Starters and Finishers: Predicting Next Assignment Completion from Student Behavior During Math Problem Solving

Taylyn Hulse
Worcester Polytechnic Institute
trhulse@wpi.edu

Avery Harrison
Worcester Polytechnic Institute
aeharrison@wpi.edu

Korinn Ostrow
Worcester Polytechnic Institute
ksostrow@wpi.edu

Anthony Botelho
Worcester Polytechnic Institute
abotelho@wpi.edu

Neil Heffernan
Worcester Polytechnic Institute
nth@wpi.edu

ABSTRACT

A substantial amount of research has been conducted by the educational data mining community to track and model learning. Previous work in modeling student knowledge has focused on predicting student performance at the problem level. While informative, problem-to-problem predictions leave little time for interventions within the system and relatively no time for human interventions. As such, modeling student performance at higher levels, such as by assignment, may provide a better opportunity to develop and apply learning interventions preemptively to remedy gaps in student knowledge. We aim to identify assignment-level features that predict whether or a not a student will finish their next homework assignment once started. We employ logistic regression models to test which features best predict whether a student will be a "starter" or a "finisher" on the next assignment.

Keywords

Student Modeling, Mathematics Education, Classification

1. INTRODUCTION

Online learning environments, paired with educational data mining research, provide student and teacher supports for learning. These environments are able to map student learning and behavior to personalize content, offer scaffolding, and provide real time support such as informational or motivational messages [11], making them nearly as effective as one-on-one human tutoring [14]. While great strides have been made in refining online learning environments to optimize learning, past work has primarily focused on student learning at the problem-level or using problem-level features within learning systems [3, 4]. These models provide immediate feedback to students and personalize learning within a user's session. Less work has been done at higher granularities, such as modeling learning from assignment to assignment, to capture broader models of student learning.

While problem-level models of student learning are important, teachers more often care about higher level aspects of student learning such as whether students will be able to complete their homework assignment and if not, why? Building on previous work to track learning in online learning environments, as well as studies that have utilized similar data, we present our first attempt to build interpretable, predictive models of next-assignment completion. These models should indicate the best predictors of next-assignment completion to interpret reasons that a student might be a "starter" who is unable to finish the next homework assignment rather than a "finisher" who will complete the next assignment.

2. LITERATURE REVIEW

Online learning environments and tutoring systems contain rich data that can be applied to any level of fine- or coarse-grained research questions pertaining to student learning and behavior [9]. Using data from online systems, researchers have modeled student learning at various levels to better understand predictive behaviors, affective states, and system features of learning. From skill-level within problems [10], to problem-level [5], and across topics [2], the educational data mining community has tracked student learning and performance in a variety of contexts. Though steady progress has been made in predicting low-level behaviors, we can also leverage the prediction power of student logs to predict higher level behaviors and outcomes [1].

Research has also turned to predicting negative student behaviors and outcomes, such as student dropout rate. For instance, modeling student dropout rates has been a focus within massive open online courses (MOOCs) to understand why students complete online courses or dropout along the way [13, 16]. Similarly, attritional behavior in MOOCs has been modeled to identify and intervene with students who appear to be most likely to "stopout" [7]. While tutoring systems developed for K-12 curricula differ from MOOCs and secondary education settings, modeling dropout rates in online assignments would be beneficial at the K-12 level. Drawing from this work, we intend to develop predictive models of assignment dropout in an online learning environment to identify students likely to dropout of future assignments with time to intervene.

To accomplish this, we will use ASSISTments, a free, web-based tutoring system for K-college curricula that primarily features middle school mathematics content [8]. The current project will focus on "Skill Builders", which are pre-built problem sets that map onto content areas to provide students with practice on topics featured on standardized tests. Skill Builders present problems from a given content area in a randomized order and are designed to challenge a student until that student achieves content mastery.

Table 1. Descriptive statistics of predictive features of assignment completion

Predictors	*N*	Minimum	Maximum	Mean	Standard Deviation
Current Assignment Completed	71,523	0	1	0.93	0.26
Assignment Mastery: 3-4 Problems	52,896	0	1	0.69	0.46
Assignment Mastery: 5-8 Problems	16,471	0	1	0.21	0.41
Assignment Mastery: 9+ Problems	2,156	0	1	0.03	0.17
Assignment Dropout: 0 Problems	3,750	0	1	0.05	0.22
Assignment Dropout: <4 Problems	1,122	0	1	0.01	0.12
Assignment Dropout: 4+ Problems	789	0	1	0.01	0.10
Average Attempt Count per Problem	77,084	0	2.91	1.15	0.42
Average Hint Count per Problem	75,130	0	1	0.59	0.18

Under default settings, students must consecutively answer three problems correctly to achieve mastery status for the assignment.

