Grade Prediction Using fastText Features Weighted Through Differential Pattern Mining

Ryota Tachi[1], Tsubasa Hirakawa[2], Takayoshi Yamashita[3], Hironobu Fujiyoshi[4]
Chubu University
{tatti[1], hirakawa[2]}@mprg.cs.chubu.ac.jp, {takayoshi[3], fujiyoshi[4]}@isc.chubu.ac.jp

ABSTRACT: With the digitization of educational materials, there is growing anticipation for predicting grade performance using learning log data. Previous studies have attempted to predict performance by inputting histogram features of the number of digital material operations into machine learning models. However, these approaches do not consider temporal sequences, making it difficult to reflect behavioral patterns in the performance predictions. To address this issue, we propose maintaining the time series using fastText, which embeds learning behaviors as features. Additionally, we employ differential pattern mining to detect behavior patterns that exhibit significant differences and then apply weighting to these patterns in fastText. Evaluation experiments show that our proposed method improves performance prediction accuracy compared to conventional methods and that weighting behavioral patterns proves effective.

Keywords: Grade Prediction, Differential pattern mining, Sequential Data

1 INTRODUCTION

As educational environments go digital, the use of machine learning models in education is drawing attention. Many studies predict performance from learning behaviors for early dropout detection and improving learning, but most rely on operation frequency histograms (Kohama et al., 2023), ignoring behavioral patterns. To address this, we propose a performance prediction method using fastText features weighted by differential pattern mining and verify the effectiveness of input data that reflects these behavioral patterns.

2 PROPOSED METHOD

We propose an operation log embedding method that applies weighting based on differential patterns. Figure 1 provides an overview of the method, which consists of three modules: differential pattern mining, operation log embedding, and classification prediction. First, differential pattern mining identifies patterns that differ significantly between high- and low-achieving students. Next, we use E2Vec (Miyazaki et al., 2024) preprocessing and embedding modules to generate operation log embedding features. E2Vec converts each operation log into a single character and aligns it with the NLP concepts of “character,” “word,” and “sentence,” then uses fastText for embedding. For instance, NEXT becomes ‘N,’ PREV ‘P,’ and OPEN ‘O.’ In E2Vec, logs within one minute and up to 15 operations are treated as a “word,” embedded, and averaged to obtain the operation log embedding vectors. In our proposed method, we acquire a sentence embedding vector that reflects learning patterns through a weighted average, where the difference in the proportion of students who performed different patterns is used as the weight. The embedding vector v_S of a sentence is computed as $v_S =$

$\frac{\sum_{i=1}^{m} w_i \cdot \frac{u_i}{|u_i|}}{\sum_{i=1}^{m} w_i}$. Here, w_i is the weight, and u_i is the embedding vector of the word. Using the operation log embedding features obtained as described above as input, we perform grade classification predictions using a classifier.

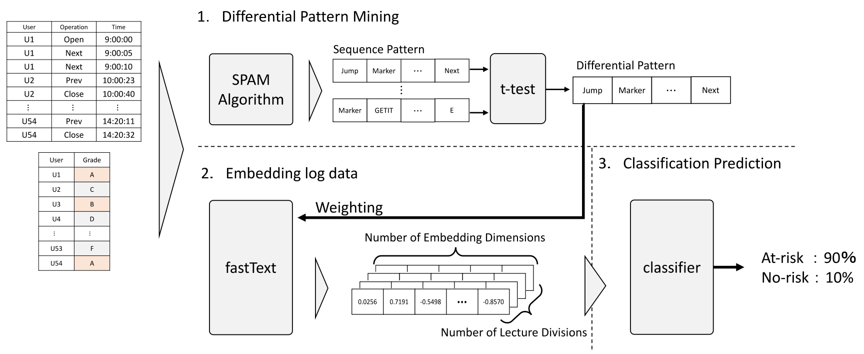

Figure 1：Overview of the proposed method

3 EVALUATION EXPERIMENT

We compare the proposed method with E2Vec, a conventional histogram-based method.

3.1 Experimental conditions

We use operation log data collected from six courses at Kyushu University. Among them, courses A-2020 and D-2020 are used solely for identifying differential patterns, while the remaining courses are used for the classification prediction task. For differential pattern mining, the SPAM algorithm is employed with a maximum pattern length of 15, a minimum support of 40%, and a maximum gap of 2. In this experiment, students are classified into two classes: No-risk and At-risk. The classification models used are RandomForest, XGBoost, and SVM.

3.2 Experimental results

We compare the grade classification accuracy of the proposed method with that of conventional methods in Table 1. Of the four courses, one is used as the training dataset, while the remaining three serve as evaluation datasets. Table 1 presents the average accuracy across evaluations conducted on each evaluation dataset. The results indicate that, except when using A-2022 as the training data, the proposed method achieved higher classification accuracy than the conventional method.

Table 1：Comparison of average accuracy per training data

Train	E2Vec	Propose_{RF}	Propose_{XGB}	Propose_{SVM}
A-2021	0.6268	**0.6280**	0.6134	0.6162
A-2022	**0.7237**	0.6087	0.5324	0.6213
D-2021	0.4869	**0.5745**	**0.6044**	**0.5528**
D-2022	0.6115	**0.6649**	**0.6696**	**0.6341**

4 ANALYSIS

This chapter analyzes patterns identified through differential pattern mining. Table 2 shows patterns that differ significantly between the two classes, as well as those that show minimal differences. According to Table 2, the patterns with a large difference between the two classes involve alternating "NEXT" and "PREV" operations. This suggests that after opening the material, students frequently use "PREV," indicating they are reviewing previously covered lecture content. In particular, high-achieving students repeatedly check preceding and following pages, implying a conscious effort to grasp the contextual flow of the material. This behavior suggests that their grade performance differences may stem from more active and intentional engagement with the course content. On the other hand, patterns with minimal differences involve repetitive "NEXT" or "PREV" operations—redundant actions often observed within the first 0 to 10 minutes of the lecture. These may reflect attempts to quickly navigate to specific pages used during the lecture. Moreover, some students may be simply tracing the instructor's own operations, repeatedly clicking "NEXT" or "PREV." Such repetitive sequences are particularly common among lower-achieving students and do not directly correlate with active learning. They may indicate a lower level of concentration or an attempt to mimic the instructor's actions rather than engaging deeply with the material.

Table 2: Patterns with Large and Small Differences

Pattern	Proportion difference	Pattern	Proportion difference
ONPPPN	1.000000	PPNNNNNNNNNNCO	0.020000
ONPPNPN	1.000000	NPPPPPPPO	0.024355
ONNPPNPN	1.000000	NNPPPPPPC	0.024355
NPNPNNC	0.669145	NNCJN	0.024355
ONNPNPNC	0.669145	PNPNNCO	0.024451

5 CONCLUSION

In this study, we demonstrated the effectiveness of weighting using differential patterns. Furthermore, we were able to identify patterns of active engagement in learning as positive features, and redundant patterns of continuous operations as negative features.

REFERENCES

Kohama, H., Ban, Y., Hirakawa, T., Yamashita, T., Fujiyoshi, H., Itai, A., & Usami, H. (2023). Recommending learning actions using neural network. In International Conference on Computers in Education 2023.

Miyazaki, Y., Švábenský, V., Taniguchi, Y., Okubo, F., Minematsu, T., Shimada, A. (2024). E2Vec: Feature Embedding with Temporal Information for Analyzing Student Actions in E-Book Systems. Proceedings of the 17th International Conference on Educational Data Mining, 434--442. https://doi.org/10.5281/zenodo.12729854