Chang, M. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

An Analysis of Learning Behavior Patterns with Different Devices and Weights

Chengjiu YIN [a*], Juan ZHOU[b], Lingyu LI[a], Koudai Yamaguchi[a], Noriko UOSAKI[c], & Hiroaki OGATA[d]

[a]*Kobe University, Japan*
[b]*Ritsumeikan University, Japan*
[b]*Osaka University, Japan*
[b]*Kyoto University, Japan*
[*]yin@lion.kobe-u.ac.jp

Abstract: With e-learning systems gradually being implemented, researchers worldwide have started devoting increasing attention to Learning Analytics. At Kobe university, a digital textbook reading system has been developed to collect learning logs in the face-to-face classroom. In a previous study, *k*-means clustering was implemented to analyze learning behavior patterns; however, there were problems such as few variables for clustering and a failure to consider weighting of the learning elements. Therefore, in this study we applied clustering by increasing the number of learning elements and assigned weights to the learning elements, then analyzed the learning behavioral patterns. We found some behavioral patterns of students who can save learning time if they effectively write memos and add markers.

Keywords: Learning behaviors, *k*-means clustering, learning log, digital textbooks

1. Introduction

In recent years, owing to the development and spread of PCs and mobile devices, many educational institutions have been investigating and implementing e-learning systems. In fact, 33.8% of the universities in Japan had instituted Learning Management Systems (LMS) in preliminary learning and post-learning in 2010. This value had risen to 44.9% in 2013. On the other hand, 36.3% of the universities had implemented blended learning (a kind of learning method that combines real courses with e-learning in which learners can use a real course and web group course together) in 2010, a figure that had risen to 44.0% in 2013.

With e-learning systems gradually being implemented, researchers worldwide have started devoting increasing attention to Learning Analytics (LA) (Yin & Hwang, 2018). Researchers have reported that LA can be used for learning support and class improvement (Mostow, 2004; Yin et al., 2014; Yin et al., 2015), and LA is positively related to student efforts (Campbell, DeBlois, & Oblinger, 2007), performances (Macfadyen & Dawson, 2012), and outcomes (Archer, Chetty, & Prinsloo, 2014; Hrastinski, 2009; Yin & Hwang, 2018). Yamada (2017) pointed out that in LMSs, administrators and teachers usually pay attention to the logs used to detected abnormal activities, such as unauthorized access and system failure. The system-use logs are frequently neglected. However, LA pays attention to the system-use logs, which are used to analyze learners' behavior to find ways to improve teaching and support learning. For example, some researchers used learning logs of online courses from MOOCs (Massive Open Online Courses) and OERs (Open Educational Resources) to perform learning analytics.

Many recent studies have focused on collecting data from online performance for learning analytics; in contrast, the application of learning analytics in face-to-face classes is rare worldwide (Yin et al., 2018). Therefore, to collect learning logs in face-to-face classes, Yin et al. (2017) developed an electronic teaching material system, DITeL (Digital textbook for Improving Teaching and Learning), and then, using learning logs collected by DITeL, *k*-means clustering was used to divide learners into clusters and analyze the relationship between learners' learning behavioral patterns and learning achievement (Yin et al., 2018).

However, few learning elements were used for clustering in the previous study, and weighting of learning elements was not considered. To solve these two problems, in this study we increased the number of learning elements, such as markers and devices, to apply learning analytics more effectively. As different learning elements have different learning effects on academic performance, we also assigned a weight to each element and divided learners into clusters according their learning logs. Cluster analysis was then performed with the statistical analysis software package SPSS (developed by IBM), and we analyzed and discussed the relationship between learning elements and learning achievement. We found some behavioral patterns for such groups as students who can save learning time if they effectively write memos and add markers.

2. Literature Review

Data analysis is an internal step in the process of data collection (Yin, Hirokawa, et al., 2013; Yin, Sung, et al., 2013). In this study, the system collects data automatically while the learner is using the system, therefore the data are objective (Yin et al., 2018).

2.1 Clustering

There are many other data mining methods used in educational research, such as Aprioi and SVM. In this study, *k*-means clustering was adopted to group large and diverse data into groups. The variables have similar values in the same group significantly different from the values for the other groups; therefore, *k*-means clustering is better than other methods (Yin et al., 2018).

