So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Exploring Temporal Study Patterns in eBook-based Learning

Gökhan AKÇAPINAR[a*], Mohammad Nehal HASNINE[b], Rwitajit MAJUMDAR[c], Mei-Rong Alice CHEN[c], Brendan FLANAGAN[c] & Hiroaki OGATA[c]

[a] *Department of Computer Education & Instructional Technology, Hacettepe University, Turkey*

[b] *Research Center for Computing and Multimedia Studies, Hosei University, Japan*

[c] *Academic Center for Computing and Media Studies, Kyoto University, Japan*

*gokhana@hacettepe.edu.tr

Abstract: In this study, approximately 2 million click-stream data of 1346 students in the eBook platform were analyzed aiming to explore the temporal study patterns of the students followed during the lectures. The data used in the study collected from Kyushu University, Japan with the help of a digital textbook reader called BookRoll. Students used BookRoll for reading learning materials in and out of the class. To analyze the data we first, converted reading sessions into the sequence data which represents student's weekly reading behavior, then we clustered students based on their study patterns. Our results revealed that three groups of students can be extracted with similar study patterns. Most of the students in Cluster 1 viewed the learning materials only during the class, without previewing and reviewing them. Students in Cluster 2 previewed the learning materials before the class, viewed learning materials during the class, and also reviewed after the class. Students in Cluster 3 viewed the learning materials during the class in the beginning but they became inactive over the period of time (week by week). Our study also showed how learning analytics can be used to understand students' study patterns which are difficult to do with self-report data. These results can help instructors while designing their courses.

Keywords: sequence mining, clustering, reading logs, eBook, educational data mining, learning analytics

1. Introduction

Pre-class reading and after class repetition are crucial for students in higher education to understand the subject being taught in class and transfer their knowledge to other domains. Typically, the purpose of the pre-class reading assignment is to expose students to background knowledge that will be useful in an upcoming class discussion or to introduce a topic that will be presented more directly by the instructor (Tomasek, 2009). Completing this pre-class work helps students to be more engaged in the in-class learning process (Ripley, 2007). In another study, researchers found that students' exam performance significantly improves by nearly 12% in the flipped-format course, due in part to students interacting with course material in a more timely and accurate manner (Gross, Pietri, Anderson, Moyano-Camihort, & Graham, 2015). Although pre-class reading has many advantages on students' learning, studies have shown that students frequently do not read their textbooks before the class (Ruscio, 2001), moreover, most of the students are not reading the textbook at all (Lieu, Wong, Asefirad, & Shaffer, 2017).

The advancement of online learning technologies such as ITSs, digital eBook systems, MOOCs, etc. opened the doors to the learners to gain new knowledge. Regardless of learners' knowledge, motivation, or engagement level, learners get the flexibility to engage with the learning systems by navigating through various learning materials (Boroujeni & Dillenbourg, 2018). As a result, learners leave various study patterns which may tell a lot about learners' learning processes. Analyzing study patterns is important and has gained significant attention of educational researchers because hidden in study patterns can provide many important insights about learners and learning environments.

In the present study, we focus on students' temporal study patterns by analyzing a large amount of click-stream data collected from university students related to their pre-class, in-class, and after-class reading behaviors. As mention by Knight, Friend Wise, and Chen (2017) learning is a process that occurs over time and online learning tools generate fine-grained data regarding the temporal aspect of learning. However, the temporal aspect of learning is often neglected while analyzing learner data. According to Chen, Knight, and Wise (2018) temporal study pattern has two features. The first feature is related to the passage of time (how long, how often students engage). The second feature refers to the sequential order in which these activities take place (Molenaar & Järvelä, 2014). Students' instructional conditions such as learning design influence both of the features. Therefore, the analysis of temporal patterns in the clickstream data tracking student actions is essential to expose the insights of students' learning processes.

2. eBook Data Analysis in Higher Education

Students' reading logs previously used to predict their end of year academic performances (Hasnine et al., 2018), developing educational early-warning systems for at-risk students (Akçapınar, Hasnine, Majumdar, Flanagan, & Ogata, 2019a), detecting off-task behaviors during the classroom teaching (Akçapınar, Hasnine, Majumdar, Flanagan, & Ogata, 2019b), modeling students' level of knowledge (Flanagan, Majumdar, Akçapınar, Wang, & Ogata, 2019), and recently for understanding students' approaches to learning (Akçapınar, Chen, Majumdar, Flanagan, & Ogata, 2020). In this study, we focus on the temporal aspect of reading logs and try to understand students' study patterns on a large amount of click-stream data.

