Identifying Beneficial Learning Behaviors from Large-Scale Interaction Data

Miruna Cristus[(✉)], Oscar Täckström, Lingyi Tan, and Valentino Pacifici

Sana Labs, Stockholm, Sweden
{miruna,oscar,lingyi,valentino}@sanalabs.com

Abstract. Understanding the effect of learning behavior is fundamental to improving learning outcomes. In this paper, we perform a behavioral analysis based on data from a large high-stakes exam preparation platform. By measuring the importance of a set of candidate learning behaviors in predicting final exam outcomes, we identify a suite of beneficial behaviors. In particular, we find that *breadth* (wide coverage of content per week) and *intensity* together with *consistency* (frequent and equal-length practice for a limited period) are most predictive of final exam success rate, among eleven studied behaviors.

Keywords: Learning behavior · Test preparation · Educational data mining

1 Introduction

Sana Labs provides personalized learning through partnership with the world's largest learning content providers. Understanding which learning behaviors lead to successful outcomes is a key focus of our research and development. The combination of online education and machine learning makes it possible to study learners' behaviors and outcomes at an unprecedented scale [17]. This combination holds promise as a way to identify key learning behaviors that can be highly beneficial or detrimental to learning outcomes [11,13]. They can in turn be used to make personalized learning more effective and enjoyable [9]. In this paper, we focus on a behavioral analysis based on data from a large high-stakes exam preparation platform, where Sana provides review sessions that help students bridge their knowledge gaps and retain their acquired knowledge.

2 Online Learning Platform

Sana powers several features on a large-scale online exam preparation platform, where students need to go through a large amount of content, typically over the course of several months. Thus, students need to actively make complex decisions about their learning schedule and what material to cover at a given point in time, to optimize learning outcomes and maximize their exam results.

To guide the students in this decision making process, Sana powers adaptive review sessions tailored to the needs of each student. In these sessions, Sana predicts the current and future knowledge gaps of each student [4,15,16, *inter alia*], and recommends the most appropriate content for remediation; previously seen content is also resurfaced at the optimal time intervals in line with spaced repetition to foster long-term knowledge retention [5,8,10, *inter alia*].

3 Analyzing Learning Behavior

While Sana's recommendation algorithms are based on established research on human learning strategies, it is important to understand student behavior in context to further tailor our recommendations for different use cases. Student interactions constitute a rich source of data from which to derive such insights.

3.1 Data

We collected data from each student interacting with learning material on the platform.[1] This interaction data was enriched with the final exam outcome of each student (pass/fail) to form the core data of our analysis.

To focus on interaction events related only to the observed exam outcome, we disregarded all events registered prior to a break of at least 30 days of studying. We further excluded infrequent users of the platform.[2] After filtering, we obtained a group of 6631 students, totaling over 35 million events over a period of 7 months.

3.2 Method

We defined features to capture different facets of student behavior that were hypothesized to have an impact on learning outcomes. Each behavior was encoded as a numerical feature and the impact of each feature was assessed with the Random Forest permutation importance measure [1].[3]

Specifically, we used Scikit-learn [14] to train a Random Forest classifier to predict the exam outcome (pass/fail) from the full set of features, using a 75–25% train–test split. We tuned the following hyperparameters (optimal value in parenthesis) on the training set: the maximum fraction of features to be considered for a split (0.2), the number of trees (400), and the maximum depth of a tree (50). The optimal setting resulted in 0.93 AUC [6] on the test set.

We repeated the analysis on a smaller group of 1158 students with similar practice frequency, in order to control for the effect of time spent on the platform.

---

[1] Including responses to multiple choice questions, viewing of theory or instructional video, along with outcome, when applicable.

[2] Infrequent users are defined as: students who did not have at least one interaction with each major topic of the exam, students who had used the platform for less than a week, and students who had covered less than 20% of the content on the platform.

[3] This has been shown to be more reliable than impurity-based measures [18].

We obtained 0.91 AUC on the test set, with the optimal hyperparameters being a split fraction of 0.4, 600 trees, and unlimited maximum depth.

Finally, to understand the direction of these effects, we performed t-tests on the averages of each feature between the groups of passing and failing students.
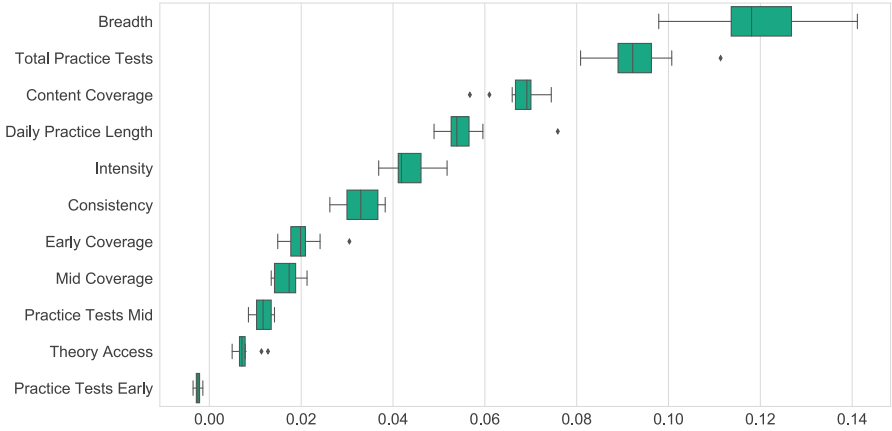
Fig. 1. Permutation importance of evaluated features. See Sect. 3.2–3.3 for details.

