1F1-OS-26a-1

Behavioral Analysis and Visualization on Learning Logs from the CALL Course

Huiyong Li[*1], Tomoyuki Tsuchiya[*2], Daiki Suehiro[*3], Yuta Taniguchi[*3],

Atsushi Shimada[*3], Yubun Suzuki[*2], Hiroshi Ohashi[*3], Hiroaki Ogata[*3]

[*1] Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan

[*2] Faculty of Languages and Cultures, Kyushu University, Japan

[*3] Faculty of Arts and Science, Kyushu University, Japan

1. Introduction

Computer-assisted language learning (CALL) is often used as an approach to foreign language teaching and learning in higher education. The CALL course is offered at a national university in Japan to allow freshman students to perform self-regulated learning (SRL) with e-learning materials for the purpose of developing language skills such as grammar, listening and reading.

Self-regulated learning is an active learning process used to regulate and monitor learning cognition, motivation, and behavior, thereby setting personal goals (Wolters, Pintrich & Karabenick, 2005). In the process of SRL, students often do not recognize how they are learning and thus do not appreciate many beneficial learning strategies (Bjork, Dunlosky, & Kornell, 2013). Students who underuse effective strategies are more likely to obtain lower achievement, and therefore it is critical to provide adaptive support to promote freshman students' self-efficacy.

Using of visualizations, it is easy to help learners and instructors understand and discover patterns in data. Visualizations of learner activities, especially the details of learning process, can motivate learners, provide prompt feedback on their work, reveal participation and detect whether the learners are learning. Thus, a visual learning support system for the freshman students is proposed.

The system is designed with the analysis results of actual learning behaviors, and to support learners and instructors using individual behavioral traces. From the huge logs in current CALL course, the feature items of personal learning activities will be extracted, and the behavioral patterns and preferences of the students in the SRL processes will be identified. Moreover, a set of the visualization tools will be shown in order to provide customized feedback on the course level and the individual level.

2. The CALL course

A total of 2,631 students enrolled in 50 CALL classes at the national university in Japan. The CALL classes were provided for freshman students of all departments, with two credits from spring and autumn semesters 2016. The students were supposed to perform self-paced language learning outside of the classroom. Table 1 shows the course schedule in the spring semester 2016. There were four sub-deadlines in one semester. The students were required to complete the assigned materials from stage 1 to stage 3, with those for stage 4 meant as an option.

Table 1. Course schedule in the spring semester 2016

		Learning materials assigned		
Stage	Deadline	Reading	Listening	Grammar
1	Week 5	Reading1	Listening1	Grammar1
2	Week 10	Reading2	Listening2	Grammar2
3	Week 15	Reading3	Listening3	Grammar3
4(optional)	Week 21	Reading4	Listening4	Grammar4

The e-learning materials of the CALL course contained grammar, listening and reading sections and included 493 units, with a total of 751 quiz items. The details of the e-learning materials are shown in Table 2.

Table 2. Unit numbers and item numbers of learning materials

Section	Part	Category	Unit #	Item #
Reading	1	Reading comprehension	6	30
	2	Reading comprehension	7	35
	3	Reading comprehension	6	30
	4	Reading comprehension	6	30
Listening	1	Short conversation	15	35
	2	Long conversation	14	46
	3	Long announcement	15	66
	4	Formal conversation	22	77
Grammar	1	Grammar and word usage	95	95
	2	Grammar and word usage	89	89
	3	Grammar and word usage	98	98
	4	Grammar and word usage	120	120
		Total	493	751

The moment students practiced quiz items online, the learning behaviors were recorded in server logs concurrently. There were three types of learning logs, including access to learning materials (access logs), completed quiz items (completion logs), and quiz answers (answer logs). A total of 7,413,397 learning

Contact: Huiyong Li, Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan, E-mail: rikeiyuu@gmail.com

logs were retrieved and analyzed from the server of the CALL course, with 1,792,277 access logs, 1,117,375 completion logs, and 4,503,745 answer logs.

The 93 students (3.53%) who scored 520 or more on the semester-initial TOEFL-ITP applied for exemption from the CALL course. Additionally, there were 39 dropout students (1.48%), who did not access the learning materials for a whole semester. Thus, the exempted students and dropout students were removed, and the remaining 2,499 students (94.99%) were analyzed in the following section.

3. Research design and method

Figure 1 illustrates the data preparation, analysis, and feedback procedures of this study.

