

Time-series Insights into the Process of Passing or Failing Online University Courses using Neural-Induced Interpretable Student States

Byungsoo Jeon
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA
byungsoj@cs.cmu.edu

Eyal Shafran
Western Governors University
4001 S 700 East
Salt Lake City, UT
eyal.shafran@wgu.edu

Luke Breitfeller
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA
mbreitfe@cs.cmu.edu

Jason Levin
Western Governors University
4001 S 700 East
Salt Lake City, UT
jason.levin@wgu.edu

Carolyn P. Rosé
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA
cprose@cs.cmu.edu

ABSTRACT

This paper addresses a key challenge in Educational Data Mining, namely to model student behavioral trajectories in order to provide a means for identifying students most at-risk, with the goal of providing supportive interventions. While many forms of data including clickstream data or data from sensors have been used extensively in time series models for such purposes, in this paper we explore the use of textual data, which is sometimes available in the records of students at large, online universities. We propose a time series model that constructs an evolving student state representation using both clickstream data and a signal extracted from the textual notes recorded by human mentors assigned to each student. We explore how the addition of this textual data improves both the predictive power of student states for the purpose of identifying students at risk for course failure as well as for providing interpretable insights about student course engagement processes.

Keywords

Student State, Clickstream Data, Mentor's Notes, LDA, Time-series Modeling, Deep Learning

1. INTRODUCTION

In online universities, modeling the population of students at scale is an important challenge, for example, in order to identify students most at-risk and to provide appropriate interventions to improve their chances of earning a degree in a timely fashion. In this respect, a plethora of approaches for clickstream analysis [11, 21, 22] have been published in

the field of Educational Data Mining, which address questions about modeling student course engagement processes [19]. While clickstream data is the most readily available, and while some success has been achieved using it for this purpose, its low level indicators provide only glimpses related to student progress, challenges, and affect as we would hope to observe and model them. In this paper, we explore the extent to which we may achieve richer insights by adding textual data to the foundation provided by clickstream data.

One advantage to modeling student behavior and states from a for-pay platform is that the level of support provided to students is greater than in freely available contexts like Massive Open Online Courses (MOOCs), and this more intensive engagement provides richer data sources that can be leveraged. In our work, we make use of a new data source provided by the Western Governor's University (WGU) platform, where each student is assigned a human mentor, and the notes from each biweekly encounter between student and mentor are recorded and made part of the time series data available for each student. Thus, even if we do not have access to the full transcript of the interactions between students and their mentors, we can leverage the documentation of provided support in order to enhance the richness and ultimately the interpretability of student states we may induce from other low level behavioral indicators we can extract from traces of learning platform interactions.

A major thrust of our work has been to develop a technique for leveraging this form of available textual data. We refer to this data as *Mentor's Notes*. In particular, we propose a sequence model to integrate available data traces over time, **Click2State**, which serves a dual purpose. The first aim is to induce predictive student states, which provide substantial traction towards predicting whether a student is on a path towards passing or failing a course. Another is to provide us with insights into the process of passing or failing a course over time, and in particular leveraging the insights of human mentors whose observations give deeper meaning to the click level behavioral data, which is otherwise impoverished from an interpretability standpoint.

Byungsoo Jeon, Eyal Shafran, Luke Breitfeller, Jason Levin and Carolyn P. Rose "Time-series Insights into the Process of Passing or Failing Online University Courses using Neural-Induced Interpretable Student States" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 330 - 335

In the remainder of the paper we motivate our specific work as situated within the literature. Next we present our modeling approach and a series of experiments that investigate the following three research questions: (RQ1) How can we extract information and meaning from mentors' notes about the formation of student states across time? (RQ2) To what extent does integrating a representation of topical insights from *Mentor's Notes* improve the ability of a time series neural model to predict whether students are on a path towards passing or failing a course? (RQ3) How can we use insights about student progress in an online course captured using student state representations from our model to understand the process of passing or failing a course? The more comprehensive version of this paper is available at Arxiv ¹.

2. RELATED WORK

One of the most important challenges in providing analytic tools for teachers and administrators [1, 19] in online university is to model the population of students in such a way as to provide both predictive power for triggering interventions and interpretability for ensuring validity. Some past research has already produced models to identify at-risk students and predict student outcomes specifically in online universities [6, 16]. For example, Smith et al. [20] proposed models to predict students' course outcomes and to identify factors that led to student success in online university courses. Eagle et al. [10] presented exploratory models to predict outcomes like overall probability of passing a course, and provided examples of strong indicators of student success in the WGU platform where our work is also situated. However, this past work has focused mainly on predictive modeling of student outcomes, whereas our work pursues both predictive power and interpretability.

