

How to Mine Student Behavior Patterns in the Traditional Classroom

Chengjiu Yin^{1, a}

¹Kobe University, Kobe, Japan

^ayin@lion.kobe-u.ac.jp

Keywords: Learning Analytics, Data of Traditional Classroom, Clustering, Backtrack Reading

Abstract. Many learning analyses focus on online learning courses, such as massive open online courses (MOOCs). The analysis of learning behaviors from access log data is expected to be of benefit to instructors and learners. However, there are few studies that focus on the reading logs of digital textbooks in the traditional classroom. This study adopts a new approach to analyzing learning behavior patterns through digital textbook use. Students were grouped into four clusters using k-means clustering to analyze their learning behavior patterns.

Introduction

Recently, there has been a growing interest among researchers in Open Educational Resources (OERs), such as OCW(Open Course Ware), and Massive Open Online courses (MOOCs) [1, 2]. Compared with OERs, traditional educational resources, such as books, textbooks, and their learning content, cannot easily be accessed online, and data on students' learning activities are unavailable. Therefore, verifying the educational effectiveness of traditional educational resources remains challenging. Despite the variety of types of traditional learning resources, research on the measurement of their educational effects remains limited.

By contrast, the use of digital textbooks in traditional classrooms enables the recording of learning logs, which can be used to analyze students' learning behaviors [3, 4]. With Moodle and digital textbooks, it is possible to collect learning log data in traditional classrooms and carry out learning analytics. Instructors' lecture materials, such as slides or other notes, can be posted to the digital textbook system, allowing students' learning behaviors to be recorded when they use this to read the learning content.

In this paper, we adopt a new approach to analyzing learning behavior patterns. First, we collected learning log data through the use of digital textbooks; then, we utilized digital textbook reading log data to identify meaningful measures that have strong correlations to learning outcomes. Through use of these measures, the students were grouped using k-means clustering to analyze their learning behavior patterns. We refer to digital textbook log entries, such as "turning to the next page," "preparing for the lesson," "spending time reading pages," and "reading pages" as "learning behaviors."

Literature Review

Data Collection in Previous Studies. Collecting data is the first step in learning behavior analysis [5, 6]. Based on the data sources, previous studies on data collection could be classified into three categories [4]: Questionnaire-based Data Collection (QDC), Manual Data Collection (MDC), and Automatic Data Collection (ADC). For categories QDC and MDC, the data are purposefully collected. For category ADC, the data are objectively collected. A digital textbook system was used for the current research, in which students' learning behavior log data are continuously recorded while they read digital textbooks. Therefore, the current work falls under the category of ADC.

Learning behavior patterns. Learning analysis has previously been undertaken by researchers utilizing these log data, who reported them to be positively related to student effort [7], performance [8], and outcomes [9, 10].

Analyzing learning behavior patterns is a critical topic in learning analysis. Some researchers have determined that writing reflective behavior patterns via a web-based portfolio assessment [11] should be valuable source of data. The research question addressed in this study is as follows:

How can we mine learning behavior patterns from digital textbooks in the traditional classroom?

Methodology

As shown in Figure 1, we propose the following approach to analyzing learning behavior patterns.

1. Collecting learning log data through digital textbooks.
2. Identifying meaningful measures from digital textbook reading log data.
3. Analyzing which of these have strong correlations to learning outcomes.
4. Using these measures to divide students into groups to analyze their learning behavior patterns.

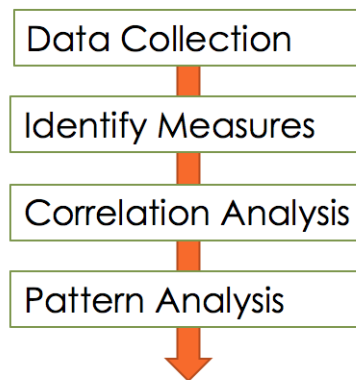


Fig. 1. Analysis of learning behavior patterns

Data collection and measures. The data were collected from digital textbooks. All actions using digital textbooks were recorded in a database. One data log contains the date, time, user, learning content, page number, and other data. Table 1 shows a sample learning log. While a user performs an action, the action and target page number is saved as one record.

