Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

Visualization and Prediction of Learning Activities by Using Discrete Graphs

Fumiya OKUBO*, Atsushi SHIMADA, Chengjiu YIN & Hiroaki OGATA
Faculty of Arts and Science, Kyushu University, Japan
*fokubo@artsci.kyushu-u.ac.jp

Abstract: This paper presents a method for visualizing students' learning logs using discrete graphs. These logs contain the following four items: attendance, time spent browsing slides, submission of a report and the quiz score for each lesson. The data were collected using learning management systems and the e-text systems. By using these data, we construct graphs for each grade of which the nodes represent all combinations of achievements and failures for the four items. The graphs enable us to observe the features of students' learning activities for each obtained grade. The order in which the above four items are presented changes the visual features of the graph. Moreover, the construction of a graph from the data of the same class held previously enables us to inform students of the learning activities they should avoid. Finally, future research plans regarding this method are presented.

Keywords: Learning log analysis, information visualizing, discrete graph

1. Introduction

The last few decades have seen intense developments in data mining technologies in many disciplines with the aim of extracting interesting patterns or features from a large amount of collected data. Such an investigation is also important for the research field of education, because the use of ICT-based educational systems, such as a learning management system (LMS) or an e-book system, has become widespread. These systems record many kinds of log data that correspond to the students' learning activities. Analysis of the collected data enables teachers to observe trends in students' learning behaviors (Baradwaj, B. & Pal, S. (2011)).

At Kyushu University, the well-known LMS Moodle and the e-book system called BookLooper were introduced in October, 2014. This e-book system has the following features: it records detailed action logs, such as moving back and forth between pages, the contents of memos, the kind of access device (PC or smartphone), etc., together with the user id and timestamp. Initial research results obtained by analyzing students' learning behaviors using these logs were published in Yin, C-J. et al. (2014).

In the field of education, an important application of data mining would be to target students who are likely to fail or drop out of class, i.e., students refered to as "at-risk" students. Methods for the early detection of at-risk students from data were investigated by Wolff, A. et al. (2014) and Hlosta, M. et al. (2014). Particularly, the method based on a discrete graph (more precisely, based on a Markov chain) introduced by Hlosta, M. et al. (2014) is interesting, because it allows us to visually observe students' behavioral patterns.

In this work, we attempt to visualize learning logs stored in the LMS and the e-book system, such as attendance, the submission of reports by students, the quiz scores in the LMS and the time spent browsing slides in the e-book system. For this purpose, the model based on discrete graphs introduced by Hlosta, M. et al. (2014) is modified and extended. The modified method enables us to construct graphs to visualize learning activities of those students who obtained each grade. These graphs are useful to determine the difference in the approach to learning activities for each grade, and to predict a next student's activity.

This paper is organized as follows. Section2 introduces the methods that were used to collect the necessary logs and to use the data to construct discrete graphs to visualize students' learning activities. The experimental results are presented in Section 3. Then, in Section 4, we briefly propose the approach we followed to predict learning activities by using the constructed graph. Finally, the concluding remarks, including future work, are given in Section 5.

2. Method

2.1 Data Collection

We collect learning logs from the designated class, in which students use the LMS and the e-book system through 14 lessons and in preparation/review for the lessons. Each of these lessons is presented so as to accompany several slides in the e-book system, although each slide is assigned to just one lesson. Using the slides, students complete a preparation and a review session before and after each lesson, respectively. In each lesson, students are required to submit a report and to answer a quiz, with each quiz containing three to five questions related to the lesson and conducted in the beginning of a lesson. Hence, the LMS and the e-book system contain the following logs for each student participating each lesson:

(i) attendance (or absence)
(ii) a sum of the time spent for browsing slides for preparation and review
(iii) the submission of a report by a student or the failure to do so
(iv) the quiz score

The class is graded by A, B, C, D or F in the usual manner, with A being the best grade and F indicating failure.

2.2 Construction of discrete graphs from data

Based on the data shown in Section 2.1, we construct discrete graphs to visualize the learning activities of the students in a class.

Here, we assume that item (ii) is achieved if a student spends more than 10 minutes browsing slides, and that item (iv) is achieved if a student scores more than 70% in the quiz. For each item, whether a student achieved an item can be represented by a bit, where 0 means an achievement and 1 means a failure. Hence, in each lesson, student's achievements of the four items can be represented by 4 bits. We call this 4 bits of (i), (ii), (iii), (iv) from a lower bit, by state number. For example, if a student achieves (i) and (iii), but fails to achieve (ii) and (iv) of the above items, then the state number "0101" is assigned. In Table 1 lists the relation between the four items and the state numbers, for which decimal instead of binary numbers are used.