Previous research using ASSISTments and Skill Builders has sought to detect and fine-tune features of ASSISTments to be most beneficial to students and educators in practice. Most notably, a recent efficacy trial found that students who used ASSISTments throughout the school year performed better on an end-of-year standardized test than their counterparts who continued to use pen-and-paper homework assignments [12]. The researchers theorized that the difference in achievement may have been attributed to teacher reports generated in ASSISTments that provide teachers with information and timely homework feedback for students. This suggests that providing predictive feedback to teachers may better prepare them to provide additional support needed before difficult assignments. Predictive reports on future assignments would enable teachers to target specific content and assist students beforehand.

3. CURRENT PROJECT

We build off previous work that has modeled next problem performance at the student level, as well as approaches to modeling dropout rates in secondary education settings, to build predictive models of completion at the assignment level. While problem-level predictive models are limited to immediate, online-tutor interventions, making predictions about student behavior at the assignment level would allocate time for teacher interventions.

We use logistic regression to predict whether a student will be a "starter" or a "finisher" on the next ASSISTments assignment. Specifically, if a student opens the next assignment, will he be able to complete that assignment, achieving content mastery? We use the predictive model to identify features most predictive of next assignment completion. We focus first on using current completion features to predict next assignment completion and then progress to the predictive abilities of student behavior metrics within the tutor. This work is guided by the following three research questions: *How well does student completion on the current assignment predict completion on the next assignment? Does the number of completed problems matter when predicting next assignment completion? Do student behaviors within the tutor predict next assignment completion above current assignment completion features?*

4. DATA AND PREPROCESSING

We used publically available ASSISTments data [6] from Skill Builders during the 2016-2017 school year. Skill Builders were restricted to mathematics content and mastery parameters of 3+ correctly solved, consecutive problems. Using descriptive statistics, outliers were trimmed from any variables with a skewness statistic > |3|. After cleaning, the dataset contained 77,200 cases of assignments started or completed by 9,231 students in grades 3-12 across 5,143 unique Skill Builder assignments. Students completed the current assignment 92.65% of the time and completed the next assignment, our outcome variable, in 88.08% of cases. We constructed nine features from the dataset using assignment-level and problem-level variables aggregated by student at the assignment level (Table 1). In addition to current assignment completion, we selected assignment mastery speed, average attempt count per problem, and average hint count per problem as features. Assignment mastery speed is based on the number of problems students solve to fulfill the requirements of the skill builder assignment and "master" the content.

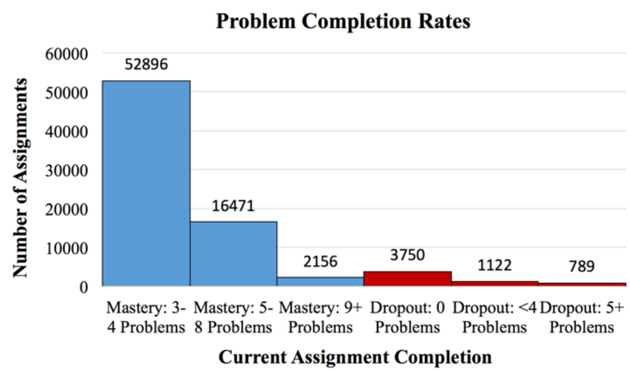

Figure 1. Rate of problem completion within complete and incomplete assignments.

We created three categories of mastery speed based on a method previously used [17]. We extended the method to account for the duration of student persistence prior to dropping out of the assignment, which we use as a dropout rate (Figure 1). To find a predictive model of future homework completion that optimized simplicity and fit, we created three models through logistic regression to compare and evaluate. We began with the completion model, with only current assignment completion as a predictor of next assignment completion. Then, we delved into student behavior within current assignments to build the binned completion model with completion status and number of problems completed within the assignment as predictors. Lastly, the binned completion and student features model used the binned completion predictors, the average hint use per problem, and average attempt count per problem as features. We used optimized thresholds for kappa and accuracy since our models were biased towards the majority class. By optimizing the threshold, we removed this bias. The optimized threshold was calculated only on training data. Five-fold cross validation was applied at the student level for each model.

5. RESULTS

Table 2 overviews the performance for each of the three models. Each model performs above chance (AUC > 0.50; kappa > 0.00) though performance slightly decreases (kappa, F_1, Accuracy) and AUC slightly increases as the models increase in complexity.

Table 2. Model comparison across performance metrics.