Table 1 Sample of Log Data

User ID	Action name	Document	Page Number	Action time
User A	Next	Lesson1	8	2016/10/22 10:40:55
User B	Prev	Lesson2	7	2016/10/22 10:42:15
User C	Marker	Lesson3	7	2016/10/22 10:42:16
User D	Memo	Lesson4	8	2016/10/22 10:42:18

The measures from the digital textbook data were Read Pages (RP), Preview Times (PT), Read Time (RT), Backtrack Reading Rate (BRR), Number of “Next” (NN), and Number of “Prev” (NP):

NN: The number of times a student turns to the next page.

NP: The number of times a student returns to the previous page.

PT: The number of times a student previews the lesson before class.

The students could preview the learning content before class. The reading action logs for “Action Time” showed which students engaged in this behavior, and we calculated the number of times this occurred. RT was used to determine if students previewed their lessons.

RP: The total number of pages that a student read. The reading action logs for “Page No.” and “Action Time” showed how many pages the students read. Many of them repeatedly read specific pages.

RT: The total time spent reading the learning content. The reading action logs “Action Time” showed the length of time students spent reading the learning content. RT was calculated on an hourly basis.

BRR: Some students often backtracked in their reading, as they tried to immediately relate the subject to what they already knew. We also calculated the rate of the backtrack reading time [4].

Correlation analysis. The PA was used to measure the learning outcomes, and we analyzed a number of variables that could affect it. These variables included behaviors and their related variables (time spent reading pages and pages read). SPSS was used to find the partial correlation of PA with these variables. The results show that the variable PA has a significant positive correlation with BRR, PT, RP, and RT [4]. In addition, based on the results of the partial correlation, a k-means clustering analysis was conducted to cluster students into groups to analyze the features of learning behaviors.

Pattern Analysis. A k-means clustering analysis was conducted in the current work. Students were clustered into groups according to similarity in learning behaviors. We then analyzed the features of learning behaviors in groups.

Results

Clusters. Table 2 presents the center, mean values, and standard deviations of each cluster, as well as comparisons of the post hoc tests (Scheffe). Clusters 1 to 4 (C1, C2, C3, C4) had 25, 29, 14, and 30 students, respectively. In order to examine the inter-cluster differences, a one-way analysis of variance (ANOVA) was conducted for each measure with C4 as a between-subject factor (the data of the four clusters satisfied the ANOVA requirements).

Statistical differences were observed in the BRR among the four clusters ($F(3,94) = 36.53, p < 0.001$), as well as with regard to RT ($F(3,94) = 63.53, p < 0.001$), PT ($F(3,94) = 120.29, p < 0.001$), and RP ($F(3,94) = 77.37, p < 0.001$).

Table 2 Clusters

	C1(n = 25) (mean/SD)	C2 (n = 29) (mean/SD)	C3 (n = 14) (mean/SD)	C4 (n = 30) (mean/SD)	F value (ANOVA)
RP	0.70/0.16	0.64/0.13	0.45/0.12	0.04/0.08	77**
RT	0.61/0.15	0.58/0.16	0.40/0.11	0.16/0.11	64**
PT	0.64/0.16	0.08/0.11	0.11/0.17	0.04/0.08	120**
BRR	0.42/0.16	0.38/0.14	0.78/0.13	0.27/0.15	37**
SC	94/6	90/7	80/9	66/22	24**

** $p < 0.001$;

SC: Final examination score; RP: reading pages; RT: reading time; PT: preview times; BRR: backtrack reading rate

Clusters Comparison. Table 2 shows the cluster comparison results comparing the students in C1 and C2. Only the “times of previewing the lesson for class” showed a significant difference (PT: 1 > 2); no significant differences were seen in RP, RT, BRR, or SC. Thus, these students’ learning achievements were not affected by preview times in the cases in which reading times, reading pages, and backtrack reading rates were not significantly different.