3. Method

3.1 Dataset

As the data source, a publicly available dataset collected from the students of Kyushu University was used. Reading logs were collected from the first-year students of the Faculty of Arts and Sciences. Students were registered to the Information Science Course. It was a face-to-face course supported by digital technologies. Weekly course materials related to information science were shared with the students with the help of a digital eBook system (BookRoll).

As mentioned by Ogata et al. (2017) BookRoll is a developed system that allows viewing digital materials used in lectures. It is an online environment that allows teachers to upload content in *pdf* format which the student can browse anytime and anywhere from a web browser in their personal devices (e.g. laptop, mobile devices, etc.). In BookRoll, there are features like bookmarks, markers, memo functions, which the students can use for learning. Overall data includes 1,914,680 rows click-stream comes from 10 different courses that use the same set of instructional design and learning materials. In each course, an instructor used 8 different learning materials and the overall length of the course is 8 weeks. The overview of the data is given in Table 1. Since all of the courses have the same structure, we merged all the data and treated it as a single course.

Table 1. *Overview of the data*

Course ID	Students	Total Event	Total Reading Session	In Class Reading Session
24a65f29b6	137	164154	2175	1590
34451e8c77	128	139976	1792	1479
39a67f80f4	131	207921	2332	1627
60ab104927	113	248599	2346	1525
65bb6224af	129	161094	1840	1461
6b1900c56c	118	263284	2830	1809
792efa2c1b	138	190070	2224	1643
86066cba6d	142	199541	2059	1673
9a683161f5	170	192844	2216	1842
dbed6c966a	140	147197	1914	1644
Total	1346	1914680	21728	16293

3.2 Data Preprocessing

The provided click-stream data has the following fields: *userid* (anonymized student id), *contentsid* (the id of the eBook that is being read), *operationname* (the action that was done, e.g. open, close, next, previous, jump, add marker, add bookmark, etc.), *pageno* (the current page where the action was performed), *marker* (the reason for the marker added to a page, e.g. important, difficult), *memo_length* (the length of the memo that was written on the page), *devicecode* (the type of device used to view BookRoll, e.g. mobile, pc), and *eventtime* (the timestamp of when the event occurred).

During the preprocessing, we assigned *session id* for each log by separating logs into the reading sessions. While dividing logs into the reading sessions, we considered the *OPEN* event as a starting point of the new reading session. In other words, we coded every sequence of logs as a reading session which starts with the *OPEN* event. By using the information provided in metadata files, we identified material used in each week of the course and assigned lecture id for each course.

Extracting Study Sequences: At the final stages of data preprocessing, we extracted students' learning states for each learning material according to states defined in Table 2. From the click-stream data, it can be understood that all learning materials are uploaded to the system at the beginning of the semester. All activities before the class counted as *Preview* activities and all activities after the class counted as *Review* activities. In other words, if a student opens the material anytime before the class, we considered it as a *Preview* activity. If a student opens the material during the class we considered it as a *Class* activity. If a student opens the material anytime after the class we considered it as a *Review* activity. Based on this information we labeled every single reading session extracted during the preprocessing with one of the states given in Table 2. At the end of this process, each student was given a state for each content. As a result, click-stream data converted into reading sessions and study sequences. Example study sequence for a student could be like this: *C1:Preview -> C2:Class -> C3:Inactive -> C4:Class -> C5:Review -> C6:Preview+Class -> C7:Class -> C8:Preview+Class*. The description of all types of activities can be seen in Table 2.

Table 2. *Students' Study States and Descriptions*

State	Description
Preview	Activity detected only during the preview period but not during the class and review period
Review	Activity detected only review period
Preview+Review	Activity detected during the preview and review period but not during the class
Class	Activity detected only during the class period
Preview+Class	Activity detected both during the preview period and the class period
Class+Review	Activity detected both during the class and review period
Preview+Class+Review	Activity detected during the preview, class, and review period
Inactive	No activity detected for a learning material

3.3 Data Analysis

We analyzed sequential data extracted during the preprocessing to understand students' study patterns across the semester. First, we looked at the general patterns across all data such as the distribution of states for each content. Second, we clustered students based on the similarity of their study patterns. This helped us to see common study patterns followed by the students. Agglomerative Hierarchical Clustering based on Ward's algorithm (Alexis Gabadinho, Ritschard, Studer, & Müller, 2009) was used to group students with similar study patterns. Optimal matching distance (OM distance) was used as a similarity calculation method. The optimal number of clusters was decided based on the dendrogram of the hierarchical cluster analysis. For labeling the obtained clusters, we compared the visualization of state distribution in each cluster. Data analysis was conducted by the R data mining tool (R Core Team, 2017) with the help of TraMiner (A. Gabadinho, Ritschard, Müller, & Studer, 2011) package.

