variable $e_u^{(t)} \in [0,1]$ in (2) characterizes the *engagement level* that learner u exhibits when watching video $v(u,t)$ at time t. This is in turn modeled as

$$e_u^{(t)} = \sigma(\boldsymbol{\beta}^T \mathbf{f}_u^{(t)}), \tag{3}$$

where $\mathbf{f}_u^{(t)}$ is a 9-dimensional vector of the behavioral features defined in Section 2.2, summarizing learner u's behavior while the video at time t. $\boldsymbol{\beta}$ is the unknown, 9-dimensional parameter vector that characterizes how engagement associates with each behavioral feature.

Taken together, (2) and (3) state that the knowledge gain a learner will experience on a particular concept while watching a particular video is given by

(i) the video's intrinsic association with the concept, modulated by

(ii) the learner's engagement while watching the video, as manifested by their clickstream behavior.

From (2), a learner's (latent) engagement level dictates the fraction of the video's available learning gain they acquire to improve their knowledge on each concept. The response model (1) in turn holds that performance is dictated by a learner's concept knowledge states. In this way, engagement is directly correlated with performance through the concept knowledge states. Note that in this paper, we treat the engagement variable $e_u^{(t)}$ as a scalar; the extension of modeling it as a vector and thus separating engagement by concept is part of our ongoing work.

It is worth mentioning the similarity between our characterization of engagement as a latent variable in the learning model and the input gate variables in long-short term memory (LSTM) neural networks [18]. In LSTM, the change in the latent memory state (loosely corresponding to the latent concept knowledge state vector $\mathbf{c}_u^{(t)}$) is given by the input vector (loosely corresponding to the video learning gain vector $\mathbf{d}_v$) modulated by a set of input gate variables (corresponding to the engagement variable $e_u^{(t)}$).

Parameter inference. Our statistical model of learning and response can be seen as a particular type of recurrent neural network (RNN). Therefore, for parameter inference, we implement a stochastic gradient descent algorithm with standard backpropagation. Given the graded learner responses $y_u^{(t)}$ and behavioral features $\mathbf{f}_u^{(t)}$, our parameter inference algorithm estimates the quiz question–concept association vectors $\mathbf{w}_v$, the quiz question intrinsic difficulties μ_v, the the video learning gain vectors $\mathbf{d}_v$, the learner initial knowledge vectors $\mathbf{c}_u^{(0)}$, the learner abilities a_u, and the engagement–behavioral feature association vector $\boldsymbol{\beta}$. We omit the details of the algorithm for simplicity of exposition.

4. EXPERIMENTS

In this section, we evaluate the proposed latent engagement model on the FMB dataset. We first demonstrate the gain in predictive quality of the proposed model over two baseline algorithms (Section 4.1), and then show how our model can be used to study engagement (Section 4.2).

4.1 Predicting unobserved responses

We evaluate our proposed model's quality by testing its ability to predict unobserved quiz question responses.

Baselines. We compare our model against two well-known, state-of-the-art response prediction algorithms that do not use behavioral data. First is the sparse factor analysis (SPARFA) algorithm [24], which factors the learner-question matrix to extract latent concept knowledge, but does not use a time-varying model of learners' knowledge states. Second is a version of the Bayesian knowledge tracing (BKT) algorithm that tracks learners' time-varying knowledge states, which incorporates a set of guessing and slipping probability parameters for each question, a learning probability parameter for each video, and an initial knowledge level parameter for each learner [13, 27].

4.1.1 Experimental setup and metrics

Regularization. In order to prevent overfitting, we add ℓ_2-norm regularization terms to the overall optimization objective function for every set of variables in both the proposed model and in SPARFA. We use a parameter λ to control the amount of regularization on each variable.

Cross validation. We perform 5-fold cross validation on the full dataset ($\Omega^{1,92}$), and on each subset of the dataset introduced in Section 2.3 ($\Omega^{20,92}$ and $\Omega^{1,20}$). To do so, we randomly partition each learner's quiz question responses into 5 data folds. Leaving out one fold as the test set, we use the remaining four folds as training and validation sets to select the values of the tuning parameters for each algorithm, i.e., by training on three of the folds and validating on the other. We then train every algorithm on all four observed folds using the tuned values of the parameters, and evaluate them on the holdout set. All experiments are repeated for 20 random partitions of the training and test sets.

