Adding duration-based quality labels to learning events for improved description of students' online learning behavior

Matthew W. Guthrie
University of Central Florida
4111 Libra Drive
PSB 319
matthew.guthrie@ucf.edu

Zhongzhou Chen
University of Central Florida
4111 Libra Drive
PSB 153
zhongzhou.chen@ucf.edu

ABSTRACT

Many existing studies analyzing log data from online learning platforms model events such as accessing a webpage or problem solving as simple binary states. In this study, we combine quality information inferred from the duration of each event with the conventional binary states, distinguishing abnormally brief events from normal or extra-long events. The new event records, obtained from students' interaction with 10 online learning modules, can be seen as a special form of language, with each "word" describing a student's state of interaction with one learning module, and each "sentence" capturing the interaction with the entire sequence. We used second order Markov chains to learn the patterns of this new "language," with each chain using the interaction states on two given modules to indicate the interaction states on the following two modules. By visualizing the Markov chains that lead to interaction states associated with either disengagement or high levels of engagement, we observed that: 1) disengagement occurs more frequently towards the end of the 10 module sequence; 2) interaction states associated with the highest level of learning effort rarely leads to disengaged states; and 3) states containing brief learning events frequently lead to disengaged states. One advantage of our approach is that it can be applied to log data with relatively small numbers of events, which is common for many online learning systems in college level STEM disciplines. Combining quality information with event logs is a simple attempt at incorporating students' internal condition into learning analytics.

Keywords

Markov chains, online learning modules, log data analysis

1. INTRODUCTION

Understanding and predicting students' learning behavior by mining the log files of online and computerized learning systems has been the focus of a significant body of research in educational data mining. For example, many studies have modeled student learning as a chain of ordered events such as opening a page, viewing a video, or solving a problem [8, 9]. Events are often represented by a binary variable, e.g. whether the student accessed a webpage or answered a problem correctly [8, 14]. While describing events using binary variables can significantly reduce the complexity of the data, doing so can remove important information from event logs. One indicator of the quality of an event is its duration. For example, abnormally short problem-solving attempts have been associated with either random guessing due to low test-taking effort [4] or answer copying [1]. In two earlier studies [6, 7], we demonstrated that learning events can be separated into "Brief" (B) and "Normal" (N) categories by applying a mixture model clustering algorithm using the time-on-task data alone, since other measurements such as the number of practice problems answered are highly correlated with time-on-task.

Matthew Guthrie and Zhongzhou Chen "Adding duration-based quality labels to learning events for improved description of students' online learning behavior" In: *The 12th International Conference on Educational Data Mining,* Michel Desmarais, Collin F. Lynch, Agathe Merceron, & Roger Nkambou (eds.) 2019, pp. 560 - 563

We combine duration-based categorical quality labels such as "Brief" or "Normal" with conventional binary event states, such as "Pass" or "Fail," to analyze the log data obtained from students' interaction with 10 Online Learning Modules (OLMs). OLMs are a form of online instructional design in which students progress through learning modules in a pre-determined order [5–7]. Students are required to attempt the assessment problems at least once before accessing the accompanying learning resources in each OLM. The restrictive structure of OLMs has major advantages for data analysis, providing more accurate estimations of duration information. Assessments and learning events are closely coupled on each module; the OLM structure itself improves the interpretability of log data events.

We combine the event logs with categorical quality labels to form a simple artificial language. Each event, such as "Brief Pass" or "Normal Fail," becomes part of a "word" that captures students' interaction with either the assessments or the learning components of a module. Four such "words" form a "phrase" that describes a student's state of interaction with one section of the OLM, and each 10 word "sentence" corresponds to a student's interaction with the entire OLM sequence. Can we gain insight into the patterns in students' learning behavior by understanding the underlying "grammar" of this artificial language? Are there certain combinations of words in these phrases that are frequently followed by other words which indicate that the student is either disengaged or highly engaged with the learning from the OLM sequence? We answer these questions by utilizing a second order Markov chain, a common technique used in natural language processing.