3.3   Results

Figure 1 shows the permutation importance of each of the analyzed features. The most effective behaviors identified, all of which are found to have positive effects, are listed below in order of importance (due to space restrictions, we are not able to describe all behaviors in detail):

– `Breadth:` the number of topics (as defined by the curriculum) the student engages with on an average week, without necessarily having completed them.
– `Total Practice Tests:` the number of practice tests taken (a practice test fully mimics the structure and timing of the real exam).
– `Content Coverage:` the percentage of unique content elements (exercises) covered across the entire study path.
– `Daily Practice Length:` the average number of content elements the student interacts with per day.
– `Intensity:` the fraction of days the student has been active on the platform out of the total number of days registered.
– `Consistency:` the reciprocal of one standard deviation of the daily practice length.

Our finding on the importance of breadth and content coverage is consistent with research on interleaved learning [3]. In the case of exam preparation, a breadth-first approach could potentially help by familiarizing the students with the structure of the content or by making associations between different topics.

Additionally, in terms of content type, we found that completing the available practice tests has a positive impact, however with diminishing returns.

Looking at daily practice length, we found that students with the highest total amount of practice do not necessarily have the best exam outcomes: these students often practice less daily, but keep studying for a longer period. It seems that learning intensity matters more than the sheer number of days of studying.

Once the effects of frequency and timing are isolated, consistency becomes important: students who spend roughly the same amount in each practice have better outcomes.[4] These finding may reflect the importance of appropriately spaced repetition on knowledge retention.

4   Limitations

Due to the limitations of data collection, we could not consider factors that are difficult or impossible to measure from the available data, for example, the knowledge state of a student prior to using the platform, their use of external resources, demographics [7], socioeconomic status [2], or motivation [12]. A causal model is also out of scope of the current study for the same reasons. Collecting relevant external information and isolating potential confounding variables would allow us to better identify beneficial learning behaviors that are addressable within the platform. Finally, we hope to verify in future studies whether the present findings are applicable to other learning platform and subjects as well.

5   Conclusion

As online learning platforms are becoming increasingly popular, there is a rising need to tailor both learning paths and content to maximize learning outcomes. In addition to personalization and adaptivity, understanding the effect of overall learning behavior is an important aspect of designing effective strategies, content organization and user experience.

Two key beneficial behaviors have been identified in this study:

1. Cover as much of the content as possible, through a breadth-first approach (interleaved learning).
2. Practice frequently and consistently (i.e. for a similar amount of time in each session).

Our findings validate the algorithms currently employed by Sana for personalized review: these sessions bring out the best content from all topics, facilitating breadth. By predicting knowledge gaps in a topic and surfacing unseen but related material from that topic, the sessions also promote content coverage. We believe that these findings provide a basis for further improvements to recommendation strategies to promote optimal learning behavior.

---

[4] The *consistency* feature has an importance score of 0.1 on the smaller dataset of 1158 students described in Sect. 3.2.

References

1. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
2. Broer, M., Bai, Y., Fonseca, F.: A Review of the Literature on Socioeconomic Status and Educational Achievement, pp. 7–17. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11991-1_2
3. Carvalho, P.F., Goldstone, R.L.: When does interleaving practice improve learning? In: Dunlosky, J., Rawson, K.A. (eds.) The Cambridge Handbook of Cognition and Education. Cambridge Handbooks in Psychology, pp. 411–436. Cambridge University Press, Cambridge (2019)
4. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modelling the acquisition of procedural knowledge. User Model. User-Adapted Interact. **4**(4), 253–278 (1995)
5. Dempster, F.: Spacing effects and their implications for theory and practice. Educ. Psychol. Rev. **1**, 309–330 (1989)
6. Hanley, J., McNeil, B.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology **143**(1), 29–36 (1982)
7. Islam, M.A., Rahim, A., Tan, C., Hasina, M.: Effect of demographic factors on e-learning effectiveness in a higher learning institution in malaysia. Int. Educ. Stud. **4**(1), 112–121 (2011)
8. Karpicke, J., Roediger, H.: Repeated retrieval during learning is the key to long-term retention. J. Memory Lang. **57**, 151–162 (2007)
9. Koedinger, K.R., Brunskill, E., Baker, R.S., McLaughlin, E.A., Stamper, J.: New potentials for data-driven intelligent tutoring system development and optimization. AI Magazine **34**(3), 27–41 (2013)
10. Melton, A.W.: The situation with respect to the spacing of repetitions and memory. J. Verbal Learn. Verbal Behav. **9**(5), 1 (1970)
11. Morris, L.V., Finnegan, C., Wu, S.S.: Tracking student behavior persistence and achievement in online courses. Internet Higher Educ. **8**(3), 221–231 (2005)
12. Orhan-Özen, S.: The effect of motivation on student achievement. Factors Effect. Student Achieve.: Meta-Anal. Empirical Stud. **5**, 35–56 (2017)
13. Osmanbegovic, E., Suljic, M.: Data mining approach for predicting student performance. Econ. Rev. J. Econ. Bus. **10**(1), 3–12 (2012)
14. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
15. Piech, C., et al.: Deep knowledge tracing. In: Neural Information Processing Systems (2015)
16. Rasch, G.: Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche (1960)
17. Romero, C., Ventura, S.: Educational data mining: A review of the state of the art. IEEE Trans. Syst. Man, Cybern. Part C (Applications and Reviews) 40(6), 601–618 (2010)
18. Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T.: Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinform. **8**(1), 25 (2007)