In a previous study, four learning elements (reading time, the number of pages viewed, page backtracking rate, and the times of preview lessons) were used to divide the data into clusters (Yin et al., 2018). However, only a few learning elements were used in the clustering, although there are other learning elements that also affect learning achievement. Therefore, in this paper, besides the reading time, the number of pages viewed, and page backtracking rate, we added the number of Markers and the number of Memos for *k*-means clustering.

Liu, Feng, Shi, and Guo (2014) insisted that weighting should be performed when dividing data into clusters. This study assigned a weight to each learning element before clustering. A previous study (Yin et al., 2018) did not consider the weighting of different variables (learning behaviors) even though different variables have different impacts on grades. Since it was confirmed that the learning achievement has a different correlation with each learning element, it is necessary to assign a weight to each element.

3. E-Book System DITeL

As motioned above, we developed an e-book system, DITeL, to collect data in face-to-face classes. DITeL is supported on a variety of devices such as PCs, Mobiles, and Tablets, because different learners learn on different devices. In addition, DITeL has various functions to make learning more efficient.

Every user action (such as learning material name, page number, action time, and device) is recorded as a learning log. Table 1 shows an example of the learning logs from which the data are aggregated. The variables for aggregating are shown in Table 2.

Table 8

An Example of the Learning Logs Recorded on the Server

Userid	*Action name*	*Learning material*	*Page Number*	*Action time*	*Device*
Student1	*Next*	*Law Course*	*16*	*2017/5/22 8:40*	*PC*
Student1	*Prev*	*Law Course*	*15*	*2017/5/22 8:42*	*Mobile*
Student2	*Add UnderLine*	*Law Course*	*15*	*2017/5/22 8:42*	*Tablet*
Student3	*Add Memo*	*Law Course*	*15*	*2017/5/22 8:42*	*Mobile*

Table 2

The Variables of the Data Recorded on the Server

Elements	Define
PC Read Pages (PRP)	The number of pages the student read on a PC
Mobile Read Pages (MRP)	The number of pages the student read on a Mobile device
Table Read Pages (TRP)	The number of pages the student read on a Tablet PC
BookMark(BM)	The number of Bookmarks created by students
Memo	The number of Memos created by students
Highlight	The number of Highlights created by students
UnderLine	The number of Underlines created by students
Prev	The number of times the student returned to the previous page
Next	The number of times the student advanced to the next page
Reading Time(RT)	The number of seconds the student spent reading the teaching material
Reading Page(RP)	The number of pages the students read of the teaching material

4. Analysis of Learning Logs and Analytical Results

4.1 Problems in Analysis

In previous studies analyzing learning behavior patterns with learning logs, *k*-means clustering was applied. However, there were some problems:

1. The number of variables collected was few, and the relationships between variables were not examined at all.
2. The differences in devices and the students' tendencies to learn had not been considered.
3. The different variables have difference influences on learning achievement, so it is necessary to assign weights to the variables.

In order to solve the above problems, the relationship between learning behavior patterns and learning achievement was analyzed after assigning weights to the variables. The analytical steps are described below.

- Step 1: Data Structuring; summarizing learning logs and formatting the data.
- Step 2: Data Correlation; in order to examine the relationship between learning achievement and other variables, correlation analysis is used.
- Step 3: Assigning weights; the variables are assigned weights based on the correlation coefficients.
- Step 4: Clustering; students were clustered into groups by these weighted variables.
- Step 5: Multiple comparison test to find and objectively consider differences between groups.

4.2 Variables

Based on the variables in Table 2, new variables were also generated and unnecessary variables deleted as described below.

- Due to the small number of users, “Tablet ReadPage” and “BookMark” were not used because their influence on the results would not be accurate.
- Since “HighLight” and “UnderLine” are similar in usage, they were combined to create a new element “Marker,” and “HighLight” and “UnderLine” were not used.
-

$$Marker = HighLight + UnderLine$$

- In order to investigate the difference in consciousness between individuals, the variables “Prev” and “Next” were converted to the new variables “PrevPer” (PP) and “NextPer” (NP).

$$PP = \frac{Prev}{Prev+Next} \quad , \quad NP = \frac{Next}{Prev+Next}$$

- “GPA” was used to represent the final grade of students.

4.3 Correlation Analysis

In this study, correlation analysis was performed to determine the degree of influence of each factor on GPA as measured by the correlation coefficient. Statistical analysis software (IBM SPSS Statistics) was used for the correlation analysis.

In this study, p-values were considered significant at the 5% level. Also, negative values were not considered, as they had negative correlations with GPA. The general results are as follows.