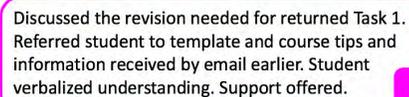
While much work in the field of Educational Data Mining explores time series modeling and induction of student state representations from open online platforms such as Massive Open Online Courses (MOOCs) or Intelligent Tutoring Systems, far less has been published from large, online universities such as WGU, which offer complementary insights to the field. Student states are triggered by students' interaction with university resources, their progress through course milestones, test outcomes, affect-inducing experiences, and so on. Affect signals in particular have been utilized by many researchers as the basis for induced student states, as this rich source of insight into student experiences has been proven to correlate with several indicators of student accomplishments [18]. Researchers have investigated affect and developed corresponding detectors using sensors, field observation, and self-reported affect. These detectors capture students' affective signals from vocal patterns [7, 17], posture [9], facial expressions [3, 17], interaction with the platform [4, 5, 12], and physiological cues [7, 15]. Although these signals provide rich insights, the requisite data is sometimes expensive or even impractical to obtain, even on for-pay platforms such as WGU, where we conduct our research.

The bulk of existing work using sequence modeling to induce student states has focused on the data that is most readily available, specifically, clickstream data. For example, Tang et al. [21] have constructed a model to predict a set of stu-

dent actions with long short-term memory (LSTM) [13] on student clickstream data from a BerkeleyX MOOC, though the basic LSTM was unable to match the baseline of defaulting to the majority class for samples of student actions. Fei et al. [11] proposed a sequence model to predict dropout based on clickstream data using recurrent neural network (RNNs) model, with more success. Wang et al. [22] also built a neural architecture using a mix of convolutional neural network (CNN) [14] and RNN for dropout prediction from clickstream data. Though these models have achieved differing success at their predictive tasks, a shared shortcoming is the lack of interpretability in the induced student state representations. Our work extends previous studies by proposing a model that enriches temporal signals from clickstream data using the textual mentor's notes to provide a means for interpreting student state representations.

3. DATA

Our study is based on data collected by Western Governor's University (WGU), an online educational platform ². To support self-paced learning, students in WGU are assigned to a program mentor (PM). The PM is in charge of evaluating a student's progress through their degree and helping to manage obstacles the student faces. A PM and a student generally have bi-weekly live calls, but this may vary depending on the student's needs and schedule. Each PM writes down a summary of what was discussed, which we refer to as a mentor's note. An example is given in Figure 1. Mentor's notes describe the status and progress of the student and what types of support was offered or what suggestions were made during the call. This information can provide meaningful cues to infer student states over time.



Discussed the revision needed for returned Task 1. Referred student to template and course tips and information received by email earlier. Student verbalized understanding. Support offered.

Figure 1: An example of mentor's notes.

In this work we specifically investigate how the use of the mentor's note data alongside the more frequently used clickstream data might enable that important goal. Clickstream data in WGU also provides us with information on how active students are and where in the WGU platform they spend their time. We collect clickstream data from four different types of web pages in the WGU platform: course, degree plan, homepage, and portal. The course web pages cover all pages related to courses in WGU. Degree plan represents a dashboard where students check their progress toward a degree. Homepage is the main page that shows students' progress in each course and allows access to all provided WGU features. Portal covers any other pages for student support including technical and financial assistance.

An example of the clickstream data can be seen in Table 1. Each row represents one of five different click sources: target course page, other course page, degree plan page, portal page, and homepage. We divide the course pages into "target course" and "other course". Each column represents one of different six click types: click count (C1), focus state count (C2), keypress count (C3), mousemove count (C4), scroll

¹<http://arxiv.org/abs/1905.00422>

²<https://www.wgu.edu/>

count (C5), and unfocused state count (C6). The values in the table represent the weekly count of different type of clicks from each different source.

