

Predicting student performance based on Lecture Materials data using Neural Network Models

Sukrit Leelaluk¹, Tsubasa Minematsu¹, Yuta Taniguchi¹, Fumiya Okubo¹, Atsushi Shimada¹

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

Abstract

Student Performance Prediction is essential for learning analysis of the students' learning behavior to discovering at-risk students for the early invention to support students. This study transforms the students' reading behavior into a two-dimensional matrix input based on each lecture material's reading behavior. The matrix input will be updated by accumulating the value for each week for performance prediction week by week. The multilayer perceptron neural network is employed to receive the matrix input and give feedback as a student's criteria consist of at-risk or no-risk students. This study considers the accuracy of a model considering between contents information and weekly information. We also investigate the switching of learning materials' order, the feature importance of the reading operation on an event stream, and the difference in reading behavior between at-risk and no-risk students. These can help the instructors for an early invention to support at-risk students.

Keywords

Learning Analytics, Performance prediction, Neural Networks

1. Introduction

Performance prediction is education mining data by collecting students' learning activity logs and analysis. Then, we receive output to analyze students' learning behavior. The purpose of performance prediction is to increase students' academic success or discover at-risk students early and invent them against drop-out or failure [1, 2].

Recently, Machine Learning (ML) and Deep Learning (DL) have been used for performance prediction in previous research. This is because ML and DL can learn and recognize the data pattern with a vast volume and predict the outcome with better accuracy [3]. Many pieces of research are proposing to predict students' performance early by using students' activities logs based on each week to create a time sequence. However, collecting the data from the education systems each week maybe find the break weeks or the same courses on each semester have a period difference. These problems will have the possibility to make accuracy drop. Furthermore, instructors can adjust the teaching styles of the courses every time. Therefore, using the past activities data as training data to predict the students' performance with different formats at the current time makes the prediction performance drop possible. This situation is similar with using a single-course predictive model to predict another course [8].

From the reason above, this study decided to transform the data to be based on lecture materials for students' performance prediction since the lecture materials are a fundamental object used to hold the lecture. In addition, this paper relies on the neural network (NN) to derive the high-performance classification model for predicting students' reading behavior. This can support instructors to identify at-risk students during the period of courses and invent students early.

2. Related Works

The prediction of students' performance becomes the challenge of learning analytics. The previous literature was recently analyzed to present different techniques to predict students' performance, such as statistical methods, machine learning, or deep learning.

Moises et al. [4] used ML and multilayer perceptron NN (MLP) to analyze the data on the massive learning management system (LMS) on 10%, 25%, 33%, and 50% of the course length for detect at-risk, fail and excellent students. Okubo et al. [5, 6] studied a method to predict students' final grades on multiple courses using a recurrent neural network (RNN) to analyze the time-series data. Murata et al. [7] used the knowledge distillation RNN model to compress the original model to become a student model for early detection of at-risk students. Finally, Conjin et al. [16] predict students' final grades and the predictor variables of 17 blended courses using standard and logistic regression on each week from the Moodle LMS data.

From these previous works of literature, we found they predicted students' performance by considering the event stream of reading behavior and other factors based on conventional weekly. In contrast to previous work, we present a prediction by considering the reading behavior on each learning material that instructors used in the lecture. This study will transform reading behavior data into the matrix input that can explain each student's reading information.

3. Datasets and Preparation

3.1. Data Collection

The clickstream data was received from the International Conference on Learning Analytics & Knowledge 2022 Data Challenge [9, 10]. Dataset consists of the two types of lectures (A and B) from onsite classes in 2019 and online classes in 2020 in the same institution. The summary of the detail of the experimental data shown in Table 1.

Table 1 The summary of the detail of the four courses used in experiments.

Course Name	The length of course	The number of students	The number of lecture materials
A – 2019	8 Weeks	50	21
A – 2020	7 Weeks	62	12
B – 2019	8 Weeks	163	8
B – 2020	7 Weeks	93	9

3.2. Features

The features are received from the operation on the event stream that concludes the total operations and reading time computed by using the openLA library [11]. The features are used in experiments shown in Appendix A.

