Evidence Mining Using Course Schedule

AUTHOR(S):

NAKANISHI, TARO; KUROMIYA, HIROYUKI; MAJUMDAR, RWITAJIT; OGATA, HIROAKI

Evidence Mining Using Course Schedule.

TARO NAKANISHI[1] HIROYUKI KUROMIYA[1]
RWITAJIT MAJUMDAR[2] HIROAKI OGATA[2]

Abstract: Creating evidence from learning big data has become increasingly important as we can use eLearning infrastructure and store learning log digitally. On the other hand, we need to time and effort to create evidence because it is manual. In this paper, we proposed the method to make evidence easier. Especially, we focus on procedure to automatically select the duration of intervention and comparison data based on the course schedule information. We simulated the procedure and confirmed the making a case based on course schedule information. In the discussion part, we mentioned the points that should be further improved for practical use in the future. Through our method, we will democratize the evidence-based practice to all the teachers in schools.

Keywords: Learning Analytics, Learning and Evidence Analytics Framework (LEAF), Course schedule, Evidence-Based Education

1. Background

Organizing course schedules at school is an important proposition in data utilization. In Japan, the government [1] (2020) is considering introducing unit ID to grasp and evaluate learning outcomes. In learning analytics, it is essential to match the data with the data corresponding to the course schedule in order to estimate the existence of learning based on the learning log. So, we looked at the relationship between evidence-based education and course schedule.

For evidence-based education, Davies (1996) [2] defined the concept of evidence-based education practices. evidence-based education means integrating individual teaching and learning expertise with the best available external evidence from systematic research.

In conventional practice, it is difficult to apply the concept of evidence-based education because it takes time and effort to review a paper and conduct experiments. In addition, it was difficult to use the evidence because teachers don't have time to look up past evidence although there are some websites that aggregate and share the past evidence published in the educational field.

In this study, we aim to make it easier to make evidence. We proposed a method to automatically select a data log for analysis based on the course schedule information.

2. Related Works

For example, What Works Clearinghouse (WWC) [3] is a website established by the Institute of Education Sciences at the U.S. Department of Education under the current American educational reform. It provides educators, policy makers, researchers, and the public with a source of scientific evidence on effective education programs and practices. Education Endowment Foundation [4] was Established in 2011 by the UK Department for Education. It shares useful evidence to practice.

LACE Evidence Hub [5] is a website to gather evidence in learning analytics. These allow us to aggregate and share evidence from existing research and learning analytics. Researchers and educators are now able to look at aggregated evidence. On the other hand, they require manual input of evidence. Registering evidence requires a lot of time and effort by users.

To solve the problem, in the learning analytics field, Learning and Evidence Analytics Framework (LEAF) is proposed [6 7]. LEAF compares the intervention period data with the control

3. Proposed Method

3.1 Learning and Evidence Analytics Framework (LEAF)

We had proposed an evidence extraction system called LEAF (Learning and Evidence Analytics Framework). This system integrated into the learning analytics platform as a solution to extract and store evidence from practice [8]. LEAF consists of five components: Learning Behavior Sensors, Learning Record Store, Learning Analytics Dashboard, Evidence Engine, and Case Record Store. Learning Behavior Sensors collect learner learning data. It uses a learning management system Moodle and the e-book reader BookRoll [9]. BookRoll is a learning tracker and e-book reader to make a seamless learning environment. The system makes it easy to facilitate many learning analytics studies [10]. Learning Record Store accumulates data from Learning Behavior Sensors. Learning Analytics Dashboard visualizes the analyzing result of accumulated data. Evidence engine generates a case from the learning logs in the system and aggregates cases in the Case Record Store. it also supports users to make decisions. In this paper, we focus on the function of generating a case about Evidence Engine. We propose that the system automatically selects the duration of intervention data and comparison data based on the course schedule information when it generates a case.

1 Kyoto University, Japan
2 Academic Center for computing and Media Studies, Kyoto University, Japan

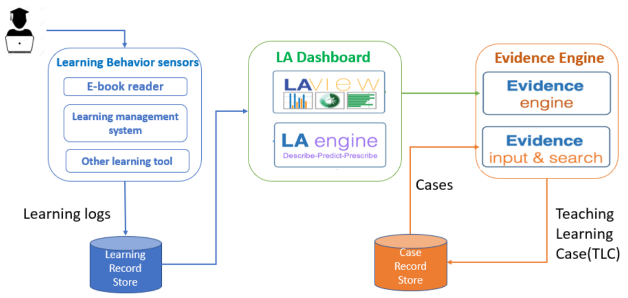

Figure1 The five components in LEAF.