For the proposed model and for SPARFA, we tune both the

Feature	Coefficient
Fraction spent	0.1941
Fraction completed	0.1443
Fraction played	0.2024
Fraction paused	0.0955
Number of pauses	0.2233
Number of rewinds	0.4338
Number of fast forwards	−0.1551
Average playback rate	0.2797
Standard deviation of playback rate	0.0314

Table 2: Regression coefficient vector $\boldsymbol{\beta}$ learned over the full dataset, associating each clickstream feature to engagement. All but one of the features (number of fast forwards) is positively correlated with engagement.

number of concepts $K \in \{2, 4, 6, 8, 10\}$ and the regularization parameter $\lambda \in \{0.5, 1.0, \ldots, 10.0\}$. Note that for the proposed model, when a question response is left out as part of the test set, only the response is left out of the training set: the algorithm still uses the clickstream data for the corresponding learner-video pair to model engagement.

Metrics. To evaluate the quality of the algorithms, we employ two commonly used binary classification metrics: prediction accuracy (ACC) and area under the receiver operating characteristic curve (AUC) [19]. The ACC metric is simply the fraction of predictions that are made correctly, while the AUC measures the tradeoff between the true and false positive rates of the classifier. Both metrics take values in $[0, 1]$, with larger values indicating higher quality.

4.1.2 Results and discussion

Table 1 gives the evaluation results for the three algorithms. The average and standard deviation over the 20 random data partitions are reported for each dataset group and metric.

First of all, the results show that our proposed model consistently achieves higher quality than both baseline algorithms on both metrics. It significantly outperforms BKT in particular (SPARFA also outperforms BKT). This shows the potential of our model to push the envelope on achievable quality in performance prediction research.

Notice that our model achieves its biggest quality improvement on the full dataset, with a 1.3% gain in AUC over SPARFA and a 5.7% gain over BKT. This observation suggests that as more clickstream data is captured and available for modeling – especially as we observe more video-watching behavioral data from learners over a longer period of time (the full dataset $\Omega^{1,92}$ contains clickstream data for up to 12 weeks, while the $\Omega^{1,20}$ subset only contains data for the first 2 weeks) – the proposed model achieves more significant quality enhancements over the baseline algorithms. This is somewhat surprising, since prior work on behavior-based performance prediction [8] has found the largest gains in the presence of fewer learner-video pairs, i.e., before there are many question responses for other algorithms to model on. But our algorithm also benefits from additional question responses, to update its learned relationship between behavior and concept knowledge.

Figure 3: Plot of the latent engagement level $e_j^{(t)}$ over time for one third of the learners in FMB, showing a diverse set of behaviors across learners.

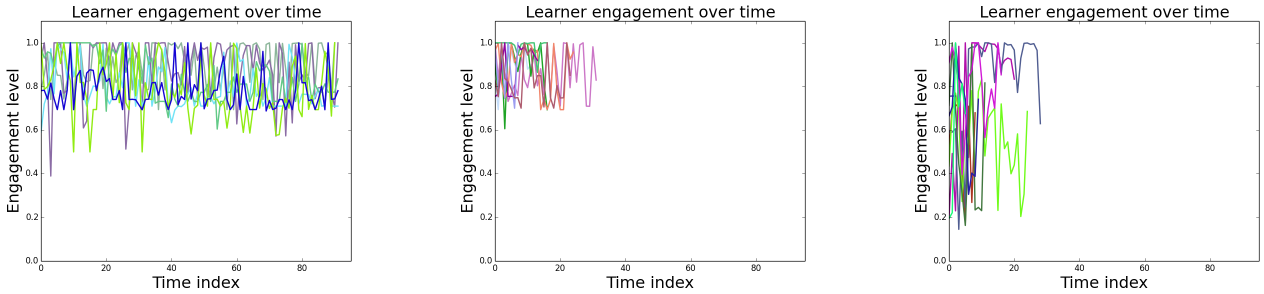
The first two weeks of data ($\Omega^{1,20}$) is sparse in that the majority of learners answer at most a few questions during this time, many of whom will drop out (see Figure 1). In this case, our model obtains a modest improvement over SPARFA, which is static and uses fewer parameters. The gain over BKT is particularly pronounced, at 5.7%. This, combined with the findings for active learners over the full course ($\Omega^{20,92}$), shows that observing video-watching behavior of learners who drop out of the course in its early states (these learners are excluded from $\Omega^{20,92}$) leads to a slight increase in the performance gain of the proposed model over the baseline algorithms. Importantly, this shows that our algorithm provides benefit for *early detection*, with the ability to predict performance of learners who will end up dropping out [8].