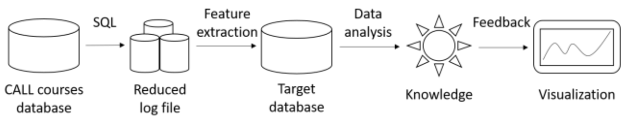

Figure 1. Data preparation, analysis, and feedback procedure

First, Search Query Language (SQL) queries were conducted to retrieve a variety of data from the CALL course server, and then log records were saved on SQL Server 2012, a database management system.

Second, the feature extraction phase was performed using reduced log files, which were cleaned by removing all useless, irregular, and incomplete data from the original CALL course logs. For example, sometimes the end time was not recorded when a student accidentally closed the web browser, or a student might do nothing for a long time with the website left open. In these cases, the related raw data stored were removed from the log files to reflect only normal learning activities of the students. Feature items were extracted through calculating or accumulating reduced log files, including daily access items, daily access time, daily completion items, and daily answer lists.

Third, the feature items stored in the target database were analyzed and interpreted. The analysis phase included three sub-phases: (a) descriptive analysis, (b) behavioral analysis on the course level, (c) behavioral analysis on the individual level. Descriptive analysis was used with summarizing, clustering, and association rules techniques to generate an overview on the dataset. Behavioral analysis, with the course level and the individual level, was focused on learning processes. To construct learning patterns and preferences, the learning progresses, active time distributions, and behavioral tendencies were shown.

Finally, in order to provide prompt feedback to students and instructors, a set of visualizations of student activities were developed. The visualizations were easy to use on a variety of web browsers, since they were developed with JavaScript, which was a major browser scripting language.

4. Analysis result and visualization

In this section, an overview of learning behaviors of all students was provided using descriptive analysis method. Then the visual learning patterns were shown and discussed, including learning progresses, active time distributions, and behavioral tendencies. In order to support instructions of the instructors and learning of the students, the analyses and discusses were conducted on the course level and the individual level.

4.1 Overview

Table 3 summarizes the ratio of task completion at four stages (N=2499). The respective ratio of task completed at stage 1, stage 2, and stage 3 was 98.08%, 93.20%, and 88.52%. In contrast, that at stage 4 significantly declined to 42.70%, since the learning materials of stage 4 were not required, but encouraged, to complete. The results reveal that most of the students completed tasks at the required stages but they need more motivation to complete tasks at the optional stage.

Table 3. Course stages and the ratios of task completion

Stage	Total	Task completed		Not completed	
	N	N	%	N	%
1	2499	2451	98.08%	48	1.92%
2	2499	2329	93.20%	170	6.80%
3	2499	2212	88.52%	287	11.48%
4	2499	1067	42.70%	1432	57.30%

To evaluate the degrees of students' activeness in the CALL course, a comparison of total active days was conducted in Table 4. One active day represented a day on which one student accessed learning materials. About half of the students (50.82%) accessed learning materials 11 to 20 active days, and most of the students (90.88%) accessed learning materials 30 active days or less.

Table 4. Comparison of total active days

Active days	N	Percent
1~10	395	15.81%
11~20	1270	50.82%
21~30	606	24.25%
31~40	160	6.40%
41~50	48	1.92%
51~60	10	0.40%
>60	10	0.40%
Total	2499	100.00%

To investigate the quality of the learning outcomes, a scatter plot was generated. Figure 2 illustrates the relationship between total time spent on the course and quiz scores. The results reveal that the students in the area A spent less time than others on the course but still obtained high scores. In contrast, the students in the area B spent above-average time on the course but scored less than 40, thus they might need more attention and support. Moreover, a small group of the students worked hard spending more than 1,800 minutes (30 hours) on the course, but failed to achieve high scores eventually. On the basis of the results, more features of different group should be extracted, considering other variables, such as language level, motivation, self-efficiency, and so on.

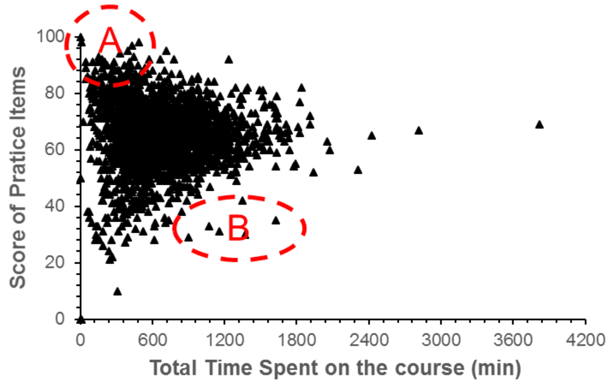

Figure 2. Scatters of total time spent on the course and quiz scores

4.2 Behavioral analysis and visualization on the course level

In order to identify learning patterns and support instructions of the instructors, the learning process of the students of one class were represented. This part would focus on feedback to the instructor of one CALL class with the visualization of learning activities, since it was not easy to provide appropriate instructions without detecting the processes of SRL.