3.2 Course schedule

In this paper, we use the course of study format from MEXT as the course schedule representation. This is a common code for the content and units of course of study. It was established by the government to enhance information technology in education.

Currently, the code doesn't contain teaching activity information because it is not only for learning analytics but also for general usage of electric materials. However, it will be important when we try to validate a specific teaching strategy. Then, we proposed a revised format of course schedule (Table1). It has three elements: Date, Study Code, and Teaching Strategy. The date represents the time when the learning took place. The Study Code represents the unit information students have studied. The Teaching Strategy represents the teaching strategies used in the classroom.

Table 1 Proposed structure of course schedule.

Date	Study Code	Teaching Strategy
2020/06/15	8454503200000000	Lecture-based teaching
2020/06/16	8454503210000000	Inquiry instruction
2020/06/17	8454503211000000	Lecture-based teaching
2020/06/18	8454503212000000	Lecture-based teaching
2020/06/19	8454503213000000	Lecture-based teaching
2020/06/19	8454503214000000	Lecture-based teaching
2020/06/23	8454503215000000	Lecture-based teaching
2020/06/24	8454503220000000	Group activity

3.3 Using course schedule to compare the teaching strategy

Multiple records are needed to compare the efficacy of teaching strategies. Currently, the computing process requires multiple parameters every time and requires knowing the period of time to be compared with the period of intervention. Therefore, comparing the effectiveness of the interventions is time-consuming. Figure 2 shows the workflow for comparing the teaching strategies using the course schedule.

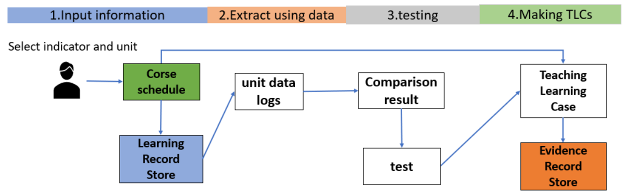

Figure 2 Proposed flow of using course schedule in Evidence Engine.

Step 1: Input information
First, we input the learning unit and indicator in the system. The learning unit we entered will be checked against the learning unit in the course schedule.
Step 2: Match the most relevant data
The log data in the Learning Record Store is to be extracted by referring to the dates and unit information. If the unit information matched, the system automatically retrieves the data during that period.
Step 3: Testing
Unpaired t-test is performed to check the significance between the intervention and comparison. In this paper, the significance level is set at 5%. The p-value is passed to TLC in the next step.
Step 4: Making a Teaching Learning Case (TLC)
TLC is primarily a single data point regarding the result of an intervention. It is created from comparison results and course schedule information. Structure of TLC using the course schedule information is displayed in table2. Context is the information regarding the context of evidence. Learning unit information is retrieved from input data. All other information is automatically retrieved from LMS. Problem means problem addressed in the classroom. For example, low engagement with homework materials can be a motivation for the intervention. Indicator is a measurable indicator that a user wants to analyze. Intervention Activity means the details of the intervention conducted by a teacher. It describes the activity of interventions performed. Comparison activity means details of the comparison data. The format is the same as the intervention activity. Result represents analysis results of comparing data.

Table 2 Structure of TLC.

Factor	Example

Context
study code:
subject:
learning unit:
intervention
context: {
school name:
grade:
class size:
dates:
}
	study code:
	XXXXXXXXXXXXXXXX
	subject: Math,
	learning unit: XXX,
	intervention context: {
	school name: XXX high school,
	grade: First grade,
	class size: 120,
	dates: “2019-05-01,” “2019-05-02,” ...
	},
	comparison context: {

comparison context: { school name: grade: class size: dates: }	school name: XXX high school, grade: First grade of high school, class size: 120, dates: "2019-04-05," "2019-04-06," ... }
Problem	Low engagement with homework materials
Indicator	Reading time on materials
Intervention Activity:	Activity: "do some group activity"
Comparison Activity:	Activity:" lecture only"
Results	Group activity intervention increased the reading time by 5.5 min (p = .01)

4. Simulation

Based on our proposed method in the previous section, we simulated whether the system could automatically select the data. Furthermore, we checked if the system could make TLC from extracted data. In this paper, the experiment was conducted using simulated data (not actual classroom data).

