Majumdar, R. et al. (Eds.) (2025). Proceedings of the 1st International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education

Grade Prediction Considering Learning Log Relationship

Taiga YAMAMOTO*, Tsubasa HIRAKAWA, Takayoshi YAMASHITA & Hironobu Fujiyoshi
Chubu University, Japan
*taiga@mprg.cs.chubu.ac.jp

Abstract: Analyzing learning log data from digital platforms helps identify at-risk students and provide personalized academic support. In this study, we aim to improve prediction accuracy by considering the temporal and contextual relationships among learning logs. We introduce a Transformer-based approach that processes sequences of tokenized learning logs. Our experiments show that the proposed method achieves higher prediction accuracy than previous methods. This result highlights the effectiveness of modeling these sequential relationships.

Keywords: Grade Prediction, Transformer, Learning Logs, Digital Textbooks, e-book EventSteam

1. Introduction

The digitization of education has led to the widespread availability of student data, including learning logs and academic records. Analyzing this data enables the early identification of at-risk students and provision of personalized support. Consequently, grade prediction using machine learning has become a key area of research.

Previous methods used features extracted by fastText (Miyazaki et al., 2024) and histograms of learning logs (Kohama et al., 2023) as input, but they fail to capture short-term patterns in learning behavior. This study aims to improve grade prediction accuracy by modeling relationships between learning logs. Specifically, each log is treated as a token and input to a Transformer (Vaswani et al., 2017) for grade prediction.

2. Proposed Method

To address the limitations of previous methods, we propose a model based on the Transformer Encoder architecture. Our approach involves tokenizing each learning action and utilizing the encoder's self-attention mechanism to effectively capture the complex sequential and contextual dependencies within the action logs. As illustrated in *Figure 1*, our model predicts student grades from their learning action logs.

To implement this, we first handle the computational cost of long sequences by segmenting the logs into overlapping subsequences of 256 tokens with a 128-token stride; padding is applied where necessary. Consequently, the number of subsequences, n, varies for each student.

Each token is then converted into a 128-dimensional embedding vector, to which positional encodings are added to preserve the sequential order. The Transformer Encoder processes these embedded subsequences to generate contextualized output vectors. These output vectors are subsequently averaged to produce a single vector representing the student's overall behavior. Finally, a linear classification head uses this representation to predict the student's grade among five classes (A, B, C, D, F). The model is trained end-to-end by minimizing the cross-entropy loss.

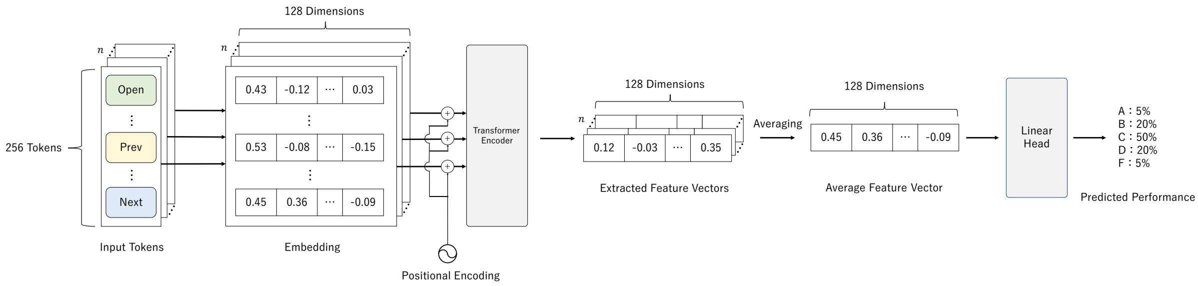

Figure 1. Overview of the proposed method

3. Evaluation Conditions and results

The dataset comprises learning logs from 1,471 students in information science courses at Kyushu University (2019–2022), with five grade categories (A, B, C, D, F). We used logs from 2019–2021 for training and those from 2022 as the test set. We trained each model 10 times and report the mean and standard deviation for the following metrics: accuracy, Macro-F1 score, and Root Mean Square Error (RMSE).

Table 2 summarizes the grade prediction results. Our proposed model outperformed both baseline methods (Histogram-based and E2Vec) across all evaluation metrics. The reduction in RMSE is particularly noteworthy, as it indicates fewer critical prediction errors (e.g., misclassifying an 'F' as an 'A') and thus more reliable predictions.

Table 1. *Accuracy Comparison*

	Accuracy	F1 Score	RMSE
Histogram	48.42% ± 1.53	46.58% ± 2.34	1.048 ± 0.040
E2Vec	54.36% ± 0.00	51.52% ± 0.00	1.311 ± 0.000
Ours	**55.60% ± 0.74**	**57.12% ± 0.57**	**0.725 ± 0.028**

4. Conclusion

In this study, we proposed a Transformer-based method to predict student grades by modeling relationships in their learning logs. Our experiments confirmed that this approach outperforms existing baselines on all metrics. For future work, we plan to enhance predictive accuracy by integrating contextual data, like the content and structure of lecture slides.

Acknowledgements

This work was supported by JST CREST Grant Number JPMJCR22D1, Japan.

References