1) Since “RP” and “PRP,” “MRP” and “Prev,” and “MRP” and “Next” were strongly correlated with each other, their effects on GPA overlapped. It was found that these variables were extracted from the same element of the number of pages. In this paper, in order to focus on devices, “RP,” “Prev,” and “Next” were not considered, but “RT” was retained because it was strongly correlated with GPA, and because it represents the time of extracting variables differently.
2) Since “NP” had negative correlations with all variables, it was not considered in this research.

4.4 How to Assign Weights to Variables?

Correlation analysis found that “GPA” could be modeled by six variables: “Memo,” “Marker,” “RT,” “PP,” “PRP,” and “MRP.” Considering the relationships between them, these six variables were roughly divided into three groups. We defined learning with “Memo” and “Marker” as plus-α learning, “RT” (reading time) and “PP” (rate of return to the previous page) were learning traces, and “PRP” (read pages using PC) and “MRP” (read pages using mobile devices) were devices. These same groups did not affect each other. We then analyzed the relationship between those groups and found two sufficient conditions:

1. If the value of device is larger, then learning trace and plus-α learning are larger. For example, if the student reads many pages of content, then he will return to the previous page many times and write memos or add markers many times.
2. If the value of learning trace is larger, then positive plus-α learning is larger. For example, if the student reads the content for a long time, then he will write memos or add markers many times.

The weights were calculated based on these two sufficient conditions. For example, “GPA” was correlated with “Memo” with a strength of 0.294. Also, “Memo” was correlated with “RT,” “PRP,” and “MRP,” respectively, with strengths of 0.146, 0.36, and 0.427. That is, “RT,” “PRP,” and “MRP,” respectively, were correlated with “GPA” with strengths of (0.294 * 0.146), (0.294 * 0.36), and (0.294 * 0.427). It was possible to obtain weights by this calculation method and sum them all. The values of the weights are listed in Table 3.

Table 3

The Values of the Weights

Memo	Marker	RT	PP	PRP	MRP
0.294	0.339	0.601	0.498	0.816	0.984

5. Discussion

To further understand students' possible behavior patterns, cluster analysis was employed. Students were clustered into five groups according to the similarities in their learning behaviors. We used *k*-means clustering to group learners.

Table 4 presents the center, mean values, and standard deviations of each cluster as well as comparisons with a post-hoc test (Scheffe). Clusters 1–5 (C1, C2, C3, C4) included 65, 54,43, 51, and 22 students, respectively. We then analyzed the learning behavior features in each group.

Table 4

Assigned Weighted k-Means Clustering Results and Analysis

	Learning Behavior Clusters						
	Cluster 1 (*n* = 65) (mean/*SD*)	Cluster 2 (*n* = 54) (mean/*SD*)	Cluster 3 (*n* = 43) (mean/*SD*)	Cluster 4 (*n* = 51) (mean/*SD*)	Cluster 5 (*n* = 22) (mean/*SD*)	*F*-value (ANOVA)	Post-hoc (Scheffe) tests
PRP	0.034/0.034	0.047/0.04	0.155/0.093	0.222/0.126	0.123/0.093	50.776**	4>3.5>1.2
MRP	0.03/0.034	0.186/0.082	0.04/0.047	0.094/0.076	0.443/0.163	140.882**	5>2>4>1,2>3
Memo	0.01/0.024	0.04/0.047	0.104/0.054	0.018/0.034	0.116/0.064	52.14**	3.5>2>1,3.5>4
Marker	0.01/0.018	0.034/0.03	0.09/0.054	0.023/0.033	0.141/0.082	60.241**	5>3>2>1,5.3>4
PP	0.167/0.099	0.189/0.072	0.207/0.074	0.278/0.054	0.251/0.062	17.697**	4.5>1.2,4>3
RT	0.035/0.024	0.12/0.056	0.107/0.047	0.159/0.086	0.253/0.088	65.171**	5>4>2.3>1
GPA	5.357/2.355	6.548/2.364	6.842/1.508	7.125/2.017	7.782/1.036	8.642**	2.3.4.5>1

***$p < 0.001$.*

PRP: PC Read Pages; MRP: Mobile Read Pages; RT: Reading Time; PP: The rate of returns to previous page

Cluster 3: The reading time (RT) of Cluster 3 is lower than Clusters 4 and 5; however, "Memo" and "Marker" are higher than in Clusters 1, 2 and 4, and "Memo" was almost the same as in Cluster 5, while "MRP" was as low as in Cluster 1, meaning that the students in Cluster 3 were more likely to use a PC device and were actively involved in the learning process, as they actively wrote memos and added markers. They were thus characterized by plus-α learning.

