

A VISUALIZATION SYSTEM FOR PREDICTING LEARNING ACTIVITIES USING STATE TRANSITION GRAPHS

Fumiya Okubo¹, Atsushi Shimada², Yuta Taniguchi³ and Shin'ichi Konomi¹

¹*Faculty of arts and science, Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka 819-0395, Japan*

²*Graduate School of Information Science and Electrical Engineering, Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka 819-0395, Japan*

³*Research Institute for Information Technology, Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka 819-0395, Japan*

ABSTRACT

In this paper, we present a system for visualizing learning logs of a course in progress together with predictions of learning activities of the following week and the final grades of students by state transition graphs. Data are collected from 236 students attending the course in progress and from 209 students attending the past course for prediction. From these data, the system constructs a state transition graph, where the prediction is based on the Markov property. We verify the performance of predictions by experiments in which the accuracy of prediction using the data of the course in progress and the one by 5-fold cross validation.

KEYWORDS

Learning log, predication of learning activity, state transition graph

1. INTRODUCTION

In recent years, the use of the ICT based educational system, such as learning management systems and e-book systems, has been widely spread, especially in higher education. One of advantages to use such a system is that we can automatically collect several kinds of log data of students' learning activities on a system. Analyzing the collected learning logs by data mining technologies, teachers can observe students' typical learning behaviors (Baradwaj et al. (2011)).

At Kyushu University, a learning support system called the M2B system was introduced in October 2014. The M2B system consists of three subsystems, that is, the e-learning system Moodle, the e-portfolio system Mahara, and the e-book system BookLooper provided by Kyocera Maruzen, Inc. The feature of BookLooper is that detailed operation logs can be collected, 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. This feature enabled us to analyze the learning behavior of students both inside and outside the university classroom. Our research group has conducted various investigations on learning analytics with the collected data. The details of the M2B system and our investigations are summarized in Ogata et al. (2015) and Ogata et al. (2017).

In the field of learning analytics, finding "at-risk" students who are likely to fail or drop out of class is important issue. A lot of methods for the early detection of at-risk students from data were intensively investigated, for example, in Baker et al. (2015), Romero & Ventura (2010) and Marbouti et al. (2016), where Logistic Regression, Support Vector Machine, Decision Tree, Multi-Layer Perceptron, Naïve Bayes Classifier, and K-Nearest Neighbor are guided. Moreover, Okubo et al. (2017) introduced the method by using Recurrent Neural Network that is known as a one of variants of Deep Neural Network. You (2016) showed meaningful learning logs based on the theory of self-regulation to predict students' final achievement by the statistical analysis method. Generally, the prediction of students' final achievement is based on data of previous courses, while in Hlosta et al. (2017) presented a method for finding "at-risk" students, which is depending only on the current course.

Not only analyzing data, giving appropriate feedback on the result of analysis to teachers and students is also important to improve their teaching activities and learning activities, respectively. For this purpose, in Okubo et al. (2015), a method for visualizing the four types of learning logs stored in the e-learning system and the e-book system namely, attendance, time spent for browsing slide, submission of a report, and quiz score, was presented. This method was realized by state transition graphs, where a state and an edge represent learning activities of a week and the number of students of such a transition, respectively, referring to the method proposed in Hlosta et al. (2014). Moreover, in Okubo et al. (2016), a method for identifying learning activities from the above four types of learning logs that are important for students to achieve a good final grade.

In this work, extending the method presented in Okubo et al. (2015), we propose a system for visualizing learning logs of an ongoing course together with predictions of student's learning activities of the following week and the final grade by state transition graphs. For predictions, learning logs of a course that has already finished are utilized. These methods are introduced in Section 2. Then, in Section 3, we verify the performance of prediction by experiments. Then, we compare the performance by using data of the past course for predicting with one by 5-fold cross validation. Moreover, we show results of questionnaire for students about this system. Finally, we give the conclusion and future research plans in Section 4.

2. METHOD

2.1 Data Collection

In order to employ the method of visualization, we need to collect learning logs from two courses, that is, one course that is in progress and another course that has already been finished. We expect that the two courses have a similar construction, for example, the past course is the same course as the course in progress by the same teacher opened one year ago.

In the two courses, the teacher and students use the LMS and the e-book system during lectures and preparation at all weeks of the courses. The lectures are presented by using several slides in the e-book system, each slide being associated with only one lecture. The slides are used by the students to complete their preparation sessions before each lecture. Furthermore, on some weeks, the students are required to submit a report and answer a quiz related to that week's lecture through the LMS.

