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

Using Sequence Clustering to Unveil Students' Learning Strategies and Explore the Relationship with Cognitive Load

Ching-Hsiang Kang[a], Anna Y.Q. Huang[a*], Owen H.T. Lu[b], Bin-Shyan Jong[c], Irene Y.L. Chen[d], & Stephen J.H. Yang[a]
[a*]*Computer Science & Information Engineering, National Central University, Taoyuan City, Taiwan*
[b]*College of Computer Science, National Pingtung University, Pingtung City, Taiwan*
[c]*Information & Computer Engineering, Chung Yuan Christian University, Taoyuan City, Taiwan*
[d]*Department of Accounting, National Changhua University of Education, Changhua City, Taiwan*
*anna.yuqing@gmail.com

Abstract: The generation of learning strategies is related to students' cognitive abilities, and students with similar cognitive abilities tend to adopt similar learning strategies. This study conducted an experiment of a system programming course with 113 college students and collected logs from a learning management system. There are five learning strategies that can be identified through sequence clustering analysis, they are comprehensive video-oriented learning, comprehensive slide-oriented learning, offline learning, selective learning, and reflective learning. Among the five learning strategies, selective learning strategy is positively correlated with students' cognitive load, which can be explored that these students make practice of quiz without reading learning materials, and tend to select unfamiliar concepts to read based on quiz results. Students who use selective learning strategy are referred to as incomprehensive learners, and the selective learning strategy prone to increase students' cognitive load due to overflying unclear and segmented information.

Keywords: Cognitive Load, Sequence Clustering, Learning Strategies

1. Introduction

There are some learning strategies or patterns will significantly affect learning performance, such as cognitive load (Paas, 1992), which makes it necessary to reveal the student 's condition immediately. A large number of students' interactions with computers can be summarized and extracted from the logs of the digital learning environment. Therefore, these logs that are generated in real time and accurately recorded may be used as an auxiliary mechanism for questionnaire measurement. Many studies have started from software logs, such as: Akçapınar et al. (2019) developed an early warning system for students by logs which generated from e-Book software. However, students' learning strategies is a potential structure and cannot be observed directly in learning logs, but instead must use the appropriate analysis methods and techniques of data mining, past research has shown that by converting students 'learning logs into sequences, and the learning strategies of students can be well found through sequence clustering (Akçapinar et al., 2020; Jovanović et al., 2017).

In summary, this study focuses on identifying a most suitable sequence clustering for extracting students' learning strategy from learning logs which collected from learning management system (LMS). In addition, this study examined the correlation between students' learning strategies and cognitive load. Therefore, there are two research questions in this study.

RQ1: How to extract students' learning strategy from the logs collected from LMS?
RQ2: What is the correlation between students' learning strategy and cognitive load?

2. Methodologies

2.1 Participants

This study is based on a system programming course at a university in Taiwan. The experiment duration ran for 9 weeks. There are 113 students participating in the course. Moreover, this course used a software named i-learning as the LMS platform, which includes teaching material videos, slides, discussion areas, teaching material downloads, and test.

2.2 Instrument

We adopted a cognitive load instrument for measuring students' learning status. The number of valid questionnaires is 88 out of 113 students. It is used to evaluate the load generated by students' cognition of the system program course. The cognitive load questionnaire in this study was based on Sweller et al. (1998) and Paas (1992). The questionnaire is a 6-point Likert scale, a total of eight questions, ranging from 1 (very disagree) to 6 (very agree). The cognitive load questionnaire contains two parts: questions 1-5 were a group as the first part named mental load, and questions 6-8 were a group as another part named mental effort.

2.3 Constructing students' learning sequence

In order to perform sequence analysis, this session will introduce the process and method for students to extract the raw data from the i-learning LMS to the sequence, and collect students' learning logs for 9 weeks. This study uses final exam score to represent students' learning performance. Students' learning logs on the i-learning LMS will be recorded in the database. Each event record is represented by three fields, including username, event begin time, and event title as shown in Table 1.

Table 1. *The Classified Event Categories*

Reclassified event category	**Related event categories in this category**
Regulations	Information about course, grading standards, teaching objectives
Test	Events related to online test
Course key points video	Related key points videos by the teacher in course
Teaching material download	Course notes, reference materials, program execution examples, noun explanation
Previous exam	Previous exam by year
Course video	Video teaching materials for teachers' various chapters
Course slide	Slides of the teaching materials of the teachers' various chapters
Discuss	Course discussion, software development discussion, course content explanation
Questions related to the course	Learning review videos

2.4 Sequence clustering methods

We used agglomerative hierarchical clustering method to cluster students' learning sequences. For each iteration of agglomerative hierarchical clustering method, the clustering task starts from the tree leaves and then combines two nearest clusters into one. This technique is considered particularly suitable for detecting student groups in online learning environments (Kovanović et al., 2015). Aggregated hierarchical clustering method starts aggregation from the bottom of the tree structure, and the algorithm steps are as follows (Murtagh, 1983):

1. At the beginning, each data point is a cluster denoted by $\mathbf{C_i}$, $i = 1,2,\ldots,n.$
2. Find the nearest two clusters $\mathbf{C_i}$ and $\mathbf{C_j}$ within all clusters.
3. Combine $\mathbf{C_i}$ and $\mathbf{C_j}$ into a new cluster.