Table 1: A relation between four items of learning logs and the state numbers

State number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(i) Attendance	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
(ii) Browsing time	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
(iii) Report	-	-	-	-	+	+	+	+	-	-	-	-	+	+	+	+
(iv) Quiz score	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+

The nodes of the graphs consist of these 2^4=16 kinds of states per a lesson, i.e. 224 states. For the sake of visibility, the nodes are arranged in 14 columns of lessons from left to right and in 16 rows of state numbers from bottom to top. For example, the state in the i-th column and the j-th row is of state number j-1 in the i-th lesson.

An edge between a state x of the i-th lesson and a state y of the (i+1)-th lesson is constructed if there exists a student whose activities correspond to states x and y. The edge is colored (light) yellow if only one student meets the condition. As the number of people who meet the condition increases, the color of the edge approaches to orange.

We construct the six graphs from the learning logs of all the students and of those who obtain the grade A, B, C, D, or F.

This method for constructing discrete graphs can easily be customized by changing the number of items (hence, the number of states) or the order in which the items occur. In this work, we only consider the latter case. The order in which the items occur is important, because it influences the relation between the state numbers and the items, hence, and also changes the visual appearance of the constructed graph. Generally, the last of the four items is rather important, because the fourth item is assigned the value 1 in the states 8 to 15, i.e., the upper half of the constructed graph. Hence, we can easily observe the effects of the fourth item on the grade, by only observing the upper half (or the lower half) of the graphs.

For comparison with the case in Table 1, we consider the four items in the order (iv), (iii), (ii), and (i) for which the corresponding states are listed in Table 2.

Table 2: A relation between the four items ordered by (iv), (iii), (ii), (i) and the state numbers

State number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(iv) Quiz score	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
(iii) Report	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
(ii) Browsing time	-	-	-	-	+	+	+	+	-	-	-	-	+	+	+	+
(i) Attendance	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+

3. Experimental Results

We applied the proposed method to the learning logs of about 100 students attending an “information science” class that started in October, 2014. Discrete graphs were constructed from the learning logs of all students, students who obtained the grade A, B, C, D, or F in these classes, respectively.

Figures 1 and 2 show the graphs that were constructed from the four items ordered by (i) attendance, (ii) browsing time, (iii) report, and (iv) quiz score for all the students and those who obtained the grade F, respectively. In Figure 1, it can be seen that orange edges appeared for the states 13 and 15 from the 4^{th} lesson to the 8^{th} lesson. Here, the state 13 means a student achieved (i), (iii), and (iv), and the state 15 means a student achieved the four items. Hence, Figure 1 indicates that the students who achieved (i), (iii), and (iv) were likely to continue to achieve these 3 items to the 8^{th} lesson. Moreover, we can see that the 9^{th} lesson was canceled. In Figure 2, many edges appear in the lower half of the graph for grade F, as students who achieved a poor grade were unlikely to achieve the four items. Conversely, a good grade produced many edges in the upper half of the graph.