For this paper, we have collected the mentor’s notes and clickstream data from two courses conducted in 2017: Health Assessment (HA) and College Algebra (CA). We choose these two courses because they are popular among students and represent different levels of overall difficulty. Table 2 shows the statistics for the dataset. “Average prior units” is the average number of units students transferred to WGU from prior education when they started the degree, and functions as a proxy for the level of student’s prior knowledge. We split the dataset for each course into a training set (80%), a validation set (10%), and a test set (10%). For training, to avoid a tendency for trained models to over-predict the majority class, we have resampled the training set so that both the pass state and the fail state are represented equally.

	C1	C2	C3	C4	C5	C6
Target course	53	61	0	168	904	1732
Other courses	177	167	0	455	2301	4887
Degree plan	0	0	0	0	0	0
Portal	21	89	0	263	3862	2440
Homepage	36	69	0	122	72	1581

Table 1: Example of clickstream data.

	HA	CA
# of students	6,041	4,062
Length of a term	25 weeks	25 weeks
Avg prior units	62 ± 39	11 ± 23
Fail rate	0.185	0.509
Avg # of notes per student	10.9 ± 5.7	11.0 ± 5.8
Avg length of notes (chars)	198 ± 47	194 ± 55

Table 2: Data Statistics

4. PROPOSED METHOD

As we have stated above, in our modeling work, we propose a sequence model, **Click2State**, with two primary purposes. The first is to form a student state representation that will allow us to better identify students at risk of failing a course than a baseline model that does not make use of rich textual data. The second is to provide us with a means to interpret the meaning of a student state representation.

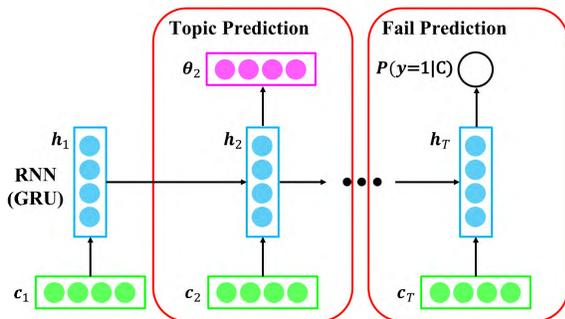


Figure 2: Architecture of Click2State Model.

Figure 2 provides a schematic overview of our proposed model. Note that it is first and foremost a sequence model that predicts whether a student will pass or fail a course

based on an interpretable student state that evolves from week to week as each week’s clickstream data is input to the recurrent neural model. A summary of the content of the mentor’s note for a week is constructed using a popular topic modeling technique, specifically Latent Dirichlet Allocation (LDA) [2]. In the full model, an intermittent task to predict the topic distribution extracted from the mentor’s notes associated with a time point is introduced. The goal is to use this secondary task to both improve the predictive power of the induced student states over the baseline as well as to enhance the interpretability of the state representation.

Feature Vector Design We train our model using clickstream feature vectors (as input) and topic distribution vectors (as output for the topic prediction task). We design the clickstream feature vector to include both an encoding of click behavior of students from a time period as well as a control variable that represents the prior knowledge of students as estimated by the number of units they were able to transfer in. The full clickstream feature vector contains thirty weekly counts for each different type and source of click, in addition to the single control variable just mentioned, which is the number of transferred units. We use min-max normalization to scale the values between 0 and 1. To extract a topic distribution vector for each mentor’s note, we run Latent Dirichlet Allocation (LDA) over the whole set of mentor’s notes from the entire training dataset.

Formal Definition of the Model Denote the student’s clickstream features by $C = (c_1, c_2, \dots, c_T)$, where c_t is the clickstream feature vector of t th week, and T is the number of weeks for the term. The clickstream feature vectors are encoded via Gated Recurrent Units (GRU) [8], which are variants of the Recurrent Neural Network (RNN). At each time step t , this network constructs a hidden state of the student for the t th week, $h_t \in \mathbb{R}^H$, where H is the dimensionality of the hidden state. We consider h_t as the student state representation at t th week. Based on the generated student state representation from RNN (h_t), our model is trained to predict a topic distribution of a mentor’s note and the probability of the student failing that course.