Model	AUC	Kappa	F_1	Accuracy
Completion	0.617	0.281	93.04	87.42
Mastery & Dropout	0.632	0.258	91.82	85.44
Mastery, Dropout, & Student Features	0.641	0.250	91.41	84.79

Model 1: Completion Model

The first model was a logistic regression that used completion on the current assignment to predict completion on the next assignment. This served as a base rate model that simply consists of completion as the predictor for future completion. For the Completion Model (kappa=0.281, AUC=0.617), completion was a positive predictor of next assignment completion, $\boldsymbol{b}$= 2.10, $z(77,200) = 71.05$, $p < 0.01$.

Model 2: Mastery and Dropout Model

The second model expanded on the base rate model by categorizing completeness and number of problems solved. We applied logistic regression using completeness and incompleteness categories as features to predict next assignment completion. Mastery speeds were significant, positive predictors of next assignment completion while dropout rates were not statistically significant (Table 3).

Table 3. Logistic regression with Mastery Speeds and Dropout Rates.

Feature	b	B	SE	z value	p value
Intercept	0.51	2.13	0.00	175.35	0.00
Mastery (3-4)	1.83	0.85	0.52	3.53	0.00
Mastery (5-8)	1.69	0.69	0.52	3.27	0.00
Mastery (9+)	1.26	0.21	0.52	2.43	0.02
Dropout (0)	-0.46	-0.10	0.52	-0.88	0.38
Dropout (1-3)	-0.10	-0.01	0.52	-0.20	0.84
Dropout (4+)	-0.05	-0.01	0.52	-0.10	0.92

Model 3: Mastery, Dropout and Student Features Model

The final model incorporates two student-related features: hints and attempts. We applied logistic regression using completeness and incompleteness categories, as well as the two student features, to predict next assignment completion. Table 4 shows that in addition to mastery speeds, average attempts was also a significant, negative predictor of next assignment completion. This suggests that more attempts in the current assignment results in a lower likelihood of finishing the next assignment.

Table 4. Logistic regression with Mastery Speed and student features.

Feature	b	B	SE	z value	p value
Intercept	0.67	2.13	0.00	3.53	0.00
Mastery (3-4)	1.82	0.84	0.52	3.52	0.00
Mastery (5-8)	1.75	0.72	0.52	3.38	0.00
Ave Attempt Count	-1.40	-0.06	0.04	-3.92	0.00
Mastery (9+)	1.34	0.22	0.52	2.57	0.01
Dropout (0)	-0.58	-0.13	0.52	-1.12	0.26
Ave Hint Count	-0.02	-0.00	0.07	-0.29	0.77
Dropout (4+)	0.07	0.00	0.52	0.14	0.89
Dropout (1-3)	-0.06	-0.00	0.52	-0.11	0.91

6. DISCUSSION

The models presented in this paper predict next assignment completion, which compared to predicting next problem completion, could provide more timely and practical information about student learning that could be applied through teacher intervention. Though most of the performance measures slightly decreased as more features were added, all three models performed similarly as a whole. Out of the student features we analyzed, completeness on the current assignment is the most prominent predictor of completion on the next assignment, which answered our first research question. This suggests that a simple model using only completeness as a predictor would be appropriate for uses such as creating an alert in a teacher dashboard to signal when students may not complete their next assignment.

That said, Models 2 and 3 add a more detailed explanation regarding how completeness breaks down and what other features may contribute to next assignment completion. We answered our second research question with Model 2 by categorizing completeness into mastery speed and dropout rate based on how many problems students completed. Though dropout rates were not significant predictors, higher mastery speeds in the current assignment increased the likelihood of students completing their next assignment. This is to be expected, as students who complete their current assignment in fewer problems are generally performing more efficiently, which may be suggestive of future performance due to underlying knowledge levels, motivational and behavioral tendencies, or other student-level characteristics.

To answer our third research question, Model 3 incorporated within-problem student behaviors, average attempts made, and average hints used per problem. The number of attempts was a negative predictor of next assignment completion, suggesting that students who make more attempts per problem are less likely to complete the next assignment. It seems that lower performing students (based on those who finish the assignment in more problems and take more attempts per problem) are less likely to complete the next assignment. While this is not a surprising finding, it brings us closer to teasing apart the integral facets of students' knowledge, behavior, and interaction with the system which leads them to become starters or finishers on their next assignment.

Though the models presented herein serve as a valuable first step towards understanding student and system contributions to student learning at a higher level, we acknowledge limitations in our dataset and analyses. We had a limited sample of incomplete current assignment behavior, as the majority of next-assignments were completed. When binning into completeness and incompleteness categories, the majority of data fell in the first two categories of completeness (3-4 problems solved 69%, 5-8 problems solved 21%), resulting in disproportionate pools for other categories (≤ 5%). These characteristics of the data made it more difficult to predict next assignment incompletion and instead, bias towards the larger current completeness categories. These models also only included measures of completeness (mastery speed and dropout rate) with a small selection of student behavior features. We started with simple models to identify the most logical predictive features. Moving forward, our analyses will include more holistic models of learning with parameters based on student and assignment features. Prior student knowledge and exposure, as well as problem content and difficulty, could be logical predictors of assignment completion and student learning. As such, future work will assess the generalizability of our three models while working to extend our predictive capacity further through the addition of new parameters.