4. Results

The distribution of students' reading activities across the semester is illustrated in Figure 1. According to data presented in Figure 1, most of the activities are in Class (60.6%). On the other hand, *Preview*, *Review*, and *Preview+Review* activities are limited in the dataset. Only 1.7% of all activities include these three reading activities. To make output models simpler we merged these activities into the closest activity. Hence, we merged *Preview* activities with *Preview+Class* activities, *Review* activities with *Class+Review* activities, and *Preview+Review* activities with *Preview+Class+Review* activities.

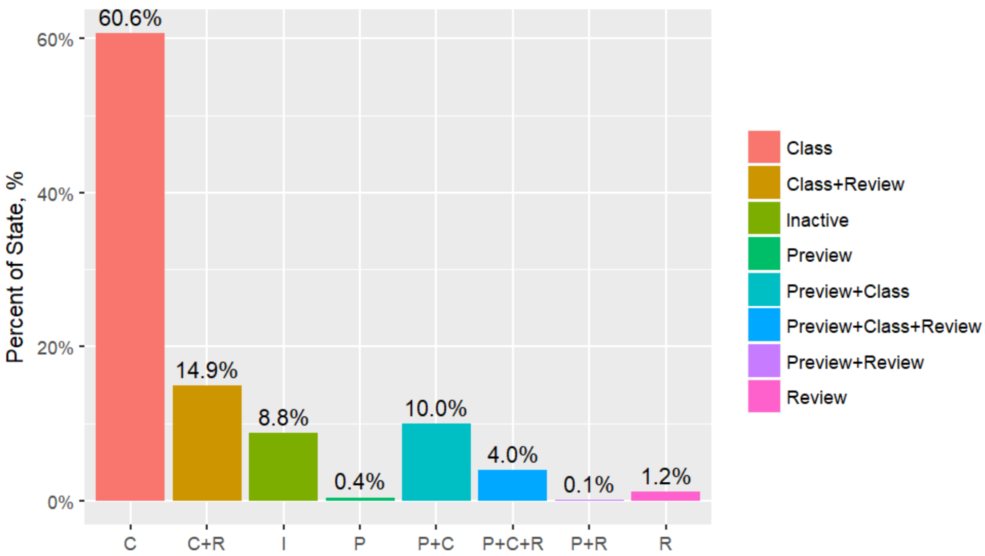

Figure 1. Distribution of Study Activities.

Agglomerative hierarchical clustering based on Ward's algorithm was used to group students based on their similar study patterns. The optimum number of clusters was decided to 3 based on the dendrogram of the hierarchical cluster analysis (refer to Fig. 2).

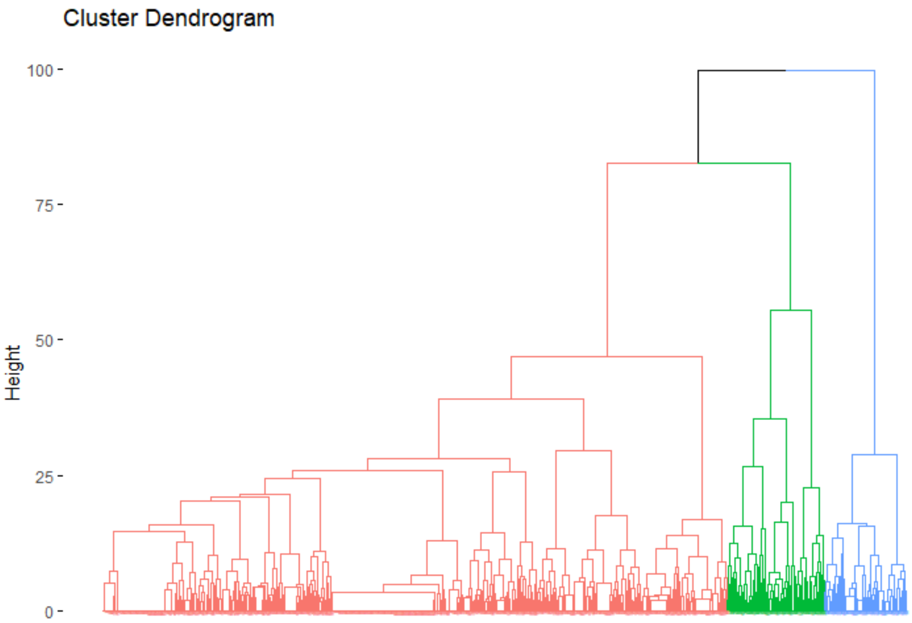

Figure 2. Dendrogram of the Hierarchical Cluster Analysis.