4. If the number of clusters is equal to the desired one, stop. Otherwise go back to step 2 to create more clusters.

3. Results and discussion

The preprocessed sequence is clustered by agglomerative hierarchical clustering method, and the result shows the number of clusters is 5, and the smallest cluster size is 30 and the largest cluster size is 284. In order to answer the RQ1 (How to extract students' learning strategy from the logs collected from LMS), we cluster a total of 621 learning sequences after pre-processing to detect students' learning strategies on the LMS. The results of the sequence clustering show that students' learning process will produce different types of learning strategies, which including:

Comprehensive video-oriented learning strategy: The learning sequence in this strategy is mainly composed of the interlaced reading of course video and course slide. Most of the first event in this learning sequence is watching the course video. We then sorted out the top three most frequently learning sequences in this learning strategy. The most common sequence was students stopped after course video, with a sequence length of 1, and the second most common sequence was to watch the course video followed by the course slide, the sequence length is 2, and we can find that the remaining sequences follow the course video connection to any event and then the course video, the sequence length is equal or larger than 3. From the above observation, we can consider that this learning strategies is relatively complete learning. Therefore, we named this strategy as comprehensive video-oriented learning.

Comprehensive slide-oriented learning strategy: The learning sequence in this strategy is mainly composed of course slide and course video. According to the statistics results, this strategy has higher number of the event of course slide than the strategy of comprehensive video-oriented learning. The average length of the learning sequence of this strategy is the longest than any other strategies. Most of the first event of this strategy are students will watch the course video. It can be seen from the above discussion that this learning strategies is similar as pervious one: starting with basic teaching material. However, we named this strategy as comprehensive slide-oriented learning due to the major chrematistic of this strategy is the slide-oriented learning.

Offline learning strategy: According to the statistics results of this strategy, it can be observed that the major sequences are teaching material download, which is considering as offline learning. Students will download teaching material at the first beginning of the class. Students in this strategy were search the chapter videos or slides that they want to watch. Therefore, we named this strategy as offline learning.

Selective learning strategy: According to the statistical results, it can be observed that the learning sequence in this strategy is mainly consists of previous exam. Most of the first event in this strategy is students tends to do previous exam. We further sorted out the top three most frequently learning sequences in this strategy and found that the most common sequence is to stop after doing the previous exam. The remaining two common learning sequences are to watch the course video after doing the previous exam, and to watch the course video after doing the previous exam. Both of remaining two common learning sequences' length are 2 or over 2. The selective learning strategy represents that students are reading for the purpose of being able to pass the course, and it can be found that the average number of times of the course video and the course slide in the sequence are lower than the overall average. This indicating that the selective learning strategy is lacks of learning from the fundamental concepts, considering that the behavior in this strategy is only for performance-oriented learning. Therefore, we named this strategy as selective learning.

Reflective learning strategy: The diversity of learning sequences in this strategy is relatively higher than other strategies. According to the statistical results, it can be observed that the proportions of the two events of course key points video and discussion are the highest in the learning strategy. This strategy means students will find key points to organize and participate in the discussion after watching the slide and video. It means that students have the ability to reflect on the teaching materials and can review the past experience to improve learning behavior. Therefore, we named this strategy is reflective learning.

To reply RQ2 (What is the correlation between students' learning strategy and cognitive load?), we used Spearman correlation coefficients to explore the correlation between learning strategies and cognitive load. We test the significance of the correlation between the sum of the number of sequences

in each learning strategy of each student and cognitive load. The results of Spearman correlation coefficients indicated that the selective learning strategy has a significant positive correlation with mental load, while mental load refers to working memory load. The correlation is significant at the 0.05 levels, and the correlation coefficient value is 0.22. Cognitive load as the amount of load generated by applying specific tasks to students' cognitive systems. To further understand how selective learning strategy affects mental load, we test the significance of the correlation between the top three most frequently learning sequences of selective learning strategy and mental load.

Statistics results indicated that the mental load is significantly positively related to the sequence of watch the course video after doing the previous exam. The correlation is significant at the 0.05 levels, and the correlation coefficient value is 0.21. This sequence shows that the students tend to be performance-oriented first, and then find important information in the teaching material based on the previous exam, relying heavily on the exam content for learning, lack of basic concepts, students hope to receive important information about the exam in a short period of time. This learning strategy may cause students to be unable to load knowledge for a while, and according to past research, mental load is considered to be an important indicator that affects student' learning performance. We hope that by correcting this kind of learning strategy of the students to improve students' learning performance.

4. Conclusion and future research

After a series of log data processing processes, we have summarized five learning strategies for students: Comprehensive video-oriented learning, Comprehensive slide-oriented learning, Offline learning, Selective learning, and Reflective learning. The essence of the difference between the five strategies is the completeness of watching movies, learning materials, and the order of actions. The most critical one of the five strategies are Selective Learning. The main reason is that the strategy has a significant positive correlation with students' cognitive load. More specifically, when students skip watching videos and learning material and take quizzes directly, or the student takes the quiz first and then returns to the video or learning material to find the answers to the questions in the quiz; this is a symptom of the student's cognitive overload. The above summary clearly shows that in the design of learning activities or the design of teaching materials, to avoid the situation we mentioned above, it must induce or guide students to read at least once. Additionally, with a mechanism of encouragement to enable students to go back and reread the textbooks or videos under specific conditions, it will effectively reduce cognitive load and have the opportunity to improve learning outcomes. In the future, we will design and implement classroom activities under the concept mentioned above and encourage, guide and inducing students to read the content in a precise number of times through intervention or prevention mechanisms and expose the actual benefits to students during this process.

Acknowledgements

This work is supported by Ministry of Science and Technology, Taiwan under grants MOST-109-2511-H-008-007-MY3, MOST-108-2511-H-008-009-MY3, MOST-106-2511-S-008-004-MY3, and Ministry of Education, Taiwan.

References

Akçapinar, 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.

Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Developing an early-warning system for spotting at-risk students by using eBook interaction logs. *Smart Learning Environments, 6*(1), 4.

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education, 33*(4), 74-85.

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education, 27*, 74-89.

Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The computer journal*, *26*(4), 354-359.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*, *84*(4), 429.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251-296.