Topic Prediction Given the generated hidden states from RNN (h_t) for the t th week, the model estimates the true topic distribution ($\theta_t \in \mathbb{R}^{N_t}$) of a mentor’s note on t th week where N_t is the number of topics. The estimated topic distribution ($\hat{\theta}_t \in \mathbb{R}^{N_t}$) is computed by taking h_t as an input of one fully connected layer (weight matrix: W_θ) whose output dimensionality is N_t followed by a softmax layer.

$$\hat{\theta}_t = \text{Softmax}(W_\theta h_t)$$

Fail Prediction As data from a student’s participation in a course is fed into the RNN week by week, the model estimates the probability of the student failing that course ($P(y = 1|C)$) at the last timestep T . The estimated probability is computed by taking h_t as an input of one fully connected layer (weight matrix: W_y) whose output dimensionality is one followed by a sigmoid layer.

$$P(y = 1|C) = \text{Sigmoid}(W_y h_t)$$

Loss The loss function is composed of KL divergence loss for the topic prediction and binary cross-entropy loss for the fail prediction. Assume there are a total of N students. The

KL divergence loss of topic distribution of the mentor’s note for n th student at time t is defined as:

$$KLD_{n,t} = D_{KL}(\theta_{n,t} \parallel \hat{\theta}_{n,t}),$$

where $\theta_{n,t}$ and $\hat{\theta}_{n,t}$ are the true and estimated topic distribution of the mentor’s note at time t for n th student. The binary cross-entropy of the n th student measures the similarity between $P(y_n = 1|C)$ and the true y_n as:

$$BCE_n = -y_n \log P(y_n = 1|C) - (1 - y_n) \log(1 - P(y_n = 1|C)),$$

Assume that there are a total of N_n mentor’s notes for n th student. Combining the two losses, our final loss is

$$\frac{1}{N} \sum_{n=1}^N [\lambda BCE_n + (1 - \lambda) \frac{1}{N_n} \sum_{t=t_{n,i}}^{t_{n,N_n}} KLD_{n,t}],$$

where $t_{n,i}$ is the timestep when n th student has i th mentor’s note, and λ is the rescaling weight.

5. RESULTS

In this section, we answer our aforementioned research questions one by one with the experiment results.

RQ1. What types of information about student states can we extract from mentor’s notes? We answer the question of how student state information may be extracted from mentor’s notes through application of LDA to the notes. We set the number of topics to ten to maximize the interpretability of the results. Table 3 shows the learned topics with manually assigned labels, topical words, and text. Topical words are the top ten words with the highest probability of appearing within each learned topic, and are presented in decreasing order of likelihood. The topical text column contains an example snippet from one of top ten mentor’s notes for each topic. We exclude the one topic that was incoherent out of the 10 learned topics.

Note that there are four topics related to student progress and plan, term plan (T5), course progress (T6), term progress (T7), and goal setting (T9). Course progress and goal setting (T6, T9) focus on progress towards modules in a particular course, along with past and present goals about the course itself. Term plan and term progress (T5, T7) emphasize discussions about plans for a term, such as courseload within in a term, course selection, and long-term degree planning. There is a clear utility to these topics as an interpretation tool for regulation of the student’s process moving through the curriculum—if a student hits an impasse, mentor’s notes are expected to focus on what challenges the student experienced and how to address these challenges.

The remaining six topics provide insight on specific issues and circumstances a student may be facing at a particular time, and which may end up impacting their overall progress. In revision (T1), we discover students seeking feedback on revisions, suggesting significant engagement with the platform. In question (T2), students ask for tips on using WGU platforms, course logistics, and how to succeed in a given course. In time constraint (T8), students point out time constraints in their daily life to explain why goals were not met. The time constraint (T8) topic may explain

abnormal absence or dropout. Assessment (T3) contains the result or plan of assessments and review for exam (T4) includes progress or plans of review for exam preparation.

RQ2. Does the task of topic prediction construct better student state representation than our baseline, as evaluated by the ability to predict student failure? We measure the predictive power of learned student state representation from our model and compare with that of our baseline, which shares the same neural architecture but is not trained on the extra task of topic prediction. The specific predictive task is to determine whether a student fails a course within a given term given a sequence of weeks of student clickstream data. We trained separate models to make a prediction after a set number of weeks so that we could evaluate the difference in predictive performance depending on how many weeks worth of data were used in the prediction. We measure the AUC scores of our model and baseline using data from two WGU courses: Health Assessment (HA) and College Algebra (CA).

Figure 3(a) shows the AUC scores across time steps for the HA course while Figure 3(b) shows the AUC scores across time steps for the CA course. For HA, our model achieves a statistically significant improvement (p-value < 0.05) in performance over the baseline model after the 5th week. For CA, our model achieves a statistically significant improvement after the 17th week. This difference in model performance between the HA and CA courses suggests the result from CA-specific topic data adds limited predictive power to the model. It is possible the clickstream data of students taking CA already contains enough information about whether a student is going to fail, a conclusion supported by the fact that AUC scores of the baseline model for CA are always better across time steps than those for HA.