4.2 Analyzing engagement

Given predictive quality, one benefit of our model is that it can be used to analyze engagement. The two parameters to consider for this are the regression coefficient vector $\boldsymbol{\beta}$ and the engagement scalar $e_u^{(t)}$ itself.

Behavior and engagement. Table 2 gives each of the estimated feature coefficients in $\boldsymbol{\beta}$ for the full dataset $\Omega^{1,92}$, with regularization parameters chosen via cross validation. All of the features except for the number of fast forwards are positively correlated with the latent engagement level. This is to be expected since many of the features are associated with processing more video content, e.g., spending more time, playing more, or pausing longer to reflect, while fast forwarding involves skipping over the content.

The features that contribute most to high latent engagement levels are the number of pauses, the number of rewinds, and the average playback rate. The first two of these are likely indicators of actual engagement as well, since they indicate whether the learner was thinking while pausing the video or re-visiting earlier content which contains knowledge that they need to recall or revise. The strong, positive correlation of average playback rate is somewhat surprising though: we may expect that a higher playback rate would have a



(a) Learners that consistently exhibit high engagement and finish the course.

(b) Learners that exhibit high engagement but drop out early.

(c) Learners that exhibit inconsistent engagement and drop out.

Figure 4: Plot of the latent engagement level $e_j^{(t)}$ over time for selected learners in three different groups.

negative impact on engagement, like fast forwarding does, as it involves speeding through content. On the other hand, it may be an indication that learners are more focused on the material and trying to keep their interest higher.

Engagement over time. Figure 3 visualizes the evolution of $e_u^{(t)}$ over time for 1/3 of the learners (randomly selected). Patterns in engagement differs substantially across learners; those who finish the course mostly exhibit high engagement levels throughout, while those who drop out early vary greatly in their engagement, some high and others low.

Figure 4 breaks down the learners into three different types according to their engagement patterns, and plots their engagement levels over time separately. The first type of learner (a) finishes the course and consistently exhibits high engagement levels throughout the duration. The second type (b) also consistently exhibits high engagement levels, but drops out of the course after up to three weeks. The third type of learner (c) exhibits inconsistent engagement levels before an early dropout. Equipped with temporal plots like these, an instructor could determine which learners may be in need of intervention, and could design different interventions for different engagement clusters [8, 36].

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a new statistical model for learning, based on learner behavior while watching lecture videos and their performance on in-video quiz questions. Our model has two main parts: (i) a response model, which relates a learner's performance to latent concept knowledge, and (ii) a learning model, which relates the learner's concept knowledge in turn to their latent engagement level while watching videos. Through evaluation on a real-world MOOC dataset, we showed that our model can predict unobserved question responses with superior quality to two state-of-the-art baselines, and also that it can lead to engagement analytics: it identifies key behavioral features driving high engagement, and shows how each learner's engagement evolves over time.

Our proposed model enables the measurement of engagement solely from data that is logged within online learning platforms: clickstream data and quiz responses. In this way, it serves as a less invasive alternative to current approaches for measuring engagement that require external devices, e.g., cameras and eye-trackers [6, 16, 35]. One avenue of future work is to conduct an experiment that will correlate our definition of latent engagement with these methods.

Additionally, one could test other, more sophisticated characterizations of the latent engagement variable. One such approach could seek to characterize engagement as a function of learners' previous knowledge level. An alternative or addition to this would be a generative modeling approach of engagement to enable the prediction of future engagement given each learner's learning history.