Hence, in this study, we refer to the following four kinds of data stored in the LMS and the e-book system in this paper:

- (i) attendance or absence,
- (ii) the total time spent browsing the slides for preparation which reaches 600 seconds or above, or failure to do so,
- (iii) the submission of a report or failure to do so,
- (iv) the quiz score that reaches 70% or above, or failure to do so,

of each student participating in each week of the course. For each of the four items, we consider whether or not it is achieved. The achievement of an item is coded by "1" and the failure by "0". The courses are graded as A, B, C, D or F in the usual manner, with A being the best grade and F indicating failure of a course.

For an experiment, we collected learning logs from the following courses by the same teacher:

- (1) The "Information Science" course, which is attended by 236 students and opened in the first term (8 weeks) in 2016, is treated as the course in progress,
- (2) The "Information Science" course, which is attended by 209 students and opened in the first term (8 weeks) in 2015, is treated as the past course.

These courses are mainly for the first-year students of various departments. We note that in the second week of course (2), the submission of a report is required; however, in course (1), the submission of a report is not needed. This difference causes a difficulty in visualization in our method. Due to this reason, in order to adjust the logs of the two courses, we treat the data for item (iii) (the submission of a report) of the second week of course (1) as "0" for each student. Similarly, the data for item (iii) of the fifth week of course (2) is treated as "0" for all students.

2.2 Construction of a State Transition Graph

Using the logs of learning activities shown in Section 2.1, we construct a state transition graph to visualize the learning activities of the students in the course in progress and the prediction of learning activities. For this purpose, we assume that the target courses are opened through n weeks and the course in progress is finished until the m -th week.

A state in a state transition graph represents the learning activities of each week using four digits which refer to the achievements or failures in the four items (i), (ii), (iii), and (iv). For example, the state 1011 means that a student (i) attended the lecture, (ii) browsed the slides for preparation for less than 600 seconds, (iii) submitted a report, and (iv) obtained 70% or above in a quiz. The number of states is $2^4 = 16$ per week. Hence, the state transition graph consists of $16n$ states. The states are aligned such that the horizontal axis represents weeks from left to right, and the vertical axis represents states from bottom to top. In addition, the states representing the final grades are added to the right to the states of the n -th week.

An edge represents a transition of learning activities from the i -th week to the $i+1$ -th week. An edge between state x of the i -th week and state y of the $i+1$ -th lesson is constructed if there exists a student whose learning activities correspond to such a transition. The number of students having the same transition is expressed by the thickness of an edge. The edge is thin if the corresponding number is small. When the number of students who meet the condition increases, the edge thickness increases correspondingly. Similarly, edges from the states in the n -th week to the final grades are constructed. Figure 1 illustrates the example case where the number of transition from “1011” in the i -th week to “1011” in the $i+1$ -th week is larger than the one from “1011” in the i -th week to “1010” in the $i+1$ -th week. It is noted that if the i -th week is before the m -th week, we use the logs of the courses in progress; otherwise, we use the logs of the past courses.

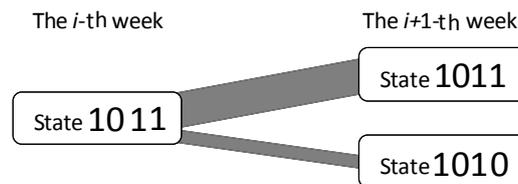


Figure 1. An example of a part of a state transition graph

2.3 Predication of Learning Activities and Final Grades

The logs of learning activities and the state transition graphs constructed by the above method can be utilized for predicting the learning activities and the final grades of students. Similar to Section 2.2, We assume that the target courses are opened through n weeks and the course in progress is finished until the m -th week.

From the state x in the m -th week of the course in progress, we can predict the next state y in the $m+1$ -th week of this course by selecting the state y in the $m+1$ -th week of the past course, where the number of students corresponding to this transition is the largest among all the transitions from the state x in the m -th week of the past course. This is represented by the thickest edge from the m -th week to the $m+1$ -th week in the state transition graph.

The final grade is predicated by using the state x in the m -th week of the course in progress and the logs of the past course which indicates the final grade that is most likely to be obtained by a student who is in state x in the m -th week. This is represented by the thickest edge only from the last week to the final grades in the state transition graph.

3. IMPLEMENTATION AND EXPERIMENTS

3.1 Implementation

The system for constructing the state transition graphs that visualizing the learning activities of students was implemented as a plugin of Moodle (the e-Learning System in Kyushu University).