3.3. Data Preprocessing

In this study, the raw data are transformed into two-dimensional matrices for inserted into the model [12, 13]. The matrix consists of m rows with the size of the number of lecture materials (C), and n columns with the size of the number of features (F) is displayed in figure 1. Where, its element X_{mn} represents the value of the clickstream on each action and lecture material.

$$\begin{array}{c}
 \\
 \\
C_m
\end{array}
\begin{array}{c}
F_1 \\
F_2 \\
\cdots \\
F_n
\end{array}
\begin{array}{c}
\left[\begin{array}{cccc}
\chi_{11} & \chi_{12} & \cdots & \chi_{1n} \\
\chi_{21} & \chi_{22} & \cdots & \chi_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\chi_{m1} & \chi_{m2} & \cdots & \chi_{mn}
\end{array} \right]
\end{array}$$

Figure 1: The Example of Matrix Input by 'C' and 'F' refers to the order of contents and feature, respectively.

The contents information matrix can input the feature in order of learning materials based on the previous data. Then, this matrix can accumulate values of features depending on students' reading behavior independent of time series. This can protect instructors from switching and not using content for a lecture. So, we can use the previous data to predict students' performance that differs from the model depending on time. Since instructors do not use a learning material to lecture students on that week, the value of the data has a significant change that may cause prediction performance to change.

From the observation data in table 1, we found that the amount of the contents is different. We thus made the matrix for each course is adjusted to have rows and columns equivalent using the matrix size of a course with most lecture materials on the same lecture-type as the main. If the missing value is found, this will be imputed by the average of each feature's value of students on the same course. All value on the matrix is normalized by robust standardization for rescaling features using statistics that are robust to outliers.

3.4. Model

This study creates MLP prediction models for each type of lecture is trained by the learning behavior in all lecture weeks. The MLP will consist of two hidden layers with the activation function are ReLU. The output layer had two units for classification between no-risk and at-risk students with the activation function is softmax. The optimizer function is Adam for the loss. The matrix will be adjusted to become the vector of size $m \times n$ by listing m rows next to each other and fetching them into the same model.

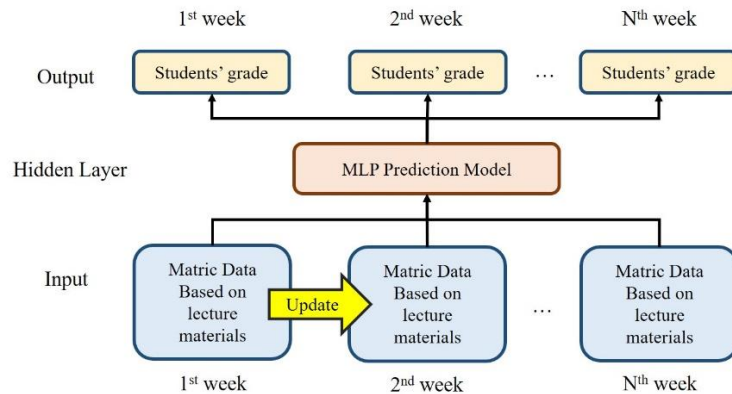


Figure 2: The concepts of input a matrix on the MLP independent model.

The experiment will predict the students' performance every week during the course period. The matrix will be updated by accumulating data week by week until the course's last week that is shown in figure 2.

3.5. Evaluation Criteria

The evaluation criteria will class the student's performance as a binary classification. The class will separate students who receive grades A, B, and C into the no-risk group and students who receive grades

D and F into the at-risk group [7]. The matrix's performance is evaluated using the accuracy. Table 2 shows the distribution of students' final grades in each course.

Table 2 The summary of the distribution of students' final grade.

Course Name	A	B	C	D	F
A – 2019	24	6	4	6	10
A – 2020	22	24	6	3	7
B – 2019	26	104	30	2	2
B – 2020	37	38	12	2	4

4. Experimental & Results

4.1. The efficiency of the model considering content information

This experiment presents the model's prediction result that considers weekly behavior with content information (Weekly w/ Content) in contrast with weekly behavior without content information (Weekly w/o Content). Many researchers predicted students' performance using the data based on conventional weekly in previous. For instance, Okubo et al. [5, 6] predicted students' final grades using RNN for each week.

The weekly prediction data is collected from the reading behavior data from the starting of a lecture of the current week until the starting of next week's lecture every week. Then, this is input into the prediction model as a vector size 1 row with n columns that uses features like a content information model. We evaluate the proposed method where A-2019 and B-2019 courses are used as the training data, and A-2020 and B-2020 courses are used as the test data.