The Markov model can be trained using a relatively small number of events from a student population of approximately 250, owing to the increased information by introducing the quality labels. This advantage is critical for modeling student learning behavior for many STEM disciplines, where solving one problem can take 5 to 10 minutes. A typical weekly online homework assignment of 10 problems may only generate 30 to 50 major events. We demonstrate that it is possible to construct a Markov model using log data from 10 OLMs for a college physics class of approximately 250 students. The resulting Markov chains are visualized to reveal combinations of states that lead to states associated with either disengagement or high levels of engagement in following modules. Engagement is a complex concept that has different definitions depending on context and measurement [2]. We adopt a pragmatic definition of engagement to indicate that students spent a normal or extended amount of time (and likely cognitive resources) on consecutive modules, emphasizing the cognitive and behavioral aspect of engagement, bearing similarity to the definitions proposed by Miller [11] and described by Motz, et al. [12].

2. METHODS

2.1 Structure of OLM and OLM Data

Data analyzed in this study were collected from student interaction with 10 OLMs assigned as homework to be completed over a period of two weeks in a calculus-based college physics class. Students were not required to finish the entire sequence in one session, but the modules must be completed in the order given. As described in detail in several earlier papers [5–7], each OLM consists of an assessment component (AC) and an instructional component (IC). Students are required to attempt the AC at least once before being able to learn from the IC and can make additional attempts after interacting with the IC. Most interactions with each OLM can be divided into three stages: ***Pre-Learning***: attempting the AC once or twice before accessing the IC, ***Learning***: interacting with the IC after one or two initial failed AC attempts, and ***Post-Learning***: making additional attempts on the AC after learning from the IC. If a student passes the AC during the Pre-Learning stage, they will not have the following two stages as the student will immediately proceed to the next OLM.

2.2 Combining Quality Labels with Events

For each Pre-Learning and Post-Learning stage, students' interaction with the AC is captured by the attempt outcomes: "Passing" (P) or "Failing" (F). The quality of each attempt is estimated from the duration of the attempt, t, which is classified into three categories: "Brief" (B: $t < 40$ s), "Normal" (N: $40\text{ s} \leq t < 180$ s), and "Extensive" (E: $t \geq 180$ s). These cutoffs were determined based on a mixture-model clustering method applied to log-transformed attempt duration data [7]. Combining the quality categories with the attempt outcomes results in six different states: BF, BP, NF, NP, EF, and EP. The EF and EP states are only assigned to the Post-Learning stages, since there were significantly fewer attempts with $t \geq 180$ s in the Pre-Learning stages, and it was less clear whether those longer attempts resulted from longer problem-solving time or students leaving the system. Previous work explains these categories and cutoffs in detail [7].

For the Learning stage, students' interaction with the IC was modeled as a single learning event described by a binary variable. The duration of the learning event is classified as "Brief" or "Normal" according to cutoffs determined for each module by a mixture-model clustering analysis of learning time distribution [7]. An isolated "Brief" category does not necessarily imply that the event is of lower quality. Brief learning can result from a student having a high level of incoming knowledge and only needed to quickly view the learning resources to answer the problem. In a small number of cases, students made 3 or more failed attempts on the AC before accessing the IC or kept attempting the AC until all attempts were used up without accessing the IC. Those cases are classified as "Other." In even fewer cases, due to a corrupted log file or other system glitch, some students were able to proceed to the next module without finishing the current module. Those cases were classified as "NAOther," making a total of 28 possible states, listed in Table 1.

Table 1: All possible interaction states. D: Disengaged. E: Highly Engaged. See section 2.2 for other acronym definitions.

State	Rank	Indication	State	Rank	Indication
NAOther	0	D	BF-N-EP	14	E
Other	1	D	NP- -	15	E
BP- -	2	D	NF-B-BF	16	
BF-B-BF	3	D	NF-B-BP	17	
BF-B-BP	4	D	NF-B-NF	18	
BF-B-NF	5		NF-B-NP	19	
BF-B-NP	6		NF-B-EF	20	
BF-B-EF	7		NF-B-EP	21	
BF-B-EP	8		NF-N-BF	22	
BF-N-BF	9		NF-N-BP	23	
BF-N-BP	10		NF-N-NF	24	E
BF-N-NF	11		NF-N-NP	25	E
BF-N-NP	12		NF-N-EF	26	E
BF-N-EF	13	E	NF-N-EP	27	E

2.3 Defining States Associated with Either Disengagement or High Engagement

It is difficult to estimate the level of engagement associated with a state, and the same state can be observed from students with different levels of engagement, though there are several states that are more likely to be associated with either a very low or a very high level of engagement with the learning process. For example, "Brief" learning events are more likely than "Normal" learning events to come from students who skimmed through the content [7]. Students displaying consecutive "Brief" events on the same module, such as in state BF-B-BF, are more likely to be disengaged with the learning process.