There was no significant difference between Clusters 2, 3, 4, and 5 in "GPA" through a ceiling effect: The upper limit of "GPA" was fixed, so even extensive study could not greatly increase the GPA; any such growth would be small. In other words, for example, the members of Cluster 5 could acquire knowledge with less learning time.

Cluster 3 should be noted. Despite the short learning time, this cluster effectively used the plus-α learning method to good effect. Clusters 2 and 4 would have been able to reduce their learning time if they had effectively handled "Memo" and "Marker." Also, if the students in Cluster 1 also had a bottleneck when learning, it would be best for them to try the plus-α learning method first.

6. Conclusion

In this paper, we have described the need to weight the variables, which was confirmed by the experimental results.

In this paper, we particularly focused on the learning effects of "Memo" and "Marker." The use of these functions led to reduced learning times and made it possible to intensively learn some contents or other subjects. Also, it is good for learners who do not like learning, because it has been verified as reducing the learning time. Thus, students can try this learning method using "Memo" and "Marker." In addition, if students want to use “Memo” or “Marker,” it is suggested that they use a PC with a large screen, whereas if they want to use this system in their free time or in a conveniently carried form, they can use mobile devices.

Acknowledgements

Part of this research work was supported by the Grants-in-Aid for Scientific Research Nos. 16H03078, 16H06304, and 18H04125 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan. We would like to thank all the people who prepared and revised previous versions of this document.

References

Archer, E., Chetty, Y. B., & Prinsloo, P. (2014). 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*(1), 62-83 .

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *Educause Review*, *42*(4), 40–57.

Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education*, *52*(1), 78–82.

Macfadyen, L., & Dawson, S. (2012). Numbers are not enough: Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, *15*(3), 149–163.

Liu, R., Feng, S., Shi, R., & Guo, W. (2014). Weighted graph clustering for community detection of large social networks. *Procedia Computer Science*, *31*, 85–94.

Mostow, J. (2004). Some useful design tactics for mining its data. *Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes*. In: Lester J.C., Vicari R.M., Paraguaçu F. (eds) Intelligent Tutoring Systems. ITS 2004. Lecture Notes in Computer Science, 3220. Springer, Berlin, Heidelberg (pp. 20–28).

Yamada, M. (2017). The recent trends of learning analytics research and future directions (in Japanese). *Educational Technology Research*, 41(3), 189–197.

Yin, C., Hirokawa, S., Yau, J., Nakatoh, T., Hashimoto, K., & Tabata, Y. (2013). Analyzing research trends with cross tabulation search engine. *International Journal of Distance Education Technologies*, *11*(1), 31–44. doi:10.4018/jdet.2013010103.

Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning*, *10*(4), 455–468.

Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., Yamada, M....Ogata, H. (2015). Identifying and analyzing the learning behaviors of students using e-books. In: Ogata, H. et al. (Eds.) *Proc. of 23rd International Conference on Computers in Education*. Hangzhou, China: Asia-Pacific Society for Computers in Education (pp. 118–120).

Yin, C., Okubo, F., Shimada, K., Yamada, M., Fujimura, N., & Ogata, H. (2014). Mobile devices based data collecting system for analyzing learning behaviors. In C.-C. Liu et al. (Eds.), *Proc. International Conference of Computers on Education*, Nara, Japan: Asia-Pacific Society for Computers in Education (pp. 575–577).

Yin, C., Sung, H. Y., Hwang, G. J., Hirokawa, S., Chu, H. C., Flanagan, B., & Tabata, Y. (2013). Learning by searching: A learning approach that provides searching and analysis facilities to support research trend surveys. *Journal of Educational Technology & Society*, *16*(3), 286–300.

Yin, C., Uosaki, N., Chu, H., Hwang, G., Hwang, J., Hatono, I., Kumamoto, E., Tabata, Y. (2017). Learning behavioral pattern analysis based on students’ logs in reading digital books. In : Chen, W. et al. (Eds.) *Proceedings of the 25th International Conference on Computers in Education*, (2017). Christchurch, New Zealand: Asia-Pacific Society for Computers in Education (pp. 549–557).

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2018). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human Factor Approach. *International Journal of Human–Computer Interaction*, *35*(4–5), 313–322. DOI: 10.1080/10447318.2018.1543077