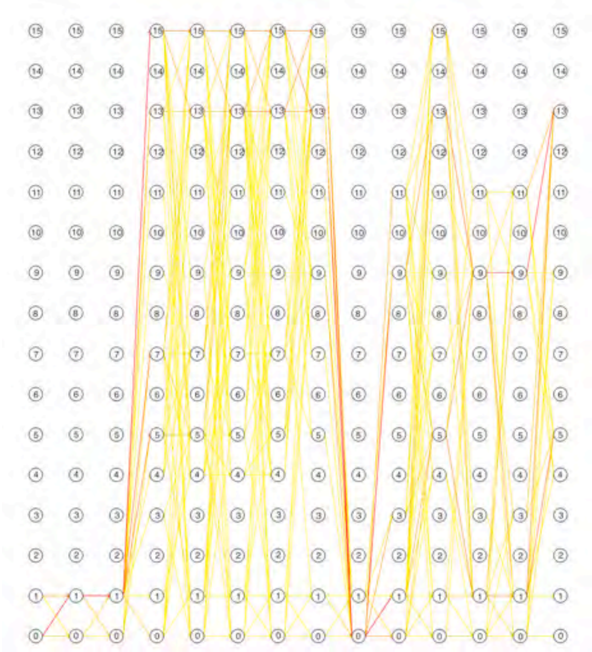

Figure 1. The graph constructed from the logs of all students

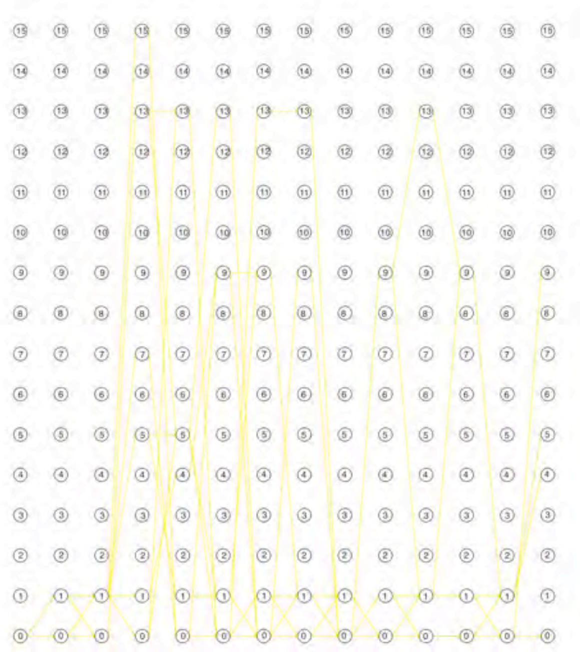

Figure 2. The graph constructed from the logs of the students who obtain grade F

Similarly, Figures 3 and 4 show the graphs that were constructed from the four items ordered by (iv) quiz score, (iii) report, (ii) browsing time, (i) attendance for all students and for those who achieved grade F, respectively. Compared with Figure 2, in Figure 4, more edges colored deep yellow appear in the lower half of the graph. This suggests that obtaining a grade F for the class was more strongly related to absence than to the quiz score.

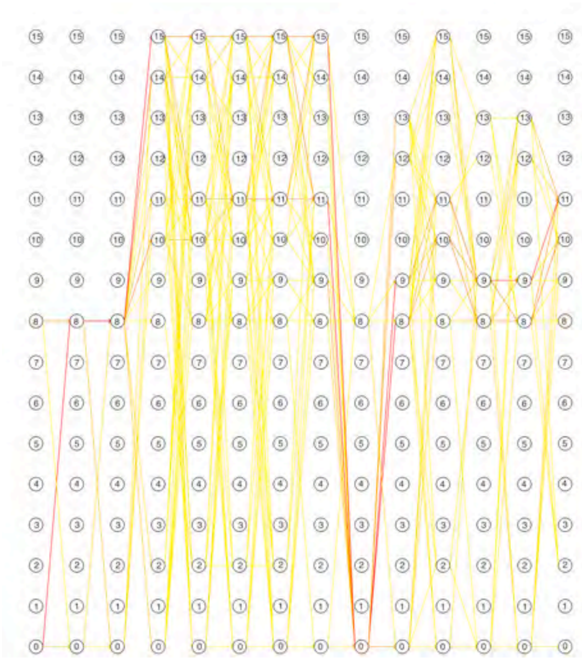

Figure 3. The graph constructed from the 4 items ordered by (iv), (iii), (ii), (i) of all students

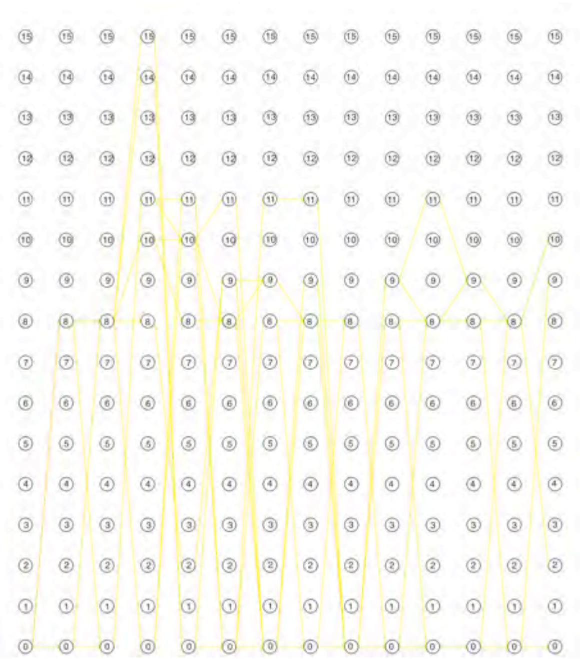

Figure 4. The graph constructed from the 4 items ordered by (iv), (iii), (ii), (i) of the students who obtain grade F

