Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Early Detection of At-risk Students Through Leaning-Activity Forecasting

Yuya OZAKI[a*], Daisuke DEGUCHI[a], Haruya KYUTOKU[b] & Hiroshi MURASE[a]
[a]*Graduate School of Informatics, Nagoya University, Japan*
[b]*Department of Media Informatics, Aichi University of Technology, Japan*
*ozakiy@vislab.is.i.nagoya-u.ac.jp

Abstract: With the widespread adoption of digital technologies such as digital textbooks, it has become feasible to collect daily logs of students' learning activities. Accordingly, there has been a growing trend in research using these logs. One of these areas is focusing on predicting grade of each student based on these learning activity logs. However, previous research focused on detecting At-risk students when learning activity logs for all lectures are available. This is not applicable for detection at the first few lectures (i.e. weeks) required in practical usage scenarios. We call this scenario as “early detection” in this paper. However, in early detection, the accuracy of at-risk detection tends to decrease. To solve this problem, we propose a Learning-activity Forecasting Network (LFNet) that improves the accuracy of early detection by aligning the embedding of the first few lectures with that of all lectures. Through experiments on learning activity logs of actual lectures, we confirmed that the proposed method could achieve high At-risk detection accuracy even from the first few lectures of learning activity logs.

Keywords: At-risk detection, E2Vec, learning analytics, learning log, student's performance prediction

1. Introduction

In recent years, digital technologies such as learning management systems and digital textbooks are becoming popular. It has become possible to collect students' learning activity logs digitally and automatically, such as Sakai (Apereo Foundation (2024)), Blackboard (Bradford et al. (2007)), and Moodle (Dougiamas et al. (2003)). These platforms provide learning content over the Internet and enable the collection of students' learning activity logs.

Many researchers try to utilize this information for actual educational improvement. One of such tasks is detection of At-risk students. Detecting At-risk students can enable early intervention from teachers to prevent them from dropping out. For detection, E2Vec (Miyazaki et al. (2024)), which can vectorize while preserving the characteristics of learning activities, has been proposed as an alternative to traditional statistical methods. E2Vec converts each learning activity into its corresponding symbol, and a sequence of them are treated as a text and then utilizes the fastText model (Bojanowski (2017)) to embeds them into a vector, resulting in a fixed-dimensional feature vector. This feature vector is called “embedding” in the machine learning research field, so we use the term “embedding” in this paper. By employing E2Vec, it becomes feasible to generate fixed-dimensional embedding that preserves more information from variable-length learning activity logs. Miyazaki et al. used E2Vec to perform At-risk detection using learning activity logs up to the final lecture. However, At-risk detection at the final lecture is not helpful in assisting At-risk students. In practice, it is necessary to identify At-risk students as early as possible, such as at the end of the third week when there are eight weeks of lectures in total. Therefore, for practical application, it is indispensable to

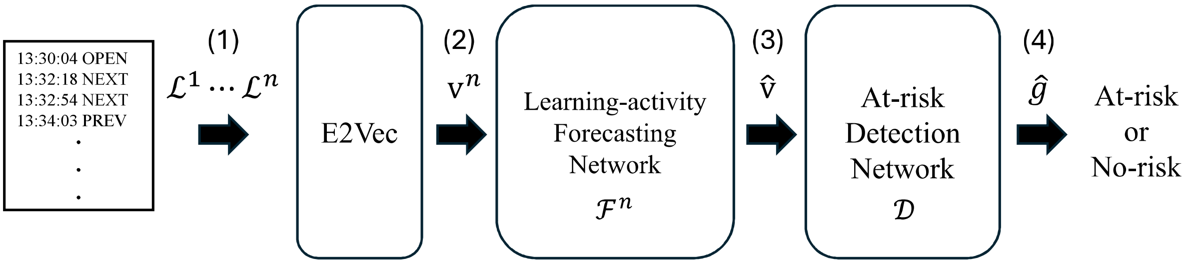

Figure 1. At-risk Detection Procedure.

develop a method that enables early detection. To address this issue, the proposed method introduces a Learning-activity Forecasting Network (LFNet). This network converts embedding generated from $1, 2, \cdots, n$-th ($n < N$) lectures into a pseudo whole lecture embedding corresponding to that generated from all N lectures. As a result, it prevents At-risk Detection Network from mistaking No-risk students for At-risk students.

2. Method

For At-risk detection, we use both Learning-activity Forecasting Network $\mathcal{F}^n$ and the At-risk Detection Network $\mathcal{D}$. These are multilayer perceptron which have dropout and batch normalization. Learning-activity Forecasting Network $\mathcal{F}^n$ is trained to predict a pseudo whole lecture embedding from the embedding v^n generated from activity logs of n-th lectures. At-risk Detection Network $\mathcal{D}$ is trained using whole learning activity logs. Figure 1 shows the overview of At-risk detection. Following procedures show the At-risk detection when activity logs of $1, 2, \cdots, n - th$ $(n < N)$ lectures are available.

1. Input the learning activity logs $\mathcal{L}^1 \cdots \mathcal{L}^n$ corresponding to $1, 2, \cdots, n$-th ($n < N$) lectures.
2. Generate embedding v^n from $\mathcal{L}^1 \cdots \mathcal{L}^n$.
3. Predict the pseudo whole lecture embedding $\hat{\mathrm{v}}$ using Learning-activity Forecasting Network $\mathcal{F}^n(\mathrm{v}^n|\Theta_1)$ (Θ_1 is parameter).
4. Predict $\hat{g}$ (At-risk or No-risk) using At-risk Detection Network $\mathcal{D}(\hat{\mathrm{v}}|\Theta_2)$ (Θ_2 is parameter).