Teachers can add this system to their courses. After selecting the course that has already been finished in the setting screen of the system (Figure 2), the state transition graph at that time is displayed.



Figure 2. The setting screen of the system

We note that the states not reached by any student are omitted, for visibility. The system has two modes depending on the user: the teacher mode and the student mode. In the teacher mode, users can see all the user names displayed on the left end, while in the student mode, users can see only their own name.

Figure 3 displays the state transition graphs using the logs of course (1) as the course in progress, which finishes in the fourth week and course (2) as the past course (users' names are hidden by a black box).



Figure 3. The state transition graph after finishing the fourth week

By clicking on a user's name, the following three changes occur (see Figure 4):

- the state transitions of the of the selected student until the fourth week are highlighted in red,
- the graph from the fourth week to the last week is reconstructed by using only the logs of the students who were at the same states of the fourth week in the past course as the selected student, the percentage that the selected student is likely to obtain each final grade is calculated and displayed on the right of each grade.

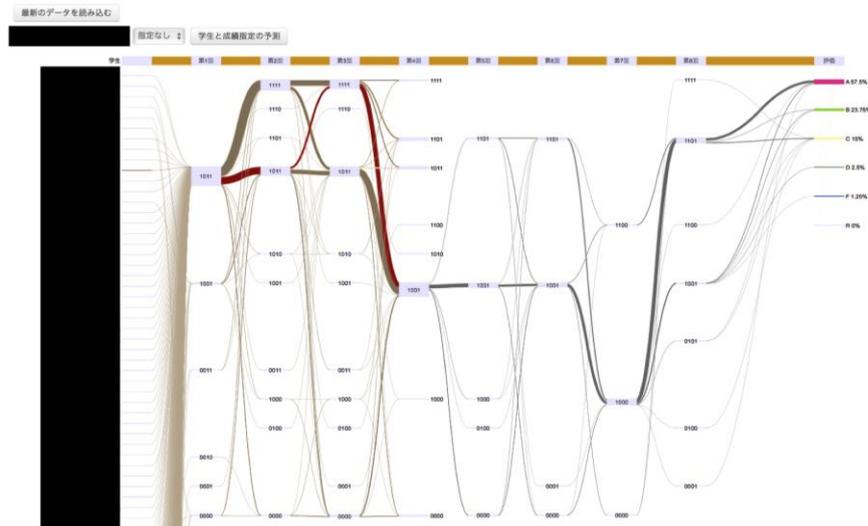


Figure 4. The state transition graph with the selected student

Students also can use this system, while users' names except oneself are anonymized.

3.2 Experiments

We conducted experiments to calculate the accuracy of the predictions of the learning activities of the following week and the final grade.

We applied our method to course (1) as the course in progress and course (2) as the past course. For each state in the i -th week with $1 < i < 7$, we predicted the state in the $i+1$ -th week and calculated the accuracy of this prediction. Similarly, for each state in the j -th week with $1 < j < 8$, we predicted the final grade and calculated its accuracy.

Moreover, we also applied our method by means of 5-fold cross validation. Therefore, the data are separated into five parts, of which four parts are used as training data and the remaining part as test data. It is noted that in our method, training data are treated as the data of the past course and test data are treated as the data of the course in progress. In this case, there are five ways to choose the parts for training data and test data. Hence, the final result of this 5-fold cross validation is the average of the results of these five ways of choosing the training data and the test data.

The results regarding the accuracy of the prediction of the learning activities of the following week are summarized in Table 1. In the column titled by " $i+1$ ", the accuracy of the prediction for the state in the $i+1$ -th week from the state in the i -th week is shown. The averages of the accuracies of the seven predictions are 37% in the case using the logs of the past course and 50% in the case of 5-fold cross validation.

Table 1. The accuracy of prediction for the following states

Method	2	3	4	5	6	7	8	Ave.
The past course	45%	45%	47%	22%	20%	58%	19%	37%
5-fold cross validation	45%	49%	47%	48%	41%	73%	46%	50%

The results regarding the accuracy of the prediction for the final grades are summarized in Table 2. In the column titled by “*i*”, the accuracy of the prediction for the final grades from the state in the *i*-th week is shown. The averages of the accuracies of the eight predictions are 73% in the case using the logs of the past course and 84% in the case of 5-fold cross validation.