Figure 3(c) exhibits the minimum KL Divergence loss of our model across time steps to determine how well our model is predicting the topic distribution of each mentor’s note for each course. Though we determined that adding this task improves the fail prediction task, results on this task specifically are not impressive, demonstrating the relative difficulty of predicting mentor’s notes from click data.

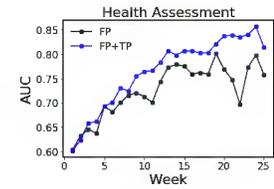
RQ3. What insights do we gain about the process of passing or failing a course over time from predicted mentor’s notes topic distributions over time from the model? We perform two different experiments on the dataset of clickstream and mentors’ notes data of students taking the College Algebra course. We choose this course because our topic prediction loss was lower (and thus, accuracy higher) for the course. First, we determine what topics inferred from our model correlate with whether a student will pass or fail a course. Then we find sequences of standardized topic probabilities of each topic inferred by our model that characterize students likely to pass or fail.

State	Assessment	Course progress	Term progress
P	-0.1486	0.6301	0.2298
F	0.1735	-0.0049	-0.1647

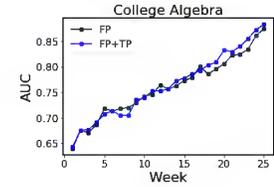
Table 4: Standard Score of Inferred Topic Probabilities from P and F State

Topic	Topical Words	Topical Text
T1. Revision	task, submit, revise, discuss, equate, complete, need, write, practice, paper	The ST and I discussed his Task 3 revisions after he made some corrections. The ST still needs to revise the task based on the evaluator's comments. He plans to do more revisions that align with the task rubric and submit the task soon.
T2. Question	student, question, call, email, send, course, discuss, appoint, speak, assist	Student emailed for help with getting started. CM called to offer support. Student could not talk for long. CM emailed welcome letter and scheduling link and encouraged for student to make an appointment
T3. Assessment	week, goal, today, schedule, pass, take, exam, final, work, talk	C278: took and did not pass preassessment, did not take final . NNP C713: took and did not pass the preassessment. Passed LMC1 PA with a 65 on 02/27. LMC1 exam scheduled for 02/27
T4. Review for exam	student, review, assess, plan, study, attempt, discuss, complete, take, report	Student scheduled appointment to review for first OA attempt but had taken and not passed the attempt by the time of the appointment.
T5. Term plan	student, discuss, course, complete, engage, college, term, plan, pass, progress	Discussed final term courses . Discussed starting C229 and working through hours and then working through C349 course .
T6. Course progress	goal, course, progress, current, complete, previous, work, date, pass, module	Current course : C349 Previous goal: completed modules 1-3 and engage shadow health by next appt date Progress toward goal: Yes New Goal: shadow health completed and engaged in video assessment
T7. Term progress	term, course, complete, date, goal, week, progress, current, leave, remain	Date: 8/22/17 Term Ends: 5 weeks OTP Progress: 5/14 cu completed Engaged Course: C785 Goal Progress: did not pass PA
T8. Time constraint	work, week, lesson, complete, go, progress, plan, finish, time, goal	NNP stated he was not able to make forward progress in course related to personal situation and time constraints from an unexpected event.
T9. Goal setting	goal, week, work, complete, task, progress, pass, accomplish, finish, contact	Previous goal: finish shadow health, finish and submit video by next call, start c228 next Progress /concerns: states working on c349 SH, discussed deadlines Goal: finish shadow health

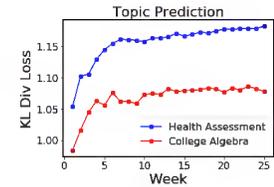
Table 3: LDA Topics Learned From Mentor's Notes