From observing the correspondence of lecture materials on both lecture types from the raw data, an instructor adjusted the teaching styles of the courses by changing the number of contents. We found that A-2019 and B-2019 courses have ten and one content canceled for usage on the lecture in 2020, respectively. Then, A-2020 and B-2020 courses have two and one new content used in the lecture, respectively. Therefore, the matrix input for training and test set is adjusted to have the row and column equivalent by using the matrix size of a course with most lecture materials on the same lecture-type as the main. The missing value thus will be imputed by the mean of each feature's value on each student. However, an A-type lecture on the 2019 course has 21 lecture materials, including 12 main contents and 9 summarized from main contents. Then, an instructor left the main content only and added new content in 2020. In addition, many students chose to read the main contents more than summarized contents significantly on students' reading behavior on A-2019 lectures. We thus transform the training matrix of A-2019 courses to have the same size as the test matrix of the A-2020 course.

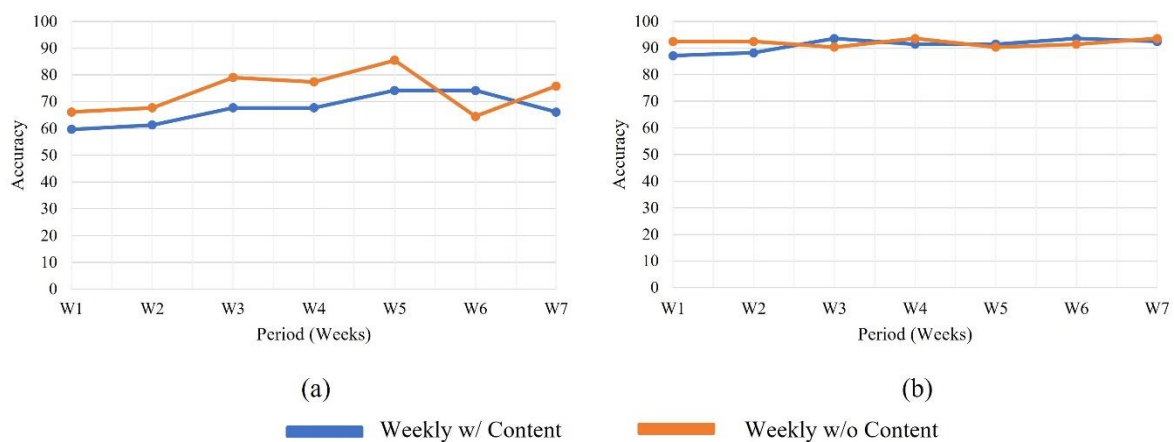


Figure 3: The graph between the model considering the content information and conventional weekly information on (a) A-2020 and (b) B-2020 courses.

The result found that the weekly w/o content model is better than the weekly w/ content model on the lecture A-type. However, these are similar to the lecture B-type. Then, we found that both prediction models had high accuracy on a lecture B-type than a lecture A-type shown in figure 3.

When we observe the result prediction on each model, we found that the weekly w/o content model can classify the final grade A and B students as no-risk students better than the weekly w/ content model on both types of lectures. However, the identity of the final grade C students is very well on both models in a B-type lecture more than an A-type lecture. The detection of the at-risk students of both models on A-type lecture is better than B-type lecture every week. This is because the training data of at-risk students on A-type lecture (16 students) is more than B-type lecture (4 students). Therefore, that affects the model better to analyze the pattern of at-risk students on A-type lecture. The pattern of reading behavior of both no-risk and at-risk students will be shown in section 4.2.

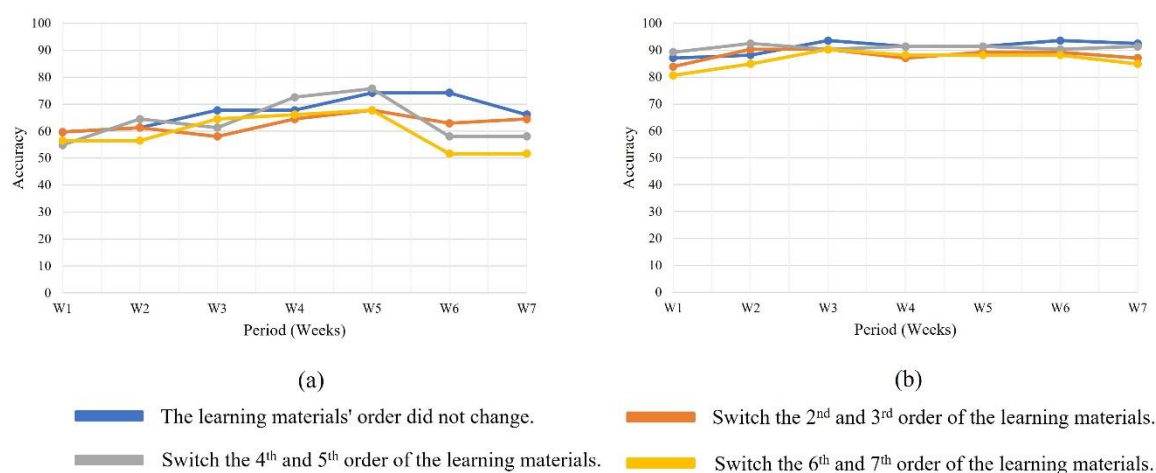


Figure 4a: The graphs show the switching of learning materials' order of a weekly w/ content model on (a) A-2020 and (b) B-2020 courses.

