Modeling Students' Behavior Using Sequential Patterns to Predict Their Performance

Mehrdad Mirzaei[(✉)] and Shaghayegh Sahebi

University at Albany – SUNY, Albany, NY 12203, USA
{mmirzaei,ssahebi}@albany.edu

Abstract. Online learning environments generate educational data that can be used to model students' behavior and predict their performance. In online learning environments, in which students are free to choose their next activity, various factors such as time spent on individual tasks and the choice of next learning material may impact students' performance. The main goal of this research is to enhance student learning by modeling students' behavior and testing whether these behavioral patterns correlate with their performance. Using sequential pattern mining methods, we will identify the most frequent patterns in students' online learning activities and test whether/which patterns correlate with higher or lower performance. By identifying which student behavioral patterns correlate with higher or lower performance, this study has the potential to inform redesign of online learning platforms and study guidelines that help students learn more and perform better.

Keywords: Sequential pattern mining · Student performance · Matrix factorization

1 Introduction

Previous research has shown that given the choice, students may take on repetitive and non-productive behaviors in solving problems [2]. Also, it has been shown that students can be grouped into various clusters according to their studying patterns, but these clusters are not directly related to students' performance. Although insightful, the past research has mostly focused on one type of learning material in analyzing student behaviors. Moreover, these studies have not drawn clear distinctions between non-productive and productive behavioral patterns. The goal of this research is to examine different factors in forming students' behavior and to draw conclusions that can improve students' learning and performance. The performance is defined as the grades of the students or learning gain in a course. Both educators and students can take advantage of these findings. Students can adopt patterns that are useful for learning and avoid patterns that may inhibit learning. Educators may also be able to use this behavioral data to identify struggling students earlier in a course and intervene

to improve their behavior before important assessments. In this work, we propose a general approach to find frequent patterns of students' behavior using online educational platforms.

2   Related Work

A group of studies aim to group students based on their behaviors. Researchers have employed data mining algorithms such as clustering in these works. In [1], they have used two different approaches to extract frequent sequences of actions in a collaborative learning environment to distinguish high achieving students from low achieving ones in small groups. In [2] patterns of student behavior with parameterized exercises are modeled and analyzed. In this work, micro patterns are extracted using a frequent mining algorithm and are used to build macro patterns to cluster students in groups with similar patterns. We extend this method by combining students' behavioral patterns on multiple types of learning materials (e.g., worked examples and problems). A number of researches have used matrix factorization to find latent factors as patterns in student data. In [5] non-negative matrix factorization is used to cluster gene expressions and find factorization with the same gene expression profiles by estimating offsets for individual genes. In [4] a method is proposed to identify common and discriminative topics among a set of given documents according to their keywords. We will adapt and extend this work to distinguish between useful and harmful patterns of students' interactions with online learning material.

3   Research Questions

Various factors may impact student performance, such as time spent on individual tasks and how a student chooses to engage with the online platform. The goal of this research is to examine such factors and draw conclusions that could improve the efficiency of the students and efficacy of online learning tools. Student activities and decision-making while functioning in a computer-based learning environment are underutilized and could be used to guide students with effective patterns in studying. The information obtained in this analysis will be used to answer the following questions:

Question 1: Do individual students exhibit stable behavioral patterns in their work with learning content, or does their learning approach depend on factors, such as time of the semester or learning material complexity?

Question 2: Are student behavioral patterns associated with their learning performance?

Question 3: How accurately can we discriminate between students' productive behavioral patterns vs. the non-productive ones?

4   Previous Work

We extracted students' behavior patterns while interacting with an online learning environment. These patterns are consecutive actions in the sequence of students' activities. To extract the patterns, we used a sequential pattern mining method (CM-SPAM) [3]. Then extracted patterns are used to build a vector for each student that contains the frequency of all different patterns and model the student's behavior. Clustering the pattern vectors, we discovered three clusters with distinct patterns. We call these clusters: "Confirmers", "Thinkers", and "Readers". "Confirmers" mostly tend to confirm their success by repeating to solve a problem again and again. "Thinkers" are the group that achieve success after some failed attempts, and have longer activities than other student groups. "Readers" usually spend more time on reading the worked examples.

5   Future Work

To cluster students' behavioral patterns, we propose a matrix factorization method (MF), extending the work by Kim et. al [4]. Having two sets of documents, the model in [4] finds topics from each document set, among which some topics are common between the two document sets and the rest of the topics are different between them. Another possible direction is to embed social networks of students in online courses to enhance the performance prediction [6].

5.1   Proposed Method

Our proposed method is based on the model in [4]. We will use MF to find common patterns and distinct patterns between two groups of students. To have similar patterns in each group, we will minimize the differences between patterns. Students' performance will be used to distinguish them in two different groups: high and low-performers. The common patterns will be considered as ordinary patterns that represent students in both groups, but distinct patterns in each group are specific to that group.

5.2   Problem Formulation

We have the pattern vectors of high-performance students and low-performance students extracted in X_1 and X_2. The purpose is to find k pattern clusters such that k_c of pattern cluster are common between two groups of students and k_d of them are different between students. So there are two matrices that should be decomposed:

$$X_1 \approx W_1 H_1^T \quad X_2 \approx W_2 H_2^T \tag{1}$$

We split W and H to have common and discriminative pattern clusters. The matrices are split in this way:

$$W_1 = [W_{1,c} \quad W_{1,d}], \quad W_2 = [W_{2,c} \quad W_{2,d}] \tag{2}$$

$$H_1 = [H_{1,c} \quad H_{1,d}], \quad H_2 = [H_{2,c} \quad H_{2,d}] \tag{3}$$

$W_{1,c}$ and $W_{2,c}$ are similar pattern clusters and $W_{1,d}$ and $W_{2,d}$ are distinct ones. We should define functions to calculate how common or distinctive the patterns are and add them to the formulation.

The model in [4] is proposed to find common and discriminative topics in two document sets. We replace documents and words with pattern vectors and patterns respectively to find patterns that are different between two groups. Moreover, we use pattern similarity matrix in the model, since we expect to have similar patterns in each group.

References

1. Martinez, R., Yacef, K., Kay, J., Al-Qaraghuli, A., Kharrufa, A.: Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In: Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011), pp. 111–120 (2011)
2. Guerra, J., Sahebi, S., Brusilovsky, P., Lin, Y.R.: The problem solving genome: analyzing sequential patterns of student work with parameterized exercises. In: Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), pp. 153–160 (2014)
3. Fournier-Viger, P., Gomariz, A., Campos, M., Thomas, R.: Fast vertical mining of sequential patterns using co-occurrence information. In: Tseng, V.S., Ho, T.B., Zhou, Z.-H., Chen, A.L.P., Kao, H.-Y. (eds.) PAKDD 2014. LNCS (LNAI), vol. 8443, pp. 40–52. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06608-0_4
4. Kim, H., Choo, J., Kim, J., Reddy, C.K., Park, H.: Simultaneous discovery of common and discriminative topics via joint non-negative matrix factorization. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 15), pp. 567–576 (2015)
5. Badea, L.: Extracting gene expression profiles common to colon and pancreatic adenocarcinoma using simultaneous nonnegative matrix factorization. In: Pacific Symposium on Biocomputing, pp. 267–278 (2008)
6. Doan, T., Lim, E.: Modeling location-based social network data with area attraction and neighborhood competition. Data Min. Knowl. Disc. **33**, 58–95 (2019)