(a) Fail Prediction Performance on HA



(b) Fail Prediction Performance on CA



(c) Topic Prediction Performance

Figure 3: Performance of Fail and Topic Prediction

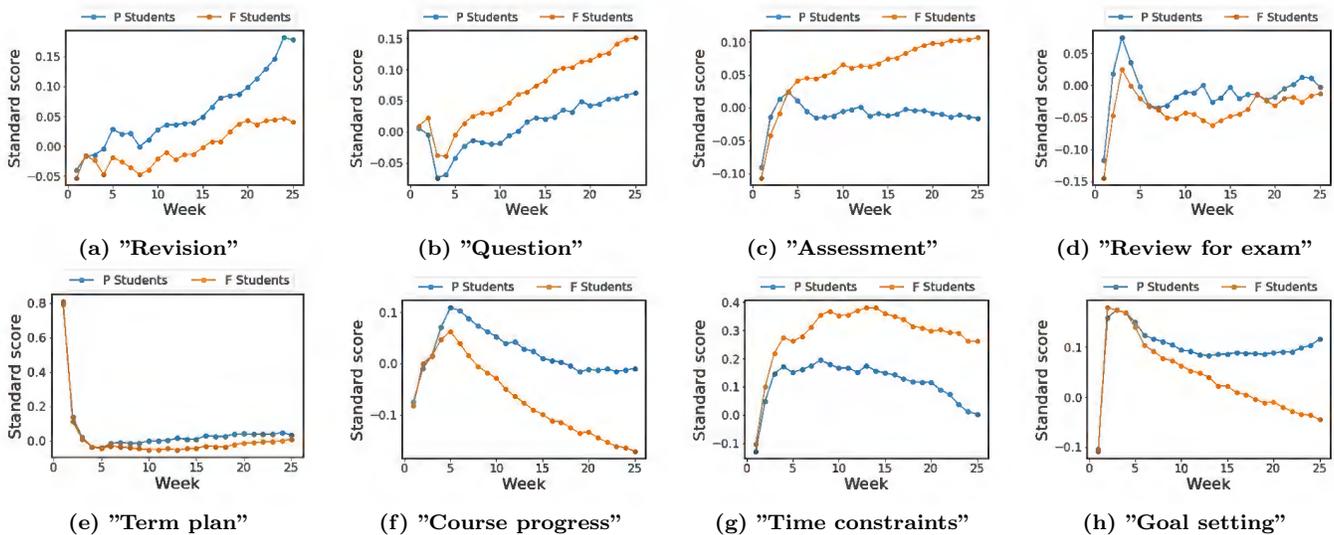


Figure 4: Standard Score of Each Topic Probability across Weeks for P and F Students

Experiment 1. In the first experiment, we find the two student state representations which minimize or maximize the probability of failing a course. We call the state representations that minimize and maximize the probability of failure as a P state and F state. Then, we show what topic distributions are inferred from each state. We represent emphasis, or a lack thereof, on a topic by standardizing topic probabilities and observing the number of standard deviations above and below the mean of a topic probability (standard score).

Table 4 shows the standard score for inferred topic probability from the P and F state. We only present the standard score of topics that vary wildly between P and F state. For example, the standard score of assessment topic (T3) for the P state is negative and for the F state is positive. One interpretation is that students likely to fail have more trouble in passing assessments, and thus talked to their mentors more about assessment topic (T3). The standard score of course and term progress (T6, T7) for the P state is positive and

negative for the F state, which shows students likely to pass report smoother progress instead of ongoing issues.

Experiment 2. We compare the trajectory of inferred probability of each topic by our model from students who passed (P students) and failed (F students) a course. Figure 4 shows the average standard score of topic probability per topic for P and F students over time.

We can see through this experiment clear, distinct patterns for the frequency of each topic over time that make intuitive sense given the format of online courses. For example, term plan (T5) is high frequency for the first week and plunges right after, since most students and mentors will naturally discuss plans for a term at the start of each term. The standard scores of other topics related to goal and progress (T6,T7,T9) also decrease over time, likely for similar reasons; the plot for T7 is omitted to save space, but it shows the similar pattern as T6. The standard scores of revision (T1), question (T2), and assessment (T3), meanwhile, increase over time, which may indicate students seek help more actively as they approach the end of a term. The standard scores of review for exam (T4) increase dramatically until the third week, decrease for few weeks, and finally level off. As the only condition for students in WGU to pass a course is to pass the final assessment, it may be that many students take their final assessments during the earlier weeks so they can pass a course as early as possible. The standard scores of time constraints (T8) steeply increase until the fourth week, and then gradually decrease over time. This suggests that when students begin a term they do not expect to have time constraints, but accumulate unanticipated issues in their personal lives as the course goes on.