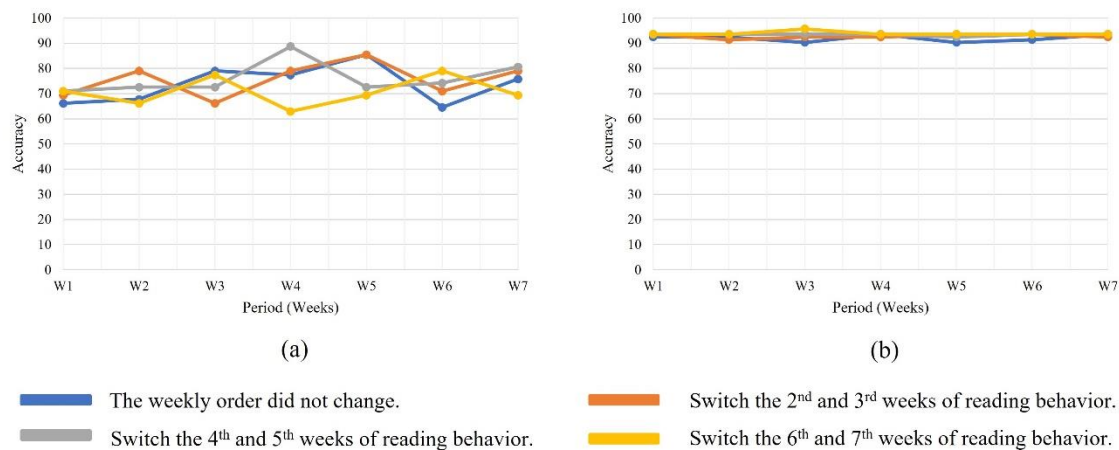


Figure 4b: The graphs show the switching of weekly order of a weekly w/o content model on (a) A-2020 and (b) B-2020 courses.

Furthermore, we simulated the example situation if instructors change the learning materials' order to make information change in an actual class in Figures 4a and 4b. We conducted these experiments by switching the order of learning materials in the weekly w/ content and the weekly information on weekly w/o content on test data every week and inputting the same model on previous experiments.

We found the switch of the contents on a weekly w/ content model made an accuracy drop on some weeks on a course period, especially on an A-type lecture. The reason is that a weekly w/ content

model considers each content's reading behavior as the main. Thus, switching the order materials makes the values of features change. In addition, instructors and students maybe give importance to each lecture equivalent each year. This made the separation of each content on each row on the matrix have new data that a weekly w/ content model has ever been trained to cause an accuracy drop (Figure 4a).

On the other hand, the switch of weekly information on a weekly w/o content model made an accuracy change on some weeks. However, this model can give an accuracy better than the weekly w/ content model. We found that the weekly w/o content model considers the operation count and reading time on each week as central. Let us consider students' behavior, such as reading time. We can find that the no-risk group takes time to read regularly. However, the at-risk group did not spend so much time reading that caused a weekly w/o content model to identify a class of students better than a weekly w/ content model (Figure 4b).

4.2. Analysis of the feature important

In the previous experiment, we conducted the matrix and vector input using features explained in section 3.2. We thus try to reduce the matrix size by exploring the feature importance by using the Random Forest classifier. This method allows us to know the insight into informative features for classification [14].