4. Prediction

Should the same class be held in the next season, the graphs constructed by the above method could be utilized for predicting students' learning activities in the following two ways:

- The state of the i-th lesson can be used to predict the next state of the (i+1)-th lesson the student is likely to reach, by using the graph constructed from the learning logs of all the students who attended the last season. The edges colored orange or deep yellow in the graph indicates these states. For example, according to the graph in Figure 1, a student who achieved state 15 in the 4^{th} lesson is likely to reach state 15 in the 5^{th} lesson. We note that the graph has Markov property, i.e., the possibility to reach the state of the (i+1)-th lesson only depends on the state of the i-th lesson (for the details of Markov property, see Norris J. R. (1998)).
- The graph of grade F enables us to inform students about the learning activities they should avoid, to ensure they do not obtain grade F. For example. Figure 2 shows that most students who achieved grade F reached state 0 in the 9^{th} lesson; hence, the students should avoid this state. Similarly, by using the graph representing a good grade, we are able to inform students of learning activities that they should do.

5. Conclusion and Future Work

In this paper, we have shown a method to visualize students' learning logs using discrete graphs. These logs contain the following four items: attendance, the time spent for browsing slides, the submission of a report and the quiz score for each lesson, all of which are collected from the LMS and the e-text system. We use these data to construct graphs for each grade, with the nodes of the graph representing all combinations of achievements and also failure to achieve the four items. In the constructed graph, the nodes are arranged in 14 columns of lessons from left to right and in 16 rows of state numbers from bottom to top. The graph enables us to observe the features of the learning activities of all the students who obtained a certain grade. The visual features of the graph can be changed by changing the order in which the above four items are presented.

We applied this method to the class that commenced in October, 2014 and considered the features of the learning activities of the students who obtained grade F. Moreover, using the graph constructed from the data of the same class held in the past enables us to propose an approach for predicting the next learning activities as well as an approach to inform students about learning activities they should avoid.

Many aspects remain to be investigated along the research direction presented in this paper. Points of particular importance include the following:

- The proposed method requires a manual decision regarding the thresholds for the achievements in items (ii) (total time of time spent for browsing slides) and (iv) (quiz score). However, determining the threshold automatically would be more effective for extracting the features of the learning activities for each grade. Similarly, it is also important to establish the best order in which to extract the four items from the learning logs.
- In our method the level of achievement in an item is expressed by a digit, i.e., 0 or 1. Extending this level to natural numbers (e.g., 0 to 100) or real numbers would enable us to construct a graph that contains more information regarding the learning activities. However, an extension such as this may cause an increase in the number of states of the graph; hence, we should explore a modified method capable of circumventing this problem.
- As stated in Section 4, in the constructed graph, the color of the edge of a state in the i-th lesson does not depend on the previous states achieved in the learning logs. In other words, the state in the (i+1)-th lesson of the student is predicted by only utilizing information of the state achieved in the i-th lesson. To realize more reliable prediction, we would have to modify the current model to take account of all the states the student passed before the (i+1)-th lesson. We note that a prediction using the modified model would require a large number of learning logs, because

the number of students who passed all the same states may be far smaller than those who only passed the same state in the i-th lesson.

Acknowledgements

The research results have been achieved by “Research and Development on Fundamental and Utilization Technologies for Social Big Data”, the Commissioned Research of National Institute of Information and Communications Technology (NICT), Japan. The work of F. Okubo was in part supported by Kyushu University Interdisciplinary Programs in Education and Projects in Research Development No.27115.

References

Baradwaj, B. & Pal, S. (2011) Mining Educational Data to Analyze Student’s Performance, International Journal of Advanced Computer Science and Applications, vol. 6, 2, pp. 63-69.

Norris J. R. (1998) Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Hlosta, M., Herrmannová, D., Váchová, L., Kužílek, J., Zdrahal, Z. & Wolff, A. (2014) Modelling student online behaviour in a virtual learning environment, in Workshop Proceedings of LAK 2014.

Wolff, A., Zdrahal, Z., Herrmannová, D., Kužílek, J. & Hlosta, M. (2014) Developing predictive models for early detection of at-risk students on distance learning modules, in Workshop Proceedings of LAK 2014.

Yin, C-J., Okubo, F., Shimada, A., Kojima, K., Yamada, M., Ogata, H. & Fujimura, N. (2014). Smart phone based data collecting system for analyzing learning behaviors, in Proc. International Conference of Computers on Education 2014, pp. 575–577.