For most topics, the P and F students exhibit distinct divergences in topic patterns. For topics related to goal and progress (T6, T7, T9), the gap between P and F students increases over time—suggesting that as time goes on F students will be reporting obstacles to their mentors instead of positive progress. The gap between P and F students for question (T2) increases over time, likely for similar reasons. For revision (T1), P students generally have higher standard scores than F students over time, supporting the idea that P students actively seek opportunities for revision towards the end of a term. For assessment (T3), standard score for F students increases over time while score for P students decreases. This could suggest that F students are more likely to procrastinate and struggle with their assessments than P students. Finally, for time constraints (T8) F students show higher standard score as time goes on. A likely interpretation is that students who encounter time constraints cannot devote focus to a course and are more likely to fail.

6. CONCLUSION

In this paper, we propose and evaluate a sequence model, **Click2State**, which aims to build an interpretable student state representation by leveraging mentor’s notes to give deeper meaning to impoverished clickstream data. We also introduce a methodology for interpreting the learned representation from our model that extracts time-sensitive insights about the process of passing or failing a given online course. Our experimental results demonstrate that student state representations learned by our model have better pre-

dictive power on the task of determining student failure rate than a baseline that only uses click stream data. We also present how individual topic-based insights into the process of passing or failing a course let us construct a rich characterization of a student likely to fail or pass an online course.

7. ACKNOWLEDGEMENT

This work was funded in part by NSF grant IIS 1822831.

References

- [1] Alison Ashby. Monitoring student retention in the open university: definition, measurement, interpretation and action. *Journal of Open, Distance and e-Learning*, 2004.
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *JMLR*, 2003.
- [3] Nigel Bosch, Yuxuan Chen, and Sidney D’Mello. It’s written on your face: detecting affective states from facial expressions while learning computer programming. In *ICITS*, 2014.
- [4] Anthony F Botelho, Ryan S Baker, and Neil T Heffernan. Improving sensor-free affect detection using deep learning. In *ICAIED*, 2017.
- [5] Anthony F. Botelho, Ryan Shaun Joazeiro de Baker, Jaclyn Ocuppaugh, and Neil T. Heffernan. Studying affect dynamics and chronometry using sensor-free detectors. In *EDM*, 2018.
- [6] Carol Elaine Calvert. Developing a model and applications for probabilities of student success: a case study of predictive analytics. *Journal of Open, Distance and e-Learning*, 2014.
- [7] Rafael A Calvo and Sidney D’Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 2010.
- [8] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv:1409.1259*, 2014.
- [9] Sidney D’Mello and Arthur Graesser. Mind and body: Dialogue and posture for affect detection in learning environments. *Frontiers in Artificial Intelligence and Applications*, 2007.
- [10] Michael Eagle, Ted Carmichael, Jessica Stokes, Mary Jean Blink, John Stamper, and Jason Levin. Predictive student modeling for interventions in online classes. In *EDM*, 2018.
- [11] Mi Fei and Dit-Yan Yeung. Temporal models for predicting student dropout in massive open online courses. In *ICDMW 2015*.
- [12] Neil T Heffernan and Cristina Lindquist Heffernan. The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *IJAIED*, 2014.
- [13] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 1997.
- [14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012.
- [15] Changchun Liu, Pramila Rani, and Nilanjan Sarkar. An empirical study of machine learning techniques for affect recognition in human-robot interaction. In *ICIRS*, 2005.
- [16] Julia M Matuga. Self-regulation, goal orientation, and academic achievement of secondary students in online university courses. *Journal of Educational Technology & Society*, 2009.
- [17] Mihalis A Nicolaou, Hatice Gunes, and Maja Pantic. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing*, 2011.
- [18] Zachary A Pardos, Ryan SJD Baker, Maria OCZ San Pedro, Sujith M Gowda, and Supreeth M Gowda. Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, 2014.
- [19] Bart Rienties, Avinash Boroowa, Simon Cross, Chris Kubiak, Kevin Mayles, and Sam Murphy. Analytics4action evaluation framework: A review of evidence-based learning analytics interventions at the open university uk. *Journal of Interactive Media in Education*, 2016.
- [20] Vernon C Smith, Adam Lange, and Daniel R Huston. Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, 2012.
- [21] Steven Tang, Joshua C Peterson, and Zachary A Pardos. Deep neural networks and how they apply to sequential education data. In *L@S*, 2016.
- [22] Wei Wang, Han Yu, and Chunyan Miao. Deep model for dropout prediction in moocs. In *ICCSE 2017*.