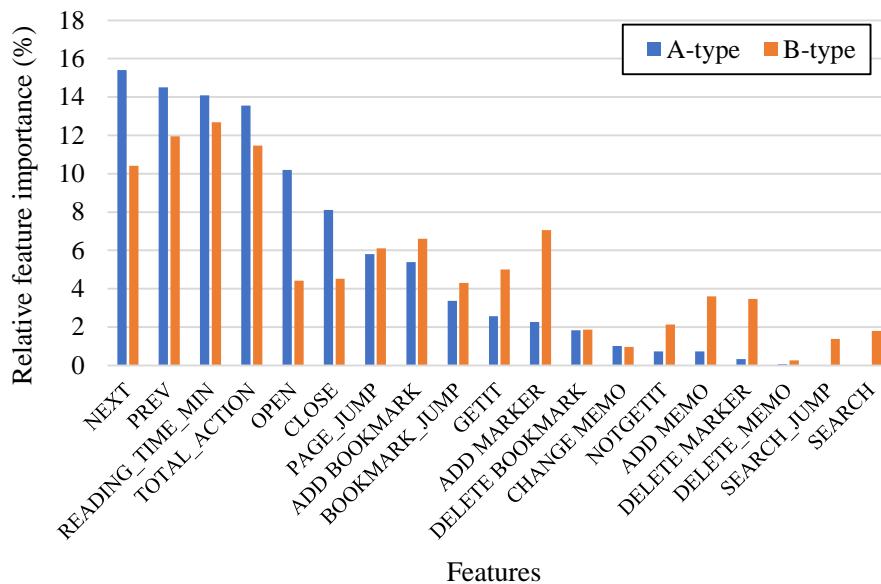


Figure 5: The feature importance of the A-type and B-type lectures by Random forest.

This experiment was conducted using the vector data from the weekly w/o content to explore the feature importance. We found the 'NEXT,' 'PREV,' 'READING_TIME_MIN,' and 'TOTAL_ACTION' are gave the most feature importance score as a percentage on both types of lecture. These features are the most helpful operation for reading learning materials. Furthermore, this will analyze students' reading behavior on each final grade by using a heatmap in figure 6. The heatmap of reading behavior was computed by the average of operations from the students on each final grade. The rich yellow color means that a feature has the most action and the purple color means that a feature has the poor action. The no-risk students had the most operation on an event stream, which means no-risk students spent much attention and time for reading learning materials on many pages. On the other hand, the at-risk students did not focus on reading lecture materials as no-risk students, which caused these students got to fail.

We thus try to adjust the matrix on weekly w/ content model by reducing the matrix size by using only the high score features composed of 'NEXT,' 'PREV,' 'READING_TIME_MIN,' and 'TOTAL_ACTION' compared with the current features. We found that the reduction of matrix size raised prediction accuracy on an A-type course due to the essential features, making the model more

easily recognize the difference between at-risk and no-risk students. However, a B-type course did not have any difference when using the high score features for prediction since the data of at-risk students was few. (Figure 7)

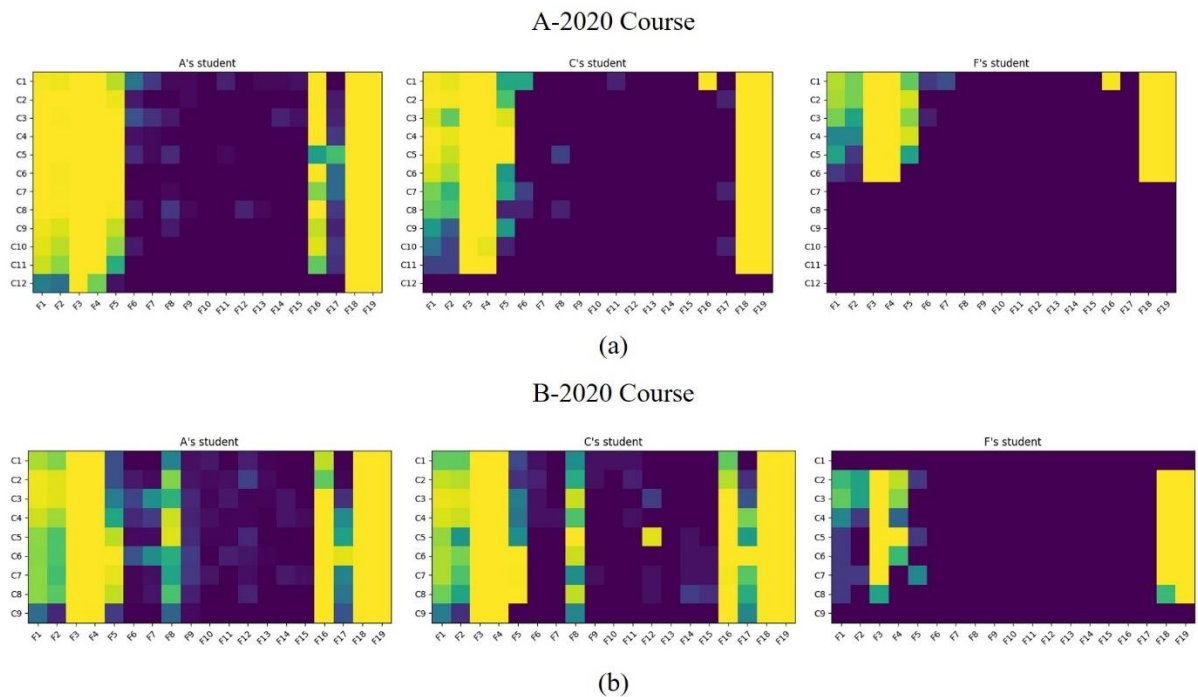


Figure 6: The heatmap of reading behavior average of students on courses (a) A-2020 and (b) B-2020 who receive grades A, C, and F, respectively.