Consecutive "Normal" or "Extensive" events are more likely to come from students who are highly engaged with learning. While individual "Extensive" events may be caused by a student leaving the computer without exiting from the module, it is much less likely that three such events occur on the same module. We assumed that states with three consecutive "B" labels or BP- - are more likely associated with disengagement. BP - - is included because "Brief" problem solving occurs in under 40 s, which likely resulting from a guessing attempt, or answer copying event [7]. States with three consecutive "N" or "E" labels or NP- - are likely associated with higher engagement. We did not distinguish between productive and unproductive engagement; failed attempts are also included in high engagement states. There are two exceptions to these rules: First, the "Other" state is classified as "Disengaged," since most engaged students should at least look at the instructional resources after two failed attempts. Second, the BF-N-EF and BF-N-EP states are classified as highly engaged, since it is possible that the student quickly decided that the assessment problem was too difficult and immediately engaged in the learning process.

2.4 Training of the Markov Model

We define four sequential states to be a "phrase," a balance between model accuracy and available computational power. Phrases are analyzed in the context of the modules in which they occurred. Ten states are defined as a "sentence" and each student contributed a ten-state sentence to the text corpus. The Markovify Python library [10] was used to parse the text corpus. We utilized second order Markov chains because student interaction on a single module is unlikely to adequately indicate the complexity of their subsequent behavior. The Markov model was used to build second order Markov chains for every combination of initial states in modules 1 and 2, modules 6 and 7, and modules 7 and 8. The Markov chains were used to investigate student behavior as they progressed through the modules and to infer how changes in behavior can be related to student engagement levels.

3. RESULTS

3.1 Outcomes of Second Order Markov Chain

For a given pair of states on the two input modules, we use the Markov model to return the probability of observing states in the two following modules. We consider a chain as "probable" if the probability of the last two states adds up to more than 100%. On average, 0.2% of all chains generated are considered probable. The three cases for which we ran the Markov model are listed in Table 2, along with the number of probable chains in each case. Those three cases are of particular interest because a previous analysis of the data [7] revealed that more students have lower levels of engagement in modules 3, 8, 9, and 10.

Table 2: Combinations of input states for modules (M1-M10) which were analyzed in this study.

Case	Input	Predict	All chains	Probable chains	Disengaged chains	Highly Engaged chains
I	M1+M2	M3+M4	55664	169	3	66
II	M6+M7	M8+M9	45472	65	18	15
III	M7+M8	M9+M10	72912	108	31	16

3.2 Markov Chains Leading to Consecutive Disengaged or Highly Engaged States

We are interested in chains that could indicate engagement or disengagement with the learning process. We consider a student disengaged from the learning process if their interaction states are indicative of disengagement on the last two modules in a chain. If a student's interaction states are associated with high engagement on the last two modules, the student is considered as highly engaged. The number of chains that lead to consecutive disengaged or highly engaged states on the last two modules are listed in Table 2. Other chains are not included in this analysis because the relation between engagement and the states on the last two modules were not as clear as the chains included. We plot all the chains indicating disengagement or high engagement for each of the three cases in Figure 1. In each case, the 28 interaction states are arranged according to the order listed in Table 1.

This ordering groups states according to similarities in Pre-learning, Learning, and Post-Learning stages, listed from low to high in the order of "B," "N," and "E" in quality labels. States with passing events are assumed to be of higher engagement than those with failing events. States associated with disengagement are placed at the bottom and states associated with high engagement are placed at the top, except for states 13-15 in Table 1. For states near the middle of the pack, the ranking does not reflect the learning effort required for each state, as it is difficult to estimate the level of effort required by 11: BF-N-NP compared to 19: NF-B-NP.