Fig. 3 shows the distribution of students' study patterns in each cluster. Each of the horizontal lines in the graph represents one student's study patterns across all contents. The y-axis shows different learning content used during the course (Content 1 to Content 8). It can be seen from Fig. 3 that students in Cluster 1 (n=1046) are mainly active in class and they are 78% of the total students. On the other hand, they do not have much activity before and after class. Students in Cluster 2 (n = 163) are highly active in-class, before the class and also after the class (12%). Students in Cluster 3 (n = 137) view the learning contents during the class in the beginning but they are becoming inactive week by week (10%).

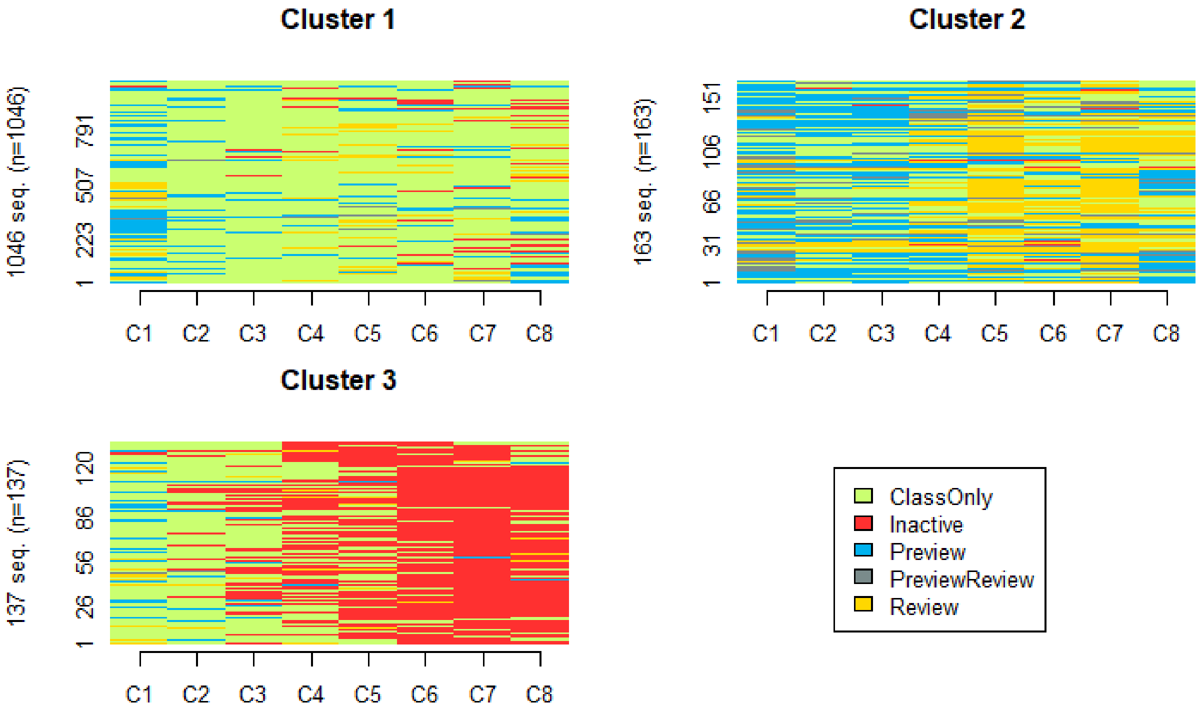

Figure 3. Distribution of Study Activities in each Cluster.

5. Conclusions

In this study, approximately 2 million click-stream data of 1346 students in the Bookroll platform were analyzed aiming to explore the common study patterns of the students followed during the lectures. The results of the analysis showed that a significant number of the students (~80%) viewed the course materials mostly in the class (e.g. Cluster 1). This result is in accordance with previous studies reporting that students frequently do not read their textbooks before the class (Lieu et al., 2017; Ruscio, 2001). On the other hand, it was observed that only a small number of students (12%) viewed the course materials before in class and after the class (e.g. Cluster 2). According to the findings obtained from a previous study (Akçapınar et al., 2020), it can be speculated that these students can be deep learners or students with high self-regulation skills. However, further studies are needed to test these assumptions and understand individual differences between students in different clusters. It is seen that the students in the last group (e.g. Cluster 3) are active in the class in the first two weeks/contents of the course, but they are mainly inactive in the following weeks/contents. These students might be students who are likely to fail the course and are named as at-risk students in the literature. The obtained results can be used to detect these students in a timely manner.