One of the long-term, end-all goals of this work is the design of a method for useful, real-time analytics to instructors. The true test of this ability comes from incorporating the method into a learning system, providing its outputs – namely, performance prediction forecasts and engagement evolution – to an instructor through the user interface, and measuring the resulting improvement in learning outcomes.

Acknowledgments

Thanks to Debshila Basu Mallick for discussions on the different types of engagement.

6. REFERENCES

[1] Networks: Friends, Money, and Bytes. `https://www.coursera.org/course/friendsmoneybytes`.

[2] L. Allen, M. Jacovina, M. Dascalu, R. Roscoe, K. Kent, A. Likens, and D. McNamara. {ENTER}ing the time series {SPACE}: Uncovering the writing process through keystroke analyses. In *Proc. Intl. Conf. Educ. Data Min.*, pages 22–29, June 2016.

[3] A. Anderson, S. Christenson, M. Sinclair, and C. Lehr. Check & connect: The importance of relationships for promoting engagement with school. *J. School Psychol.*, 42(2):95–113, Mar. 2004.

[4] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec. Engaging with massive online courses. In *Proc. Intl. Conf. World Wide Web*, pages 687–698, Apr. 2014.

[5] R. Baker, A. Corbett, and K. Koedinger. Detecting student misuse of intelligent tutoring systems. In *Proc. Intl. Conf. Intell. Tutoring Syst.*, pages 531–540, Aug. 2004.

[6] R. Bixler and S. D'Mello. Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Model. User-adapt. Interact.*, 26(1):33–68, Mar. 2016.

[7] C. Brinton, S. Buccapatnam, F. Wong, M. Chiang, and H. Poor. Social learning networks: Efficiency optimization for MOOC forums. In *Proc. IEEE Conf.*

Comput. Commun., pages 1–9, Apr. 2016.

[8] C. Brinton and M. Chiang. MOOC performance prediction via clickstream data and social learning networks. In *Proc. IEEE Conf. Comput. Commun.*, pages 2299–2307, April 2015.

[9] C. Brinton, R. Rill, S. Ha, M. Chiang, R. Smith, and W. Ju. Individualization for education at scale: MIIC design and preliminary evaluation. *IEEE Trans. Learn. Technol.*, 8(1):136–148, Jan. 2015.

[10] H. Cen, K. Koedinger, and B. Junker. Learning factors analysis – A general method for cognitive model evaluation and improvement. In *Proc. Intl. Conf. Intell. Tutoring Syst.*, pages 164–175, June 2006.

[11] L. Chen, X. Li, Z. Xia, Z. Song, L. Morency, and A. Dubrawski. Riding an emotional roller-coaster: A multimodal study of young child's math problem solving activities. In *Proc. Intl. Conf. Educ. Data Min.*, pages 38–45, June 2016.

[12] C. Coleman, D. Seaton, and I. Chuang. Probabilistic use cases: Discovering behavioral patterns for predicting certification. In *Proc. ACM Conf. Learn at Scale*, pages 141–148, Mar. 2015.

[13] A. Corbett and J. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Model. User-adapt. Interact.*, 4(4):253–278, Dec. 1994.

[14] C. Farrington, M. Roderick, E. Allensworth, J. Nagaoka, T. Keyes, D. Johnson, and N. Beechum. *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance–A Critical Literature Review*. Consortium on Chicago School Research, 2012.

[15] B. Gelman, M. Revelle, C. Domeniconi, A. Johri, and K. Veeramachaneni. Acting the same differently: A cross-course comparison of user behavior in MOOCs. In *Proc. Intl. Conf. Educ. Data Min.*, pages 376–381, June 2016.

[16] M. Gilzenrat, J. Cohen, J. Rajkowski, and G. Aston-Jones. Pupil dynamics predict changes in task engagement mediated by locus coeruleus. In *Proc. Soc. Neurosci. Abs.*, page 19, Nov. 2003.

[17] P. Guo, J. Kim, and R. Rubin. How video production affects student engagement: An empirical study of MOOC videos. In *Proc. ACM Conf. Learn at Scale*, pages 41–50, Mar. 2014.

[18] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, Nov. 1997.

[19] H. Jin and C. Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.*, 17(3):299–310, Mar. 2005.

