Profiling High-achieving Students for E-book-based Learning Analytics

Kousuke Mouri, Kyushu University, mourikousuke@gmail.com
Fumiya Okubo, Kyushu University, fokubo@artsci.kyushu-u.ac.jp
Atsushi Shimada, Kyushu University, atsushi@limu.ait.kyushu-u.ac.jp
Hiroaki Ogata, Kyushu University, hiroaki.ogata@gmail.com

Abstract: The purpose of this paper is to mine or detect meaningful learning patterns for profiling high-achieving students using e-book-based activity logs and questionnaire. The analysis of this study uses association analysis with Apriori algorithm. Logs for this analysis were collected from 99 first-year students who use a document viewer system called BookLooper, questionnaires and Moodle in an information science course at Kyushu University. From the results of the association analysis, we found that high-achieving students and BookLooer have significant relationships in terms of preparation and review time. This paper believes that the profiling and analysis can be used to predict their final grades and to detect effective learning patterns.

Keywords: Learning analytics, e-book, data mining, association analysis, user profiling

Introduction

Nowadays, majority of textbooks are not only published in printed format but also are created as electronic textbook (e-book) format available online or on mobile devices. As a Japanese government policy, they plan to introduce e-books in all K12 schools by 2020 (MEXT). Many countries' e-book policies only focus on introducing the technology of e-books into K12 schools (Fang et al., 2011), (Shin, 2012). However, little attention has been paid to analyze and mine important information for profiling from the e-book activity logs. Therefore, it is necessary to explore various analytics in this aspect.

In this paper, we call visualizing, analyzing and mining e-book activity logs "E-book-based Learning Analytics" (ELA). In such analytics, some researchers in the Kyushu University reported several analytics using a document viewer system called Booklooper (Ogata et al., 2015), (Yin et al., 2014), (Yamada et al., 2015). The objectives of their studies are as follows: (1) improving of learning materials, (2) analyzing learning patterns, (3) detecting students' comprehensive level, (4) predicting final grades, and (5) recommending e-books in accordance with personalization. This paper focuses on (2) and (4). One of the issues of (2) or (4) is how to mine meaningful learning patterns for profiling high-achieving students.

To achieve the issue, this paper describes data mining method based on ELA. The rest of this paper is constructed as follows. Section "What is BookLooper" explains the functions of BookLooper such as next page, previous page, and bookmark. Section "Data Collection" describes logs for this analysis and then how to categorize them. Section "Method" describes analysis method for profiling high-achieving students. Section "Results" describes the results of analysis, and discussion regarding high-achieving students.

What is BookLooper?

Booklooper is a commercial product designed by Kyocera Maruzen Systems Integration Co., Ltd. The system provides a cloud service. Students can download learning materials by using the BookLooper viewer. The e-books are managed in the bookshelf. If students select a book in the bookstore, the book will be downloaded into the bookshelf. The students then choose the book in the bookshelf in order to read it in the viewer. By using viewer, students can use some functions such as next page, previous page, bookmark, underline, and annotation as shown in Figure 1. For example, if a student will click button such as zoom and marker, the action will be saved into the database. In the next section, this paper describes how we categorize e-book logs accumulated in the database.