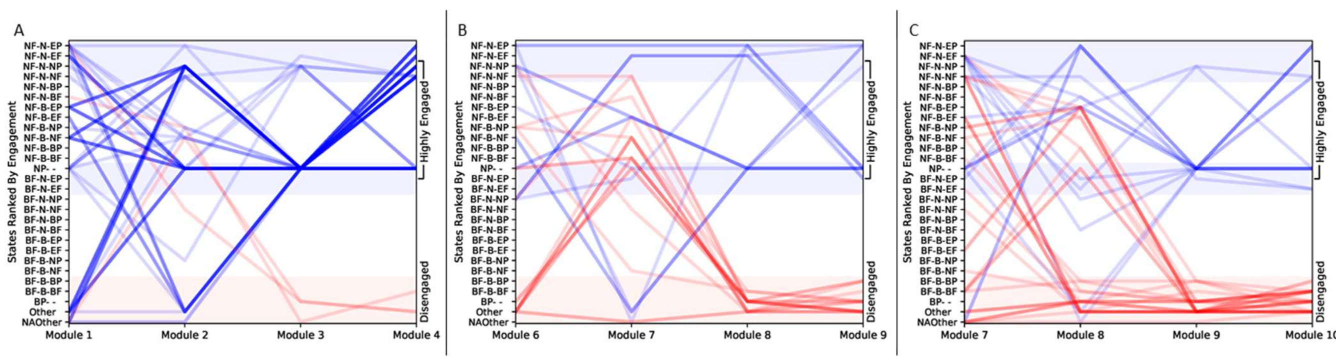

Figure 1: Probable chains that lead to consecutive disengaged states (red) or highly engaged states (blue). Darker lines indicate where multiple chains overlap; blue and red zones highlight states associated with disengagement or high engagement, respectively.

Figure 1A (M1-M4) shows significantly more chains leading to highly engaged states on M3-M4 over disengaged states. On M1, those chains started from either a variety of states above NP- -, or from Other and NAOther. On M2, most of the chains concentrated on three states: 25: NF-N-NP, 15: NP- -, and 1: Other. While 25 and 15 were highly populated states in the original text corpus, state 1 was scarcely populated for M2; very few students were consistently disengaged on both M3 and M4, which appear early in the learning sequence and cover less difficult concepts. Figure 1B (M6-M9) shows there were almost an equal number of chains leading to either engagement or disengagement on M7 and M8. Most of the chains leading to disengagement passed through one of the states between 15 and 24 on M7. More than half of those chains started in states 2 and 3 on M6. Several chains leading to high engagement started with high effort states on M6 and passed through Other or NAOther on M7. Every chain (except one) starting from or passing through one of the top three states (25-27) led to high engagement, and every chain starting with disengaged states led to disengagement on M8 and M9. Figure 1C (M7-M10) shows there were more chains leading to disengagement over engagement on M9 and M10. The chains leading to disengagement started at a variety of states on M7, and forms two groups on M8. The first group passes through disengagement between states 0 and 4, and the second group passes through states between 15-21. Most of the high engagement chains either started from state 26 on M7 or passed through state 27 on M8.

4. DISCUSSION

By comparing Figure 1A-C we identify four common patterns. **1. Disengagement happens late:** chains leading to disengagement occur much more frequently on later modules in the sequence. This type of behavior is expected since the difficulty of the modules increase towards the end, yet each module is worth the same amount of course credit. Therefore, students had less incentive to devote effort on the harder modules. **2. Disengagement-free states:** states 25-28 are seemingly "immune" to chains leading to disengagement. In all three cases, only two of those chains pass through these states, while most of the chains leading to high engagement involve those states on at least one input module. Students who spend an extensive amount of effort on one module are more likely to be more persistent, especially on difficult modules towards the end. **3. Disengagement "hot zone":** states 15-24 (the white area between two blue bands in Figure 1) seem to be a "hot zone" for chains leading to disengagement in all three cases, especially when the state appeared on the second module in the sequence. Those states have a "Brief" label on either the learning stage or the post-learning stage. A brief event in the learning and post learning stages could indicate subsequent disengagement behavior more effectively than a brief event in the pre-learning stage. **4. V-shaped high engagement chains:** several chains leading to high engagement started with states beyond 15 and passed through either Other or NAOther on the second module, forming a V-shape on Figure 1C. Even highly engaged students may occasionally display disengaged states on certain modules. It also may suggest that the two "other" states are not always associated with disengagement as previously thought. These patterns can be valuable for development of an intelligent and personalized learning system that recommends different learning resources to appropriate student populations [13]. The patterns can also help instructors prioritize efforts in improving the OLMs.