In the comparison of students in C3 and C2, significant differences were observed in their backtrack reading rate (BRR: 3 > 2), pages read (RP: 3 < 2), and reading time (RT: 3 < 2), but not in their learning achievements. C3 students tended to frequently review previous pages, clocked a shorter time for reading, and obtained satisfactory learning achievements (similar to C2 students). These findings show that BRR has a significant positive influence on learning effectiveness, whereby it helps students to efficiently manage their time for learning. BRR has a relevant correlation with learning efficiency and is thus a “good” reading strategy.

The comparison between C3 and C1 students shows significant differences in backtrack reading rates (BRR: 3 > 1), read pages (RP: 3 < 1), reading times (RT: 3 < 1), and learning achievements (PA: 3 < 1). The findings show that, although C3 students demonstrated a good reading style, they still needed to spend more time reading the learning content to ensure better learning achievements. In other words, both a good reading style and sufficient learning time are required.

Conclusions

In this study, we analyzed learning behaviors through the use of digital textbook reading log data to determine what types of learning behavior patterns are positive and significantly related to learning outcomes. The students were clustered into four clusters according to different learning behavior patterns. The main finding that emerged from the analysis results was that backtrack reading learning behavior is beneficial because it can help students save time when studying.

Acknowledgements

Part of this research work was supported by the Grant-in-Aid for Scientific Research No. 16H03078 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan.

References

- [1] Eisenberg, M., Fischer, G., MOOCs: A perspective from the learning sciences, in: J. L. Polman et al. (Eds.), *Proc. of International Conference of the Learning Sciences*, 2014, pp.190-197.
- [2] Siemens, G., Dillenbourg, P., Where are the learning sciences in the MOOC debate?, in: J. L. Polman et al. (Eds.), *Proc. of International Conference of the Learning Sciences*, 2014, pp.15-17.
- [3] Yin, C., Okubo, F., Shimada, K., Yamada, M., Fujimura, N., Ogata, H., Smart phone based data collecting system for analyzing learning behaviors, in: C.-C. Liu, et al. (Eds.), *Proc. International Conference of Computers on Education*, 2014, pp. 575-577.
- [4] Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., Yamada, M., Kojima, K., Ogata, H., Identifying and analyzing the learning behaviors of students using e-books, in: H. Ogata, et al. (Eds.), *Proc. of International Conference on Computers in Education*, 2015, pp. 118-120.
- [5] Yin, C.J., Sung, H.Y., Hwang, G.J., Hirokawa, S., Chu, H.C., Flanagan, B., Tabata, Y., Learning by searching: A learning approach that provides searching and analysis facilities to support research trend surveys, *Journal of Educational Technology & Society*, 16 (2013) 286-300.
- [6] Yin, C.J., Hirokawa, S., Yau, J., Nakatoh, T., Hashimoto, K. Tabata, Y., Analyzing research trends with cross tabulation search engine, *International Journal of Distance Education Technologies*, 11 (2013) 31-44.
- [7] Campbell, J.P., DeBlois, P.B., Oblinger, D.G., Academic analytics: A new tool for a new era. *Educause Review* 42 (2007) 40-57.
- [8] Macfadyen, L., Dawson, S., Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan, *Educ. Technol. Soc.* 15 (2012) 149-163.
- [9] Archer, E., Chetty, Y.B., Prinsloo, P., Benchmarking the habits and behaviours of successful students: A case study of academic-business collaboration, *The International Review of Research in Open and Distance Learning*, 15, (2014) 62-83.
- [10] Hrastinski, S. A theory of online learning as online participation, *Comput. Educ.* 52 (2009) 78-82.
- [11] Chang, C.C., Chen, C.C., Chen, Y.H., Reflective behaviors under a web-based portfolio assessment environment for high school students in a computer course, *Comput. Educ.* 58 (2012) 459-469.