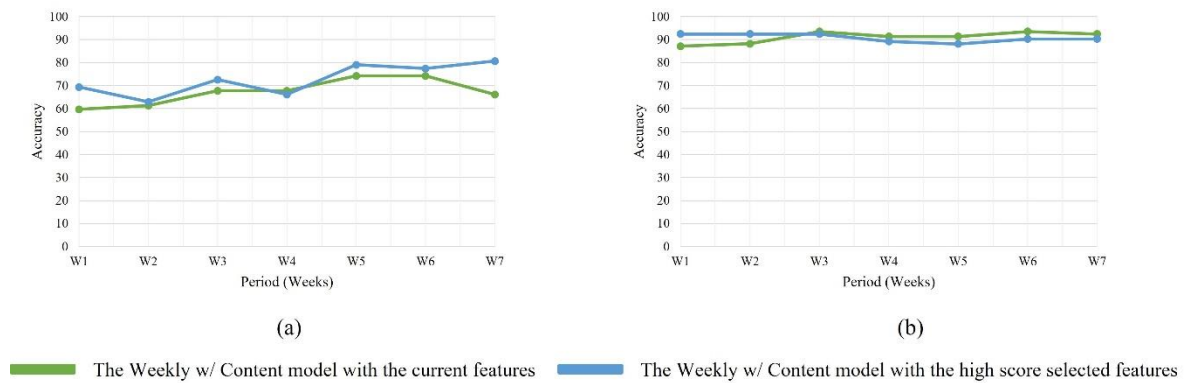


Figure 7: The graph between the selected features on (a) A-2020 and (b) B-2020 courses.

5. Discussion

From the result of the experiment, we found that predicting students' performance by considering learning content data can classify students who have a trend to fail or at-risk students if we have much reading behavior pattern of at-risk data sufficient. In addition, we used a heatmap to investigate the reading behavior on each week between no-risk students and at-risk students for a survey on how it affected the model performance. After the 3rd and 4th week of lecture, we found that no-risk students had a few actions for reading learning materials. On the other hand, at-risk students can continue reading and reviewing many learning materials until the end of the course (Figure 8). This causes a model to separate the reading behavior between at-risk and no-risk students.

The reading behavior of at-risk students has a minor operation on the reading of contents materials, which means these students spent a little time reading. We thus plot graphs of the average reading time

of all students those criteria by final grade on figure 9. Again, we found that the students in the no-risk group spent much time reading more than at-risk students significantly, according to the result on a heatmap. Let us consider the reading behavior with the prediction result based on the content information model. The instructors can invent students who can fail on half of the course period, such as 3rd or 4th week on the quarter course. Considering the early timing for interventions with past related literature, such as the author from [5] and [6], the model can also invent at-risk students around half of the course.