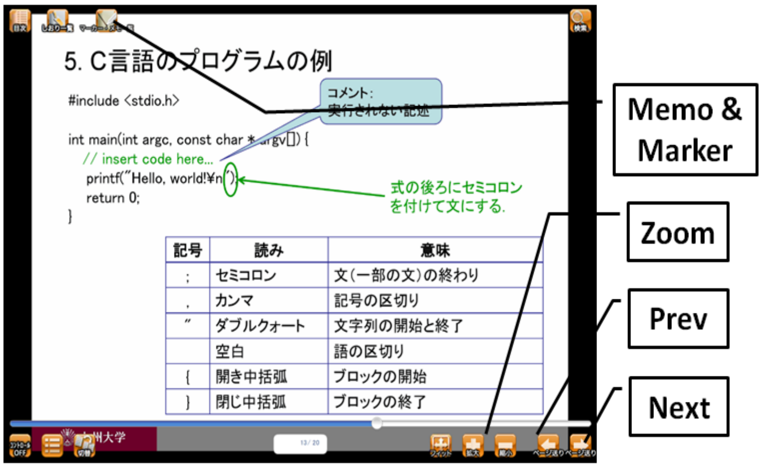

Figure 1. BookLooper interface

Data collection

Categorization of academic achievement

Logs for this analysis were collected from 99 first-year students via BookLooper and Moodle. These students took an information science course in the second semester of the 2014/2015 school year at the Kyushu University. The number of logs are collected approximately 330,000. We use Moodle to manage students' attendance, mid-semester test score, end-of-term test score, and report score. Also, BookLooper is used for collecting students' operation logs and three types of learning time of each student: Preparation Time Before Class (BTBC), Learning Time During Class (LTDC), and Review Time After Class (RTAC) using Booklooper for profiling the relationships among high-achieving students, BTBC, LTDC and RTAC because students who devoted much time to prepare and review are not necessarily good score. In addition, it is important to categorize them efficiently in order to detect or mine meaningful learning patterns for profiling high-achieving students. Therefore, we divide numerical data such as the number of attendance, lateness and absence, report scores, mid-semester test scores, end-of-term test scores, three types of learning time and final score to several categories excluding numerical data. This paper establishes criteria for categorizing them as shown in Table 1. The high-achieving students of the top 20 percent mean A rank. For example, if a student devoted much time more than 2364 seconds in order to prepare the content by the next lesson using BookLooper, we categorize the student to "BTBC = A rank".

Table 1: The criteria for categorizing the achieving rank of each student

LV	Criteria	Attendance (Scorning 30)	Report (Scoring 40)	Mid-semester (Scoring 10)	End-of-term (Scoring 20)	BTBC (seconds)	LTDC (seconds)	RTAC (seconds)
A	Top 20%	>= 23	>= 35	>= 9	>=16	>=2364	>= 32025	>=10718
B	20 ~ 40	21 ~ 23	30 ~ 35	8.5 ~ 9	14 ~ 16	676~2364	27053~32025	6705~10718
C	40 ~ 60	18 ~ 21	20 ~ 30	8 ~ 8.5	12 ~ 14	76 ~ 676s	19159 ~27053	3907 ~ 6735
D	60 ~ 80	14 ~ 18	15 ~ 20	7 ~ 8	10 ~ 12	1 ~ 76	12946 ~ 19159	785~3907
E	80~100	14>=	15>=	7>=	10>=	0	12946>=	785>=

Questionnaires

The students were required to answer questionnaires before class in order to investigate their life styles, a method and time of transportation to university, the amount of learning for one day, and satisfaction of university life. Table 2 shows the questionnaires. Q1 and Q2 ask about their life style in the morning such as breakfast and time to get up because the class of the information science course starts in the morning. Q3 and Q4 ask about their commuting method and time in order to analyze relationships among high-achieving students, commuting method and time. Q5 asks about the amount of their study time for one day. Q6 asks about satisfaction of their university life. In the next section, this paper describes how to mine meaningful learning patterns for profiling high-achieving students using these data as described Sections titled "Categorization of academic achievement" and "Questionnaires".

Table 2: Questionnaires

	Question items	Answer items
Q1	What time do you get up?	(1)before am 5:00 (2)am 5:00~6:00 (3)am 6:00~7:00 (4)am 7:00~8:00 (5) am 8:00~9:00 (6) after am 9:00
Q2	Do you eat breakfast every day?	(1)Yes (2) No
Q3	What do you use a method of transportation to university?	(1)on foot (2)bicycle (3)car (4)public transport
Q4	How many do you take to university?	(1)less than 30 minute (2)30~60 minute (3)60~90 minute (4)90~120 minute (5)more than 120 minute
Q5	How much time do you study for one day	(1)more than 3 hours (2)2~3 hours (3)1~2 hours (4)less than 1 hours
Q6	Do you feel that university life is fan?	(1)Extremely well (2)Very well (3)Moderately well (4) Slightly well (5) Not at all well

Methods

Data mining based on e-book-based learning analytics

In order to mine meaningful learning patterns for profiling high-achieving students, this paper uses an association analysis with Apriori algorithm. Association analysis is one of the popular analysis methods in order to mine regularities between some parameters of educational big data. For example, Mouri et al. (Mouri et al., 2015) use association analysis for mining useful learning patterns from learning logs accumulated in ubiquitous learning system called SCROLL. The objective of SCROLL is to support international students to learn learning object in Japanese in an informal setting. In addition, they believes that visualizing and analyzing them collected by SCROLL lead to enhancing students' learning activities in an informal setting. Unlike Informal Learning Analytics (ILA) or Ubiquitous Learning Analytics (ULA) of their focus, this paper focuses on analyzing logs collected in a formal and an informal setting. The analysis of this paper was conducted the following those criteria: Support ≧ 0.3, Confidence ≧0.6, Lift ≧1.0. The objective of the setting value is to detect many association rules as far as possible. The number of the detected association rules is 51,641. In order to find meaning learning patterns for profiling high-achieving students, this study mines association rules that the conclusion parts are score A rank as described in section titled "Categorization of academic achievement".