4.1 Simulating the data

We received the course schedule from a teacher at a Japanese public high school. We transformed it into a form that the system can use (see Table 1). It captures the date, code of learning unit, teaching strategy. We simulated dummy data using the log data from the school during this year. We assumed that in class A, the instructors did the group learning activity, and in class B, they did traditional teaching activity. Assuming that the indicator value increases with group activities, we added Gaussian noise with an average of 0.3 and a standard deviation of 0.03 to the class A data. We selected that the learning unit was the same in both classes, and the indicator was time spent on the teaching material.

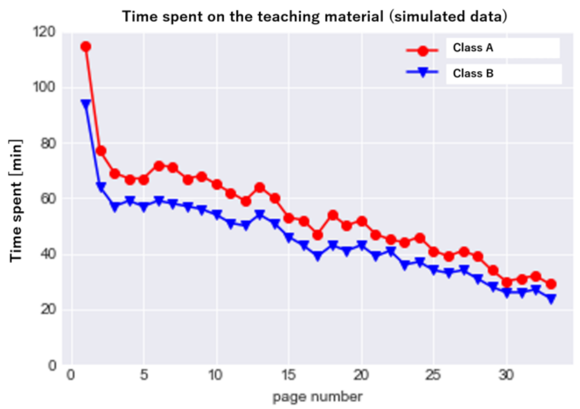

Figure 3 Time spent on the teaching material (simulated data).

4.2 Matching and comparison

4.2.1 Matching

We input an indicator and learning unit. Then the system matches the code of study with the course schedule. The system searches for the period when the code matches and extracts the log data for that period from the Learning Record Store (LRS).

4.2.2 Comparison result

When comparing different teaching methods, it is difficult to re-perform the same unit in the same class with different teaching methods. Therefore, it is necessary to use data from different classes and different years when making comparisons. Since it is not possible to perform a pre/post comparison, we perform the unpaired t-test. In order to apply the unpaired t-test, the sample must meet the following conditions. 1. The sample normality is satisfied. 2. The homoscedasticity of the variance is satisfied. First, perform the Shapiro-Wilk test to confirm that the condition 1 is satisfied. Then, perform the Levene test to confirm whether the condition 2 is satisfied. In both cases, the superiority level is 5%. If the null hypothesis is not rejected in either test, the system terminates the creation of the TLC, assuming there is no valid data to compare. If the conditions 1 and 2 are satisfied, the unpaired t-test is performed. T-test creates a TLC regardless of whether the result is significant or not. Lastly, we noted the result of statistical testing in the TLC record.

In this case, We confirmed the normality and homoscedasticity of the data. Table 3 shows the results of t-test in our simulated data.

Table 3 The result of t-test in our simulated data.

p-value	t-value	Difference in average value	Standard error
0.027	2.27	9.00	3.97

4.2.3 TLC record creation

The system created a TLC from the course information and the comparison results. Table 4 shows the created TLC. It was confirmed that the TLC could be created with the structure of the proposed method. Finally, the TLC was registered in the Evidence Record Store in LEAF.

Table 4 Created TLC in our simulated data.

Factor	Example
Context study code: subject: learning unit: intervention context{ school name: grade:	study code: XXXXXXXXXXXXXXXX subject: Math, learning unit: XXX, intervention context:{ school name: XXX high school, grade: First grade, class size: 120,

class size:
dates:
}
comparison context{
school name:
grade:
class size:
dates:
}
	dates: “2019-05-01,” “2019-05-02,” ...
	},
	comparison context:{
	school name: XXX high school,
	grade: First grade of high school,
	class size: 120,
	dates: “2019-04-05,” “2019-04-06,” ...
	}

Problem	Low engagement with homework materials
Indicator	Reading time on materials
Intervention Activity:	Activity: “do some group activity”
Comparison Activity:	Activity:”only lecture”
Results	Group activity intervention increased the reading time by 5.5 min (p = .01)

5. Discussion

5.1 Findings

In this simulation, we confirm the series of steps to create a TLC. The course schedule lets the system know the dates and study code teaching strategies. Therefore, the system can bring the matched log data from the LRS based on the input unit and course schedule. In the unpaired t-test part, we found significant differences in the results of the simulated data. In this simulation, we were able to make TLC based on only the learning unit and indicator.

5.2 Limitation and future work

However, in our method and simulation, we have some problems in the method structure and system. In the method structure, we don’t think about using multiple sets of data. Currently, the system can deal with only one intervention data and one comparison data. When we use real-world data, there may be more than one set of data that meets the criteria. we need to consider how to process when the system finds multiple matching data. To solve this problem, we plan to calculate the Euclidean distance of the data similarity. Each log data has context. We’ll use them as a parameter to calculate similarity. After calculation, the system chooses the comparative data most suitable for intervention data.