Table 2. The accuracy of prediction for the final grades

Method	1	2	3	4	5	6	7	8	Ave.
The past course	81%	71%	70%	85%	86%	25%	85%	78%	73%
5-fold cross validation	85%	84%	85%	85%	85%	83%	84%	84%	84%

3.3 Questionnaire for Students

We asked the students attending the “Information Science” course to use the system and then carried out the following questionnaire:

- (a) How easy did you use this system?
 - (b) How useful was this system for reviewing your past learning?
 - (c) How much did you would like to change your learning in order to obtain better grades by looking at the results of prediction by this system?
 - (d) How much do you would like to use this system even in another class in the future?
- 201 students answered these questions. In Figure 5, the responses are summarized.

We can see that approximately 60% of the students felt that this system was useful for reviewing past learning, changing your learning in order to obtain better grades and they would like to use this system continuously. These results can be said to indicate that the system can solve the students’ anxiety about learning strategies and final grades. On the other hand, the students who said this system is easy to use are less than 50%. Improvement of the user interface such as how to display the states of graph is an important future task.

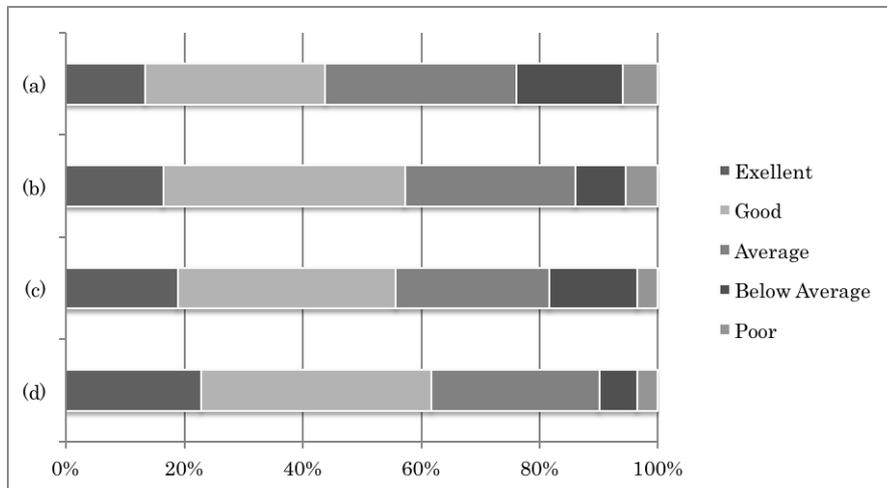


Figure 5. Summary of students’ answer to questions regarding our system

4. DISCUSSION

From the results of the experiments described in Section 3.2, we can observe that in general, the accuracy of the prediction for the state in the following week is low. We consider that the main reason for the low prediction accuracy is that the prediction of the proposed method is based on the Markov property, that is, the

possibility of reaching the state in the $i+1$ -th week of the course in progress only depends on the state in the i -th week of the past course, or the test data in 5-fold cross validation (for more details of Markov property, refer to Norris (1998)). For improving the accuracy of prediction, we suggest that using the all the data from the first week to the i -th week for predicting the state in the $i+1$ -th week may be effective. However, this plan has a limitation in that the predictions are difficult to visualize for teachers or students at least by using the state transition graphs proposed in this study.

In addition, we can observe that the accuracy of 5-fold cross validation is likely to be far higher than that of using the data of the past course in the predictions of both the following states and for the final grades. This observation implies that for the case using the data of the past course, the difference in construction (or instructions) of the course in progress and the past course sensitively affects the accuracy of the prediction, even when the two courses are the same and taught by the same teacher, although they are in different terms.

5. CONCLUSION

In this paper, we presented a system by which the learning activities of students in a course and the predictions of the learning activities of the following week and the final grades are visualized using state transition graphs. The learning activities of a week are represented by the four digits that means the achievements or non-achievements of attendance, the total time spent browsing the slides for preparation, the submission of a report, and the quiz score. A state transition graph is constructed from the logs of the two courses: the course in progress and the past course. If the course in progress is finished until the m -th week, the graph from the first week to the m -th week is constructed from the logs of the course in progress, and the rest part is from the one of the past course. A state in the graph represents the learning activities of each week using these four digits. An edge thickness between state x of the i -th week and state y of the $i+1$ -th lesson is according to the number of students whose learning activities correspond to such a transition. This system enables teachers to overview the learning activities of students until the finished week and the learning activities that students are likely to do. Moreover, in the system, by clicking on a student's name, the state transitions of him/her until the finished week are highlighted in red, and the percentage that he/she is likely to obtain each final grade is calculated and displayed. This function may help the detection of student who likely to obtain the particular final grade.