We also plan to extend our models to predict next assignment performance. Similar to how we binned completeness to predict next assignment completeness, we can also use binary or binned correctness (partial credit) [15] to predict next assignment performance. This will expand our scope on higher level learning modeling and has the potential to provide more useful feedback to teachers when deciding on which content to review to increase the number of homework "finishers."

7. CONCLUSION

We have presented three predictive models of next assignment completion in ASSISTments that vary in complexity but perform comparably to one another. By modeling student performance at the assignment level, we were able to broadly model student behavior to predict whether students will be a homework "starter" or "finisher" on the next assignment. This approach to student modeling could serve as a foundation for a predictive teacher feedback tool within ASSISTments to increase teacher ability to target key content areas in class to increase the likelihood of all students being assignment finishers.

8. ACKNOWLEDGMENTS

We gratefully acknowledge funding from multiple NSF grants (1440753, 1252297, 1109483, 1316736, 1535428, 1031398), the U.S. Dept. of Ed. (R305A120125, R305C100024, P200A120238 and GAANN), and ONR.

9. REFERENCES

[1] Beck, J. E., and Woolf, B. P. 2000. High-level student modeling with machine learning. In *International Conference on Intelligent Tutoring Systems*. Springer, Berlin, Heidelberg, 584-593.

[2] Botelho, A. F., Adjei, S. A., and Heffernan, N. T. 2016. Modeling interactions across skills: A method to construct and compare models predicting the existence of skill relationships. In *Proceedings of the 9th International Conference on Educational Data Mining*, 292-297.

[3] Cen, H., Koedinger, K., and Junker, B. 2006. Learning factors analysis–a general method for cognitive model evaluation and improvement. In *International Conference on Intelligent Tutoring Systems*. Springer, Berlin, Heidelberg, 164-175.

[4] Corbett, A. T., Anderson, J. R., and O'Brien, A. T. 1995. Student modeling in the ACT programming tutor. *Cognitively diagnostic assessment*, 19-41.

[5] Duong, H.D., Zhu, L., Wang, Y., and Heffernan, N.T. 2013. A Prediction Model Uses the Sequence of Attempts and Hints to Better Predict Knowledge: Better to Attempt the Problem First, Rather Than Ask for a Hint. In S. D'Mello, R. Calvo, & A. Olney (Eds.) *Proceedings of the 6th Int Conf on EDM*, 316-317.

[6] Harrison, A. 2018. EDMSubmission2018 Data. Retrieved March 7, 2018 from http://tiny.cc/EDMSubmission2018.

[7] He, J., Bailey, J., Rubinstein, B. I. P., and Zhang, R. 2015. Identifying at-risk students in massive open online courses. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[8] Heffernan, N. T., and Heffernan, C. L. 2014. The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24, 470-497.

[9] Pardos, Z. A. 2017. Big data in education and the models that love them. *Current Opinion in Behavioral Sciences*, 18, 107-113.

[10] Roll, I., Baker, R. S. J. d., Aleven, V., and Koedinger, K. R. 2014. On the benefits of seeking (and avoiding) help in online problem-solving environments. *The Journal of the Learning Sciences*, 23, 537-560.

[11] Romero, C., and Ventura, S. 2010. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics-- Part C: Applications and Reviews*, 40, 6, 601-61

[12] Roschelle, J., Feng, M., Murphy, R. F., and Mason, C. A. 2016. Online mathematics homework increases student achievement. *AERA Open*, 2, 4, 1-12.

[13] Taylor, C., Veeramachaneni, K., and O'Reilly, U.-M. 2014. Likely to stop? Predicting stopout in massive open online courses. *arXiv preprint arXiv:1408.3382*, 273–275.

[14] VanLehn, K. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46, 4, 197-221.

[15] Wang, Y., Ostrow, K., Beck, J., and Heffernan, N. 2016. Enhancing the Efficiency and Reliability of Group Differentiation through Partial Credit. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge*. ACM, 454-458.

[16] Xing, W., Chen, X., Stein, J., and Marcinkowski, M. 2016. Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, 58, 119-129.

[17] Xiong, X., Li, S., and Beck, J. E. (2013, May). Will You Get It Right Next Week: Predict Delayed Performance in Enhanced ITS Mastery Cycle. In *FLAIRS Conference* (Vol. 2013).