There are several caveats to be investigated and addressed in future studies. Only three pairs of modules were analyzed. Whether the patterns observed are general to all modules or specific to the selected cases can be answered by analyzing every pair of input modules. We adopted a narrow definition of engagement; students spent an expected or extended amount of time completing each component in a single module. This definition is appropriate for the purposes of the current study, but the relation between time-on-task and engagement should be investigated. The current Markov model "predicts" student behavior based on their interaction states in preceding modules, but students' decisions to engage or disengage from learning involves complex metacognitive processes influenced by multiple external factors (knowledge, instructional condition, metacognitive skills, and emotional states) [3]. We can achieve more accurate outcomes by including more factors that influence students' metacognitive processes. This study utilizes simple labels obtained from clustering algorithms on time-on-task data. Future studies should investigate the validity of those labels, and to find new and better-quality indicators for the existing events and new events in other learning systems.

We relied on the restrictive structure of OLMs, providing a regular and simple data structure and allowing for straightforward interpretation of interaction states. Events and quality labels used to generate the artificial language can be obtained from essentially any online learning platform, and more sophisticated Markov models are capable of learning languages with many more irregularities. An extension of this work will be to investigate how the current method can be modified and applied to more common learning systems that are more accessible to the average instructor.

5. ACKNOWLEDGMENTS

We thank the Learning Science and Technology team at the University of Central Florida led by Dr. Francisca Yonekura for developing the Obojobo platform and providing technical support.

6. REFERENCES

[1] Alexandron, G. et al. 2017. Copying @ Scale: Using Harvesting Accounts for Collecting Correct Answers in a MOOC. *Computers & Education*. 108, (2017). DOI=https://doi.org/10.1016/j.compedu.2017.01.015.

[2] Azevedo, R. 2015. Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist*. 50, 1 (Jan. 2015), 84–94. DOI=https://doi.org/10.1080/00461520.2015.1004069.

[3] Azevedo, R. et al. 2018. Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. *Handbook of self-regulation of learning and performance (2nd ed.)*. (2018), 254–270. DOI=https://doi.org/10.4324/9781315697048.

[4] Barry, C.L. et al. 2010. Do Examinees Have Similar Test-Taking Effort? A High-Stakes Question for Low-Stakes Testing. *International Journal of Testing*. 10, 4 (2010), 342–363. DOI=https://doi.org/10.1080/15305058.2010.508569.

[5] Chen, Z. et al. 2019. Measuring the effectiveness of online problem-solving tutorials by multi-level knowledge transfer. *2018 Physics Education Research Conference Proceedings* (Jan. 2019), 1–4.

[6] Chen, Z. et al. 2018. Re-designing the Structure of Online Courses to Empower Educational Data Mining. *Proceedings of 11th International Educational Data Mining Conference* (Buffalo, NY, 2018), 390–396.

[7] Chen, Z. and Guthrie, M. 2019. Measuring the Effectiveness of Learning Resources Via Student Interaction with Online Learning Modules. *arXiv: 1903.08003*. (Mar. 2019).

[8] Faucon, L. and Kidzi, Ł. 2016. Semi-Markov model for simulating MOOC students. *Proceedings of the 9th International Conference on Educational Data Mining (EDM)*. (2016).

[9] Hansen, C. et al. 2017. Sequence Modeling for Analysing Student Interaction with Educational Systems. *Proceedings of the 10th International Conference on Educational Data Mining, Edm 2017, Wuhan, Hubei, China, June 25-28, 2017* (2017).

[10] Markovify: 2015. *https://github.com/jsvine/markovify*.

[11] Miller, B.W. 2015. Using Reading Times and Eye-Movements to Measure Cognitive Engagement Using Reading Times and Eye-Movements to Measure Cognitive Engagement. 1520, (2015). DOI=https://doi.org/10.1080/00461520.2015.1004068.

[12] Motz, B. et al. 2019. The validity and utility of activity logs as a measure of student engagement. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (2019), 300–309.

[13] Ricci, F. et al. 2015. *Recommender systems: Introduction and Challenges*.

[14] Sahebi, S. and Brusilovsky, P. 2018. Student Performance Prediction by Discovering Inter-Activity Relations. *Proceedings of the 11th International Conference on Educational Data Mining (EDM)* (2018).