Figure 3 shows the learning progress in one class (N=53) and each curve represents one student. The horizontal axis shows the dates and the vertical axis displays the counts of completed quiz units. If a line rises steeply, it means that the student worked very hard during that period. If the line stays flat, the student did not work much during the period. The green line means the class average of learning progress.

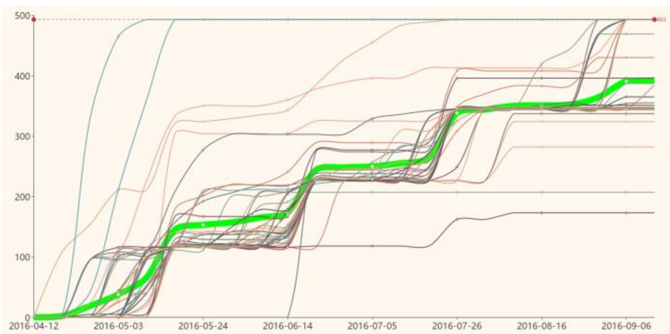

Figure 3. Learning progress in one class

As shown in Figure 3, three typical learning patterns were discovered: action just before the deadline, quit/drop out after the first deadline eventually, and complete task far before the deadline. The students with these learning patterns might not use effective strategies, such as goal setting or strategic planning.

By focusing on activeness of one class, weekly active time distributions were generated in Figure 4. The horizontal axis shows the hours, and the vertical axis displays the days of the week from Wednesday to Tuesday, since Tuesday is the last day of weekly cycles of each learning stage. The color shade represents the counts of active students. The deep shade of color represents more students were learning at that hour. As indicated in Figure 4, most of the students accessed the learning materials on the evening of the last three days.

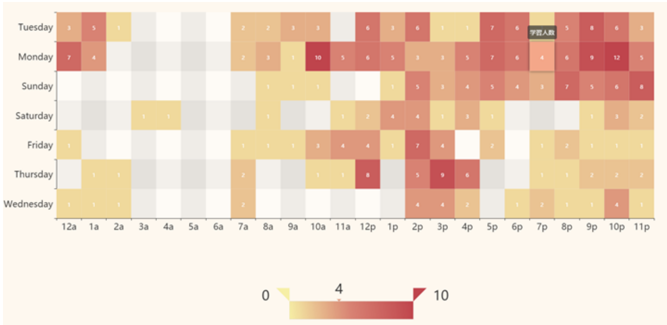

Figure 4. Weekly active time distribution in one class

Behavioral tendencies in one class are shown in Figure 5. In order to display high-dimensional data clearly, parallel coordinates plot was used. The vertical axes in Figure 5 show the student IDs, the number of completed units, the active days, the total time spent, and quiz scores. A student is represented as a polyline connecting the vertices on the vertical axes. For example, the student with 60 active days completed assigned quiz units, spent an average time, and reached an average score.

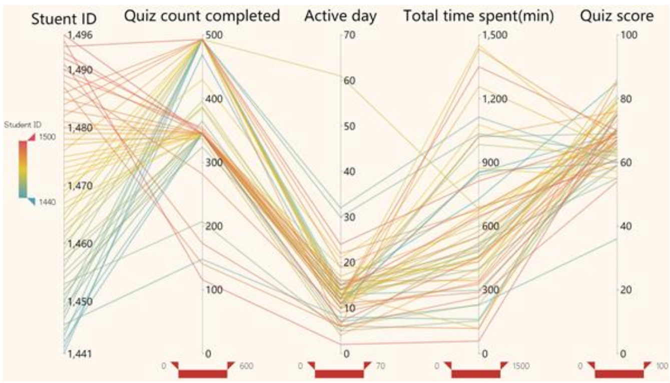

Figure 5. Behavioral tendencies in one class

4.3 Behavioral analysis and visualization on the individual level

In order to support students' self-monitoring, the learning processes of one student were represented. This part would focus on feedback to one student with the visualization of learning activities, since the freshman students often did not recognize how they are learning in the process of self-regulation.

Figure 6 shows the personal progress for each material types, a total personal progress, and a course-averaged progress. A line represents a learning material type. The red line is the total counts of completed quiz units, and the green line is the class average of learning progress. Compared with students with the class average, the represented student easily positioned himself where (s)he was, leading to increased motivation and participation.