[20] J. Kim, P. Guo, D. Seaton, P. Mitros, K. Gajos, and R. Miller. Understanding in-video dropouts and interaction peaks in online lecture videos. In *Proc. ACM Conf. Learn at Scale*, pages 31–40, Mar. 2014.

[21] R. Kizilcec, C. Piech, and E. Schneider. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proc. Intl. Conf. Learn. Analyt. Knowl.*, pages 170–179, Apr. 2013.

[22] K. Koedinger, J. Kim, J. Jia, E. McLaughlin, and N. Bier. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proc. ACM Conf. Learn at Scale*, pages 111–120, Mar. 2015.

[23] A. Lan, C. Studer, and R. Baraniuk. Time-varying learning and content analytics via sparse factor analysis. In *Proc. ACM SIGKDD Intl. Conf. Knowl. Discov. Data Min.*, pages 452–461, Aug. 2014.

[24] A. Lan, A. Waters, C. Studer, and R. Baraniuk. Sparse factor analysis for learning and content analytics. *J. Mach. Learn. Res.*, 15:1959–2008, June 2014.

[25] L. Malkiewich, R. Baker, V. Shute, S. Kai, and L. Paquette. Classifying behavior to elucidate elegant problem solving in an educational game. In *Proc. Intl. Conf. Educ. Data Min.*, pages 448–453, June 2016.

[26] J. McBroom, B. Jeffries, I. Koprinska, and K. Yacef. Mining behaviours of students in autograding submission system logs. In *Proc. Intl. Conf. Educ. Data Min.*, pages 159–166, June 2016.

[27] Z. Pardos and N. Heffernan. Modeling individualization in a Bayesian networks implementation of knowledge tracing. In *Proc. Intl. Conf. User Model. Adapt. Personalization*, pages 255–266, June 2010.

[28] J. Reich, B. Stewart, K. Mavon, and D. Tingley. The civic mission of MOOCs: Measuring engagement across political differences in forums. In *Proc. ACM Conf. Learn at Scale*, pages 1–10, Apr. 2016.

[29] M. San Pedro, E. Snow, R. Baker, D. McNamara, and N. Heffernan. Exploring dynamical assessments of affect, behavior, and cognition and Math state test achievement. In *Proc. Intl. Conf. Educ. Data Min.*, pages 85–92, June 2015.

[30] C. Shi, S. Fu, Q. Chen, and H. Qu. VisMOOC: Visualizing video clickstream data from massive open online courses. In *IEEE Pacific Visual. Symp.*, pages 159–166, Apr. 2015.

[31] S. Slater, R. Baker, J. Ocumpaugh, P. Inventado, P. Scupelli, and N. Heffernan. Semantic features of Math problems: Relationships to student learning and engagement. In *Proc. Intl. Conf. Educ. Data Min.*, pages 223–230, June 2016.

[32] S. Tomkins, A. Ramesh, and L. Getoor. Predicting post-test performance from online student behavior: A high school MOOC case study. In *Proc. Intl. Conf. Educ. Data Min.*, pages 239–246, June 2016.

[33] A. Vail, J. Wiggins, J. Grafsgaard, K. Boyer, E. Wiebe, and J. Lester. The affective impact of tutor questions: Predicting frustration and engagement. In *Proc. Intl. Conf. Educ. Data Min.*, pages 247–254, June 2016.

[34] X. Wang, D. Yang, M. Wen, K. Koedinger, and C. Rosé. Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In *Proc. Intl. Conf. Educ. Data Min.*, pages 226–233, June 2015.

[35] J. Whitehill, Z. Serpell, Y. Lin, A. Foster, and J. Movellan. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Trans. Affect. Comput.*, 5(1):86–98, Jan. 2014.

[36] J. Whitehill, J. Williams, G. Lopez, C. Coleman, and J. Reich. Beyond prediction: Towards automatic intervention in MOOC student stop-out. In *Proc. Intl. Conf. Educ. Data Min.*, pages 171–178, June 2015.

[37] D. Yang, R. Kraut, and C. Rosé. Exploring the effect of student confusion in massive open online courses. *J. Educ. Data Min.*, 8(1):52–83, 2016.