Results

Profiling and discussion

In order to find the relationships between high-achieving students and the effectiveness of BookLooper, and high-achieving students and the questionnaires as shown in Table 2, this paper investigates the association rules that the conclusion parts are "report score is rank A", "mid-semester test score is rank A", "end-of-term test score is rank A" and "final score is rank A". We found important some association rules shown in Table 3.

The rules from 1 to 5 show that the conclusion part is report score A rank. The "BTBC=A" of the rule 1 means that students devoted much time more than 2364 seconds in order to prepare by the next lesson. The relationships between "BTBC=A" and high-achieving students have a high relativity because the confidence value of the rule 1 is 1. The rule 2 and 5 show that the condition parts are "Q1= (3) && Q4= (1)" and "Q2= (1) && Q4= (1)". This means the commuting time of high-achieving students to university is less than 30 minutes. In addition, they get up early in the morning and eat breakfast every day.

The rules from 6 to 10 show that the conclusion part is mid-semester test score A rank. The rule 6 and 8 show that the condition parts are "attendance = A && report=A" and "attendance = A && Q1= (3)". In order to achieve the mid-semester test score A, it indicates that it is important to get attendance sore more than 23 points. The rule 9 and 10 shows that the relationships between "report=A && Q4= (1)" and mid-semester test score, and "report=A && Q2= (1)" and mid-semester test score.

The rules from 11 to 15 show that the conclusion part is end-of- term test score A rank. The "LTDC =A" of the rule 12 means that students devoted much time more than 32025 seconds using BookLooper during class. In addition, the "RTAC = A" of the rule 13 means that students devoted much time more than 10718 seconds using BookLooper in order to review the content after class. That means that it is important to achieve the conditions of "RTAC=A" and "LTDC=A" if students want to get the end-of-term test score A rank.

The rules from 16 to 20 show that the conclusion part is final score A rank. The rule 16 and 17 means that the condition part is “BTBC=A” and “RTAC=A”. That means that it is important to achieve the two conditions if students want to get final score A rank. Conversely, if students have “BTBC=E” or “LTDC=E”, Most of them got final score E rank as shown in Figure 2. Therefore, there is a possibility that the profiling high-achieving students lead to discoveries of students who fail to make the grade.

Table 3: The association rules among high-achieving students, BookLooper and questionnaires

No	Condition part	Conclusion part	Support	Confidence	Lift
1	BTBC=A	report=A	0.306592	1	1.1124948
2	Q1=(3) && Q4=(1)	report=A	0.401229	0.9571007	1.0647695
3	Q6=(1)	report=A	0.4007834	0.8488891	1.0443846
4	Q5=(3)	report=A	0.3476144	0.934392	1.0395062
5	Q2=(1) && Q4=(1)	report=A	0.3178796	0.9464546	1.0529258
6	attendance=A && report=A	mid-semester test score =A	0.301626	0.636362	1.0298533
7	Q4=(1)	mid-semester test score =A	0.3316247	0.6046827	1.0437379
8	attendance=A && Q1=(3)	mid-semester test score =A	0.3099994	0.6178257	1.0794652
9	report=A && Q4=(1)	mid-semester test score =A	0.3406835	0.7807266	1.1487107
10	report=A && Q2=(1)	mid-semester test score =A	0.3068324	0.8832242	1.1863906
11	Q1=(3) && Q2 =(1)	end-of-term test score=A	0.3007997	0.8368737	1.319224
12	LTDC=A	end-of-term test score=A	0.313928	0.6541544	1.1640735
13	RTAC=A	end-of-term test score=A	0.3313667	0.8179246	1.255504
14	report=A &&LTDC=A	end-of-term test score=A	0.3049478	0.6906759	1.2290638
15	attendance=A report=A LTDC=A	end-of-term test score=A	0.3049478	0.6906759	1.2290638
16	BTBC=A	final score=A	0.4007834	0.8488891	1.0443846
17	RTAC=A && Q1=(3)	final score=A	0.3476144	0.934392	1.0395062
18	Q1=(3) && Q2=(1)	final score=A	0.306592	1	1.1124948
19	mid-semester=A && end-of-term=A	final score=A	0.401229	0.9571007	1.0647695
20	Q4=(1)	final score=A	0.4007834	0.8488891	1.0443846