In summary, the overall At-risk detection process can be formulated as the following equation.

$$\hat{g} = \mathcal{D}(\mathcal{F}^n(\sum_{j=1}^{n} \mathrm{E2VEC}(\mathcal{L}_j)|\Theta_1)|\Theta_2)$$

3. Experiments

To confirm the effectiveness of the proposed LFNet for the At-risk detection task, we conducted experiments using learning activity logs collected at Kyushu University (LIMU (2020)). Among the final grades of A, B, C, D, and F, students were labeled as No-risk for grades A, B, and C, and At-risk for grades D and F. To evaluate the effectiveness of the proposed LFNet, we compared the following three methods:

Comparative 1. Learning activity logs of $1,2, \cdots, n$ lectures are used for both training and testing of the At-risk detection network.

Comparative 2. The At-risk detection network was trained by using learning activity logs of all $N = 7$ lectures, while learning activity logs of $1, 2, \cdots, n$ lectures were used for the testing.

Proposed. The At-risk detection network was trained by using learning activity logs of all $N = 7$ lectures, while pseudo whole lecture embedding converted from embedding generated from $1, 2, \cdots, n$ lectures by Learning-activity Forecasting Network was used for the testing.

4. Results and Discussions

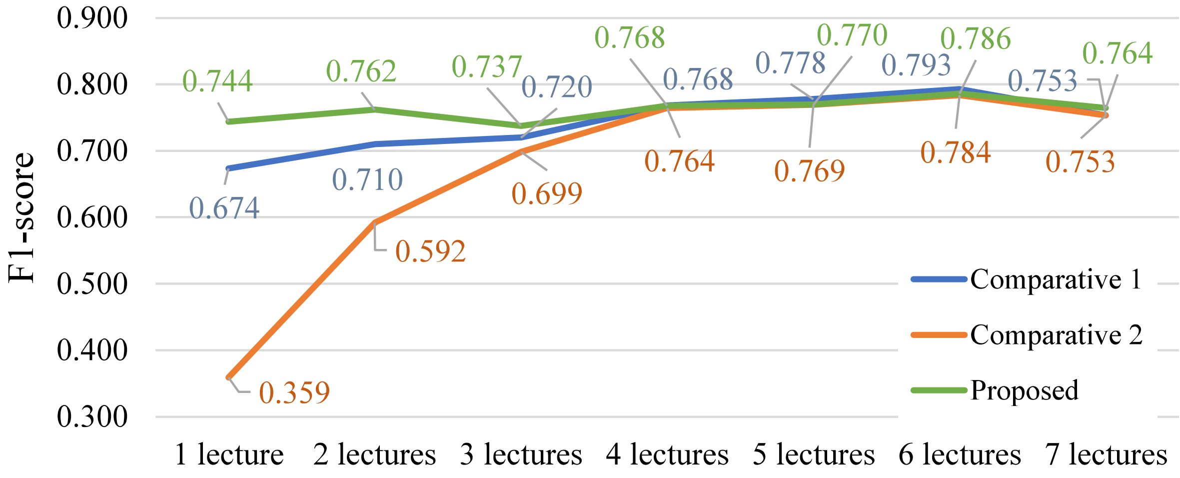

Figure 2. The results of experiments.

Figure 2 shows the results of Comparative 1, Comparative 2 and the proposed method. The vertical axis represents the average F1-score for At-risk and No-risk students, which is calculated by averaging 10 trials of cross-validations. The horizontal axis indicates the number of lectures n used for At-risk detection.

From Figure 2, it can be confirmed that the proposed method outperforms the comparative methods in terms of F1-scores from the 1st to 3rd lecture. On the other hand, after the 4th lecture, all methods achieved similar F1-scores. From this result, it is confirmed that the proposed method effectively improves the accuracy of At-risk detection in the early lecture of the course.

5. Conclusions

This paper proposed Learning-activity Forecasting Network for early detection of At-risk students. The results showed that the proposed method was more accurate for early detection than the comparative methods. Especially, we confirmed that it is better to train the At-risk detection network using learning activity log at whole lectures.
Future work will include incorporating time-series information on learning activities into the At-risk detection model and the use of multimodal information such as lecture videos and lecture materials.

Acknowledgements

This work was supported by JST CREST Grant Number JPMJCR22D1. The computation was carried out using the General Projects on supercomputer “Flow” at Information Technology Center, Nagoya University.

References

Apereo Foundation . “Sakai LMS | Apereo” . apreo , https://www.apereo.org/projects/sakai-lms, (2024-05-02).

Bradford, P., Porciello, M., Balkon, N., and Backus, D.: The Blackboard Learning System: The Be All and End All in Educational Instruction?, Journal of Educational Technology Systems, vol. 35, pp. 301-314, 2007.

Dougiamas, M., and Taylor, P. Moodle: Using learning communities to create an open source course management system., Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2003.

Miyazaki, Y., Švábenský, V., Taniguchi, Y., Okubo, F., Minematsu, T., and Shimada, A.: E2Vec: Feature Embedding with Temporal Information for Analyzing Student Actions in E-Book Systems, The 17th International Conference on Educational Data Mining, 2024.

Bojanowski, P., Grave, E., Joulin, A. and Mikolov T.: Enriching word vectors with subword information, Transactions of the Association for Computational Linguistics, vol. 5, pp. 135-146, 2017.

LIMU. “Introduction — OpenLA documentation”. OpenLA, https://limu.ait.kyushu-u.ac.jp/~openLA/introduction.html, (2024-05-02).