Although pre-class reading was found correlated with academic performance (Gross et al., 2015; Lieu et al., 2017), our results confirmed that most of the students are not willing to read course contents before and after the class. After-class activities (e.g. formative assessment) are also important for learner-centered systems such as flipped classroom settings (Gilboy, Heinerichs, & Pazzaglia, 2015). Learning analytics can be used for continuous monitoring of students' study patterns based on reading traces they left on the eBook reader. Hence, timely interventions can be designed to engage students with pre-class reading and after-class activities. It can also be used for facilitating personalization and adaptation to enhance students' learning.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S) 16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

Akçapınar, G., Chen, M.-R. A., Majumdar, R., Flanagan, B., & Ogata, H. (2020). *Exploring student approaches to learning through sequence analysis of reading logs*. Paper presented at the Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt, Germany.

Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019a). Developing an Early-Warning System for Spotting At-Risk Students by using eBook Interaction Logs. *Smart Learning Environments, 6*(4), 1-15. doi:10.1186/s40561-019-0083-4

Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019b). *Using Learning Analytics to Detect Off-Task Reading Behaviors in Class.* Paper presented at the 9th International Learning Analytics & Knowledge Conference (LAK'19), Tempe, Arizona.

Boroujeni, M. S., & Dillenbourg, P. (2018). *Discovery and temporal analysis of latent study patterns in MOOC interaction sequences*. Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge, Sydney, New South Wales, Australia.

Chen, B., Knight, S., & Wise, A. F. (2018). Critical Issues in Designing and Implementing Temporal Analytics. *Journal of Learning Analytics, 5*(1), 1–9. doi:10.18608/jla.2018.53.1

Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). *Knowledge Map Creation for Modeling Learning Behaviors in Digital Learning Environments.* Paper presented at the 9th International Learning Analytics & Knowledge Conference (LAK'19), Tempe, Arizona.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software, 40*(4), 1-37.

Gabadinho, A., Ritschard, G., Studer, M., & Müller, N. S. (2009). *Mining sequence data in R with the TraMineR package: A user's guide*. University of Geneva.

Gilboy, M. B., Heinerichs, S., & Pazzaglia, G. (2015). Enhancing Student Engagement Using the Flipped Classroom. *Journal of Nutrition Education and Behavior, 47*(1), 109-114. doi:10.1016/j.jneb.2014.08.008

Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased Preclass Preparation Underlies Student Outcome Improvement in the Flipped Classroom. *CBE—Life Sciences Education, 14*(4), ar36. doi:10.1187/cbe.15-02-0040

Hasnine, M. N., Akçapınar, G., Flanagan, B., Majumdar, R., Ogata, H., & Mouri, K. (2018). *Towards Final Scores Prediction over Clickstream Using Machine Learning Methods.* Paper presented at the 26th International Conference on Computers in Education (ICCE2018), Manila Philippines.

Knight, S., Friend Wise, A., & Chen, B. (2017). Time for Change: Why Learning Analytics Needs Temporal Analysis. *Journal of Learning Analytics, 4*(3), 7–17. doi:10.18608/jla.2017.43.2

Lieu, R., Wong, A., Asefirad, A., & Shaffer, J. F. (2017). Improving Exam Performance in Introductory Biology through the Use of Preclass Reading Guides. *CBE—Life Sciences Education, 16*(3), ar46. doi:10.1187/cbe.16-11-0320

Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning, 9*(2), 75-85. doi:10.1007/s11409-014-9114-2

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., . . . Hirokawa, S. (2017). Learning Analytics for E-Book-Based Educational Big Data in Higher Education. In H. Yasuura, C.-M. Kyung, Y. Liu, & Y.-L. Lin (Eds.), *Smart Sensors at the IoT Frontier* (pp. 327-350). Cham: Springer International Publishing.

R Core Team. (2017). R: A language and environment for statistical computing: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ripley, B. (2007). Causation, Counterfactuals, and Critical Reading in the Active Classroom. *International Studies Perspectives, 8*(3), 303-314. doi:10.1111/j.1528-3585.2007.00291.x

Ruscio, J. (2001). Faculty Forum. *Teaching of Psychology, 28*(3), 204-221. doi:10.1207/s15328023top2803_08

Tomasek, T. (2009). Critical reading: Using reading prompts to promote active engagement with text. *International Journal of Teaching and Learning in Higher Education, 21*(1), 127-132.