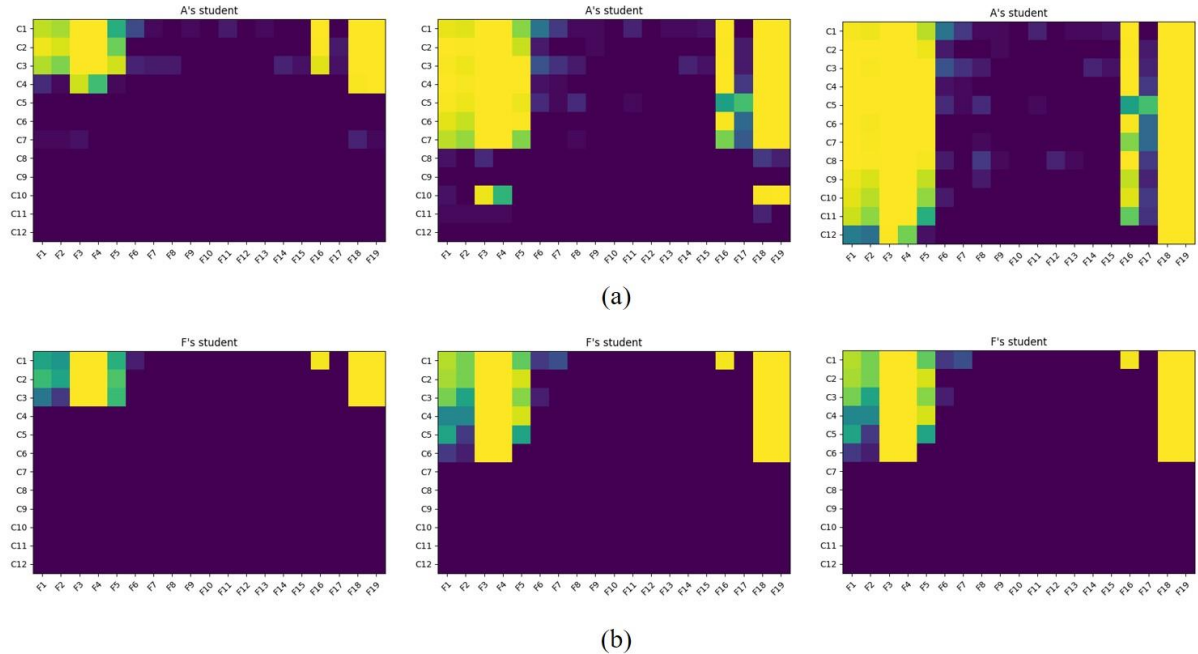


Figure 8: The heatmap of reading behavior average of students on B-2020 courses on 2nd, 4th, 7th of (a) final grade A students and (b) final grade F students.

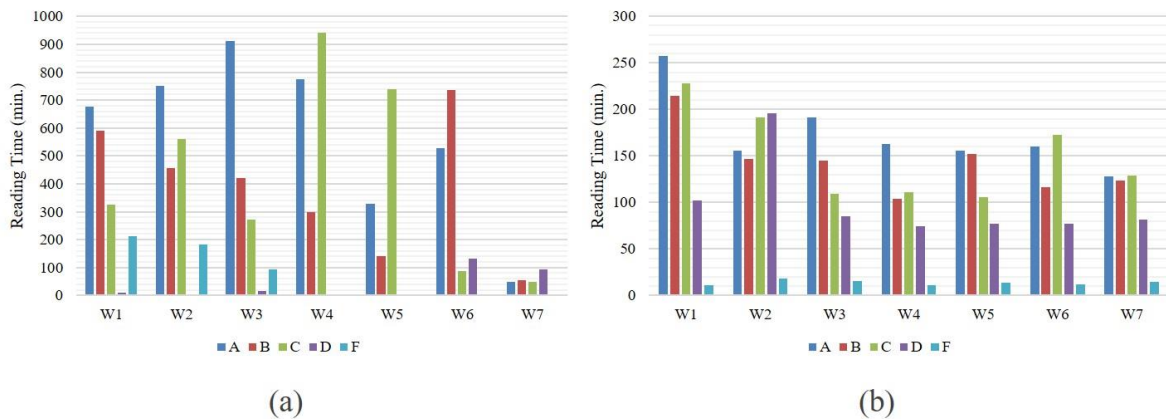


Figure 9: The graph of reading behavior time average of all students' final grades on (a) A-2020 and (b) B-2020 courses.

6. Conclusion & Future Works

This study employed the multilayer perceptron neural network to classify students' performance as at-risk and no-risk students for support instructors for early interventions. The experimental data were collected from the two different onsite and online lectures in different years. The model for identifying at-risk and no-risk students is based on the clickstream event on each learning material and transformed into a two-dimensional matrix. The features are generated by using the operation of students' reading behavior. We then contrast the model's accuracy with the model considering on weekly information

model. From the result, a weekly w/ content model can give accuracy similar to a weekly w/o model. A weekly w/ content model is a model that considers students' reading behavior depending on the content that is independent of time. However, a weekly w/ content switch made an accuracy drop since students may not focus on each content equivalent. Whereas the weekly switch information on the weekly w/o contents model can give a better accuracy since this model focuses on the total reading behavior each week.

Next, we try to reduce the size of the matrix by using the random forest to investigate the feature importance. We found that both types of lectures' high score operation features were the same as a result. These features are the most necessary action for reading learning materials on the system. Finally, we found the behavior of at-risk students gave little attention to studying and reviewing the learning materials on features on the matrix that can investigate on the heatmap after half of period so the teachers should invent to protect these students from failing before half of the course period.

In the future task, we plan to explore how to find an attention weight of each content on the same type of lecture due to instructors' changing lecture style and students' reading behavior. The purpose is to improve the weekly w/ content model because this can still perform well and explain almost all students' reading behavior well. Then, we want to optimize the model to have higher accuracy by using other deep learning models, such as the attention model [15], and extract new features from students' reading behavior.