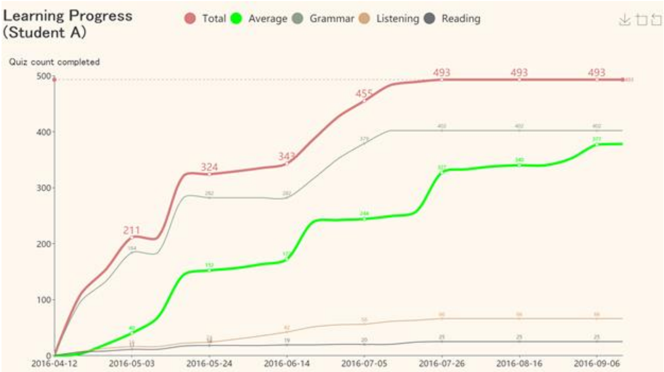

Figure 6. Personal learning progress with material types

Moreover, in order to provide an overview of individual learning activities, personal behavioral tendencies are shown in Figure 7. The vertical axes of the parallel coordinate represent the number of completed units, the active days, the total time spent, and quiz scores respectively. The results reveal that the represented student completed assigned quiz units, had average active days, spent average time, and reached a high score. On the basis of personal behavioral tendencies, the learning patterns or preferences of the students was identified.

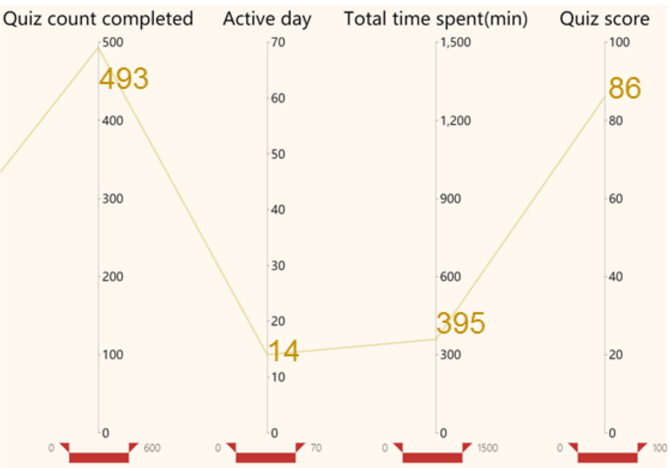

Figure 7. Personal behavioral tendencies

5. Conclusion

In order to provide appropriate feedback, this evidence-based research was conducted on a total of 7,413,397 learning logs which were retrieved from the server of CALL course. An overview of all learning behaviors in CALL classes was provided, and then the individual differences of learning process were investigated. The individual learning patterns and preferences were identified by the visualizations of actual online learning behaviors.

In current CALL environments, instructors and students are able to obtain a simple view of basic learning data, such as login time, counts of completed quiz items, etc. However, no functions or features were available to help instructors and students monitor the learning processes. Therefore, in the future research, the visualization tools will be deployed on the CALL course environments to promote students' self-monitoring and support instructors' supervision. Moreover, the evaluation of usefulness and impact of the visualization tools will be conducted.

Acknowledgments

The research has been partly supported by "Research and Development on Fundamental and Utilization Technologies for Social Big Data" (178A03), the Commissioned Research of the National Institute of Information and Communications Technology (NICT), Japan; and Grant-in-Aid for Scientific Research (S) No. 16H06304; Grant-in-Aid for Scientific Research (B) No. 25282059; Grant-in-Aid for Challenging Exploratory Research No. 26560122 from MEXT, Japan.

References

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, 64, 417-444.

Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y., & Miyagawa, H. (2015). Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes. *Learning and Individual Differences*, 37, 72-80.

Govaerts, S., Verbert, K., Klerkx, J., & Duval, E. (2010, December). Visualizing activities for self-reflection and awareness. In *International Conference on Web-Based Learning* (pp. 91-100). Springer Berlin Heidelberg.

Hung, J. L., & Zhang, K. (2008). Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*.

Ogata, H., & Mouri, K. (2015, July). Connecting dots for ubiquitous learning analytics. In *International Conference on Hybrid Learning and Continuing Education* (pp. 46-56). Springer International Publishing.

Ogata, H., Yin, C., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based Learning Analytics in University Education. In *International Conference on Computer in Education* (ICCE 2015) (pp. 401-406).

Wolters, C. A., Pintrich, P. R. & Karabenick, S. A. (2005). Assessing academic self-regulated learning. In K. A. Moore & L. Lippman (Eds), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 251–270). New York: Springer.