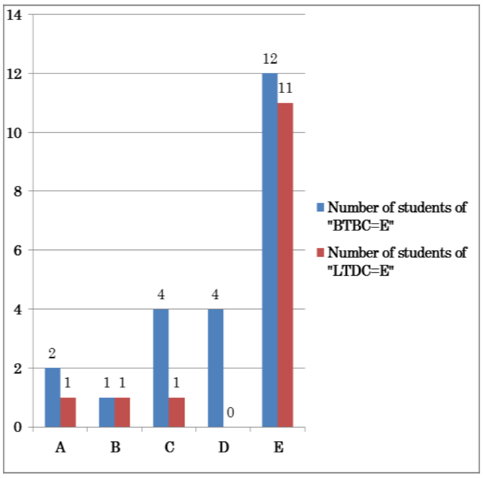

Figure 2. The number of student of “final score = E rank”: The blue bar shows BTBC = E rank”, the red bar shows “LTDC = E rank”

Conclusion

This paper describes how to mine or detect meaningful learning patterns for profiling high-achieving students using e-book-based activity logs. In order to mine the learning patterns, this paper uses association analysis with Appriori algorithm. The analysis was conducted to find the relationships between high-achieving students and the effectiveness of a document viewer system called BookLooper as shown in Table 2, and high-achieving students and the questionnaires as shown in Table 3. In addition, this paper investigated the association rules that the conclusion parts are "report score is rank A", "mid-semester test score is rank A", "end-of-term test score is rank A" and "final score is rank A". In the future, we will consider supporting students who fail to make the grade using the detected association rules. Also, we will consider visualizing various methods such as social network analysis (Ogata et al., 2015) and visualization of graph theory (Mouri et al., 2014), and then develop system for recommending to the personal learner in accordance with the detected results. In addition, we will integrate e-book and SCROLL with task-based learning called Learning Log Navigator (Mouri et al., 2013) in order to enhance learning experience.

References

Fang, H., Liu, P., Huang, R. (2011). The Research on E-book-oriented Mobile Learning System Environment Application and Its Tendency, *International Conference on Computer Science and Education* (pp. 1333-1338). IEEE.

MEXT, Japanese Ministry of Education, Culture, Sports, Science and Technology, "The Vision for ICT in Education",http://www.mext.go.jp/b_menu/houdou/23/04/__icsFiles/afieldfile/2012/08/03/1305484_14_1.pdf

Mouri, K., Ogata, H., Li, M., Hou, B., Uosaki, N. and Liu, S. (2013). Learning log navigator: supporting task-based learning using ubiquitous learning logs, *Journal of Research and Practice on Technology Enhanced Learning*, Vol.8, No.1, pp.117-128.

Mouri, K., Ogata, H.. Uosaki, N. and Liu, S. (2014). Visualizing ubiquitous learning logs using collocational networks, *Proceedings of the 22nd International Conference on Computers in Education* (pp.685-693).

Mouri, K., Ogata, H. and Uosaki, N. (2015). Analysis of Ubiquitous-Learning Logs Using Spatio-Temporal Data Mining, *Proceedings of the 15th IEEE International Conference on Advanced Technologies* (pp. 96-98). IEEE.

Mouri, K., Ogata, H. and Uosaki, N. (2015). Ubiquitous Learning Analytics in the Context of Real-world Language Learning, *International conference on Learning Analytics and Knowledge 15* (pp.378-382). ACM.

Mouri, K., Ogata, H. (2015). Ubiquitous Learning Analytics in the Real-world Language Learning, *Smart Learning Environments*, Vol.2, No.15, pp.1-18.

Ogata, H., Hou, B., Li, M., Uosaki, N., Mouri, K and Liu, S. (2014). Ubiquitous Learning Project Using Life-logging Technology in Japan, *Educational Technology and Society Journal*, 17(2), 85-100.

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

Ogata, H., Mouri, K. (2015). Connecting Dots for Ubiquitous Learning Analytics, *International Conference on Hybrid Learning* (pp.46-56).

Shin, J. A. (2012). Analysis on the digital textbook's different effectiveness by characteristics of learner, *International Journal of Education and Learning*, Vol.1, No.2, pp.23-38.

Yamada, M., Yin, C., Shimada, A., Kojima, K., Okubo, F. and Ogata, H. 2015. Preliminary Research on Self-regulated Learning and Learning Logs in a Ubiquitous Learning Environment, *International Conference on Advanced Learning Technologies* (pp. 93-95). IEEE.

Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S. and Ogata, H. 2015. Identifying and Analyzing the Learning Behaviors of Students using e-Books, *International Conference on Computer in Education 2015* (pp. 118-120).

Acknowledgments

This part of this research was supported by the Grant-in-Aid for Scientific Research No.25282059, No.26560122, No.25540091 and No.26350319 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan. The research results have been partly achieved by "Research and Development on Fundamental and Utilization Technologies for Social Big Data" (178A03), the Commissioned Research of National Institute of Information and Communications Technology (NICT), Japan.