In this paper, we proposed the procedure of evaluating the intervention based on the course schedule and code of study. We didn’t automate the creation of TLC. We haven’t implemented a system to automatically select the duration of intervention and comparison data based on the course schedule information.

In addition, we need to consider the usage of code of study. We used code of study to represent the course schedule. Each digit in the code of study has meaning (see figure 4). Although we chose the code of study in this simulation, it is desirable to identify the information for each digit using the regular expression in order to automate the process of adding context information for TLC generation. That also allows us to use other information from code of study.

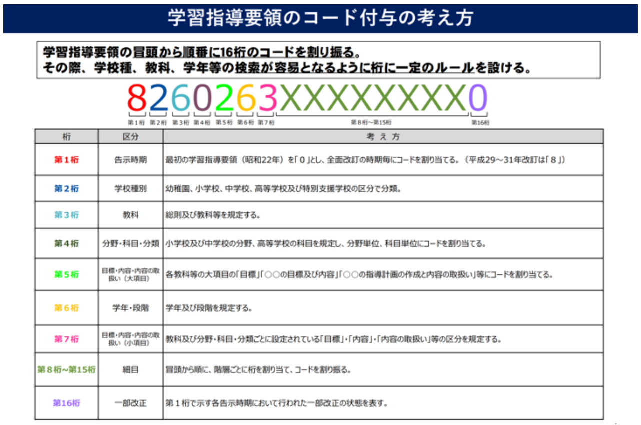

Figure 4 Meaning of each digit in code of study [11].

6. Conclusion

In this paper, we proposed the method of mining evidence easier based on course schedule information. Based on the method, we showed a series of steps to create a TLC using simulated data. We reduced the input parameters and the system can automatically extract matched data from the LRS because course schedule contains the intervention and comparison information.

As a future work, we need to conduct experiments with real data. So, we should consider what to do when there are multiple comparison data. Later, we plan to automate the steps as well so that we can create TLC automatically. We will also consider usage of code of study in the future.

Reference

[1] 文部科学省, (2020), “新時代の学びを支える先端技術活用推進方策(最終まとめ) [PDF file]”, Retrieved from https://www.mext.go.jp/component/a_menu/other/detail/__icsFiles/afieldfile/2019/06/24/1418387_01.pdf

[2] Davies, P. (1999). What is evidence‐based education? British journal of educational studies, 47(2), 108-121. Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. Educational researcher, 31(7), 15-21.

[3] Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self-management literature base. Remedial and Special Education, 34(1), 44-58.

[4] Higgins, S., Xiao, Z., & Katsipataki, M. (2012). The impact of digital technology on learning: A summary for the education endowment foundation. Durham, UK: Education Endowment Foundation and Durham University.

[5] Bienkowski, M., Feng, M., and Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of Education, Office of Educational Technology.

[6] Ogata, H., Majumdar R., Gökhan, H., Mohammad, N. Flanagan, B. Beyond Learning Analytics:Framework for Technology-Enhanced Evidence-Based Education and Learning, Wu,Y.-T. et al. (Eds.)

(2018). Workshop Proceedings. 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education.

[7] Kuromiya, H., Majumdar, R., & Ogata, H. (2020). Fostering Evidence-Based Education with Learning Analytics: Capturing Teaching-Learning Cases from Log Data. Educational Technology & Society, 23 (4), 14–29.

[8] Majumdar, Rwitajit & Akçapınar, Arzu & Akçapınar, Gökhan & Flanagan, Brendan & Ogata, Hiroaki. (2019). Learning Analytics Dashboard Widgets to Author Teaching-Learning Cases for Evidence-based Education.

[9] Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In Proceedings of the International Conference on Computer in Education (ICCE 2015) (pp. 401–406). China: Asia-Pacific Society for Computers in Education.

[10] Mouri, K., Uosaki, N., & Ogata, H. (2018). Learning analytics for supporting seamless language learning using e-book with ubiquitous learning system. *Educational Technology & Society, 21*(2), 150–163.

[11] 文部科学省初等中等教育局学びの先端技術活用推進室. (2020). “学習指導要領コードについて” [PDF file]. Retrieved from https://www.mext.go.jp/content/20201016-mxt_syoto01-000010374_2.pdf