Acknowledgements

This work was supported by JST SPRING Grant Number JPMJSP2136, JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grant Number JP18H04125, Japan.

References

- [1] Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417.
- [2] Marbouti, F., Diefes-Dux, H.A., & Madhavan, K.P. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Comput. Educ.*, 103, 1-15.
- [3] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*, 11(9), 552. doi:10.3390/educsci11090552
- [4] González, M.R., Ruíz, M.D., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Comput. Educ.*, 163, 104108.
- [5] Okubo, F., Yamashita, T., Shimada, A., & Konomi, S. (2017). Students' performance prediction using data of multiple courses by recurrent neural network. In A. F. Mohd Ayub, A. Mitrovic, J.-C. Yang, S. L. Wong, & W. Chen (Eds.), *Proceedings of the 25th International Conference on Computers in Education, ICCE 2017 - Main Conference Proceedings* (pp. 439-444). (Proceedings of the 25th International Conference on Computers in Education, ICCE 2017 - Main Conference Proceedings). Asia-Pacific Society for Computers in Education.
- [6] Okubo, F., Yamashita, T., Shimada, A., Taniguchi, Y., & Shin'ichi, K. (2018). On the prediction of students' quiz score by recurrent neural network. *CEUR Workshop Proceedings*, 2163.
- [7] Murata, R., Minematsu, T., Shimada, A. (2021). Early Detection of At-risk Students based on Knowledge Distillation RNN Models. *Educational Data Mining 2021 (EDM 2021)*.
- [8] López-Zambrano, J., Lara, J. A., & Romero, C. (2020). Towards Portability of Models for Predicting Students' Final Performance in University Courses Starting from Moodle Logs. *Applied Sciences*, 10(1), 354. doi:10.3390/app10010354
- [9] Flanagan, B., Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. In *International Conference on Computers in Education (ICCE2017)* (pp.333-338).
- [10] Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang, J., & Hirokawa, S. (2017). Learning analytics for e-book-based educational big data in higher education. In *Smart Sensors at the IoT Frontier* (pp. 327-350). Springer, Cham.

- [11] Murata, R., Minematsu, T., Shimada, A. (2020). OpenLA: Library for Efficient E-book Log Analysis and Accelerating Learning Analytics. In International Conference on Computer in Education (ICCE 2020), 301-306.
- [12] Daniusis, P., Vaitkus, P. (2008). Neural Network with Matrix Inputs. Informatica, Lith. Acad. Sci.. 19. 477-486. 10.15388/Informatica.2008.225.
- [13] Dalal, A., Imtiaz, A., Ebrahim, A. (2021). Deep learning based network traffic matrix prediction. International Journal of Intelligent Networks, 2, 46-56.
- [14] Altaf, S., Soomro, W.J., & Rawi, M.I. (2019). Student Performance Prediction using Multi-Layers Artificial Neural Networks: A Case Study on Educational Data Mining. Proceedings of the 2019 3rd International Conference on Information System and Data Mining.
- [15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
- [16] Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. IEEE Transactions on Learning Technologies, 10, 17-29.

Appendix A. Features related to reading behavior

The following table presents the variables describing the reading operation collected from the students' reading behavior on the e-book system. The description is referred to the dataset received from LAK22 Data Challenge.

Feature	Operation Name	Description
F1	OPEN	The amount of opened the e-book
F2	CLOSE	The amount of closed the e-book
F3	NEXT	The amount of went to next page
F4	PREV	The amount of backed to previous page
F5	PAGE_JUMP	The amount of jumped to a particular page
F6	ADD BOOKMARK	The amount of added a bookmark
F7	BOOKMARK_JUMP	The amount of jumped to a particular page from the bookmark
F8	ADD MARKER	The amount of added a marker
F9	ADD MEMO	The amount of added a memo
F10	CHANGE MEMO	The amount of edited an existing memo
F11	DELETE BOOKMARK	The amount of deleted a bookmark
F12	DELETE MARKER	The amount of deleted a marker
F13	DELETE_MEMO	The amount of deleted a memo
F14	SEARCH	The amount of searched for something within the e-book
F15	SEARCH_JUMP	The amount of jumped to a particular page from the search results
F16	GETIT	The amount of clicked the "GETIT" button in the e-book
F17	NOTGETIT	The amount of clicked the "NOTGETIT" button in the e-book
F18	TOTAL_ACTION	The total of operation on the e-book
F19	READING_TIME_MIN	The reading time of each lecture materials (minutes)