We verified the performance of predictions of the following state and the final grades by experiments applying method using the data of the two same courses opened in the different term, namely, the first term in 2015 and the first term in 2016. The experiments by 5-fold cross validation was also conducted, where only the data of the course in 2016 were used, to comparing the results using the data of the past course. The accuracy of prediction for the final grades was fairly high, exceeding 70%, in both the cases of using the past data and 5-fold cross validation; however, the accuracy of prediction for the following states is low, because we use only the information of the previous one state for prediction. In addition, we can observe that the result of 5-fold cross validation is better than the one using the logs of the past course for prediction. This situation may be caused by the difference in construction or instructions of the course in progress and the past course. For enhancing the performance of prediction using the logs of the past course, we need to consider the model that is robust against such a difference.

The questionnaire for students about this system was also conducted. The results indicates that this system is useful for changing learning activities in order to obtain better grades, while improvement of the user interface is necessary.

There remain many subjects to be investigated along the research direction regarding visualization and prediction presented in this paper. Points of particular importance includes the followings:

- As stated in Section 3.3, for improving the accuracy of prediction, it may be effective to use the all the data from the first week to the i -th week for predicting the state in the $i+1$ -th week or the final grade. The problem of this method is that the visualization of prediction is difficult by the similar way. For example, after finishing until the second week, for students "A" who is in state x_A in the first week and y in the second week and "B" who is in state x_B in the first week and y in the second week with $x_A \neq x_B$, the predictions for the third week of students A, B are different in general. But, this difference cannot be reflected in the state transition graph proposed in this paper.

- The thresholds which distinguish the achievement or the failure for the total time spent browsing the slides for preparation and the quiz score are manually decided in this paper. Finding the thresholds by which the accuracy of prediction is maximum is very important. It is also important to use another types of learning logs, such as the indicators based on the theory of self-regulated learning introduced in You (2016).

ACKNOWLEDGEMENT

The research results have been 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.

REFERENCES

- Baker, R. S., Lindrum, D., Lindrum, M. J. & Perkowski, D. 2015. Analyzing Early At-Risk Factors in Higher Education e-Learning Courses, *proc. EDM2015*, pp.150-155.
- 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., Kužilek, J., Zdrahal, Z. & Wolff, A. 2014. Modelling student online behaviour in a virtual learning environment, *Workshop Proc. LAK 2014*.
- Hlosta, M., Zdrahal, Z., Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data, *proc. LAK2017*, pp.6-15.
- Marbouti, F., H. A. Diefes-Dux & Madhavan, K. (2016) Models for early prediction of at-risk students in a course using standards-based grading, *Computer & Education*, pp. 1-15.
- Ogata, H., Taniguchi, Y., Suehiro, T., Shimada, A., Oi, M., Okubo, F., Yamada, M., Kojima, K. (2017). M2B System: A Digital Learning Platform for Traditional Classrooms in University, *Practitioner Track Proc. of LAK2017*, pp.154-161.
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K. & Yamada, M. 2015. E-Book-based Learning Analytics in University Education, *proc. ICCE2015*, pp.401-406.
- Okubo, F., Hirokawa, S., Oi, M., Shimada, A., Kojima, K. & Yamada, M. & Ogata, H. 2016. Learning Activity Features of High Performance Students, *Proc. the 1st International Workshop on Learning Analytics Across Physical and Digital Spaces (Cross-LAK 2016)*, pp.28-33.
- Okubo, F., Shimada, A., Yin, C. & Ogata, H. 2015. Visualization and Prediction of Learning Activities by Using Discrete Graphs, *proc. ICCE2015*, pp.739-744.
- Okubo, F., Yamashita, T., Shimada, A. & Ogata, H. (2017). A Neural Network Approach for Students’ Performance Prediction, *proc. LAK2017*, pp.598-599.
- Romero, C. & Ventura, S. (2010) Educational Data Mining: A Review of the State of the Art, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), pp.601-618.
- Wolff, A., Zdrahal, Z., Herrmannová, D., Kužilek, J. & Hlosta, M. 2014. Developing predictive models for early detection of at-risk students on distance learning modules, *Workshop Proc. LAK 2014*.
- You, J. W. 2016. Identifying significant indicators using LMS data to predict course achievement in online learning, *Internet and Higher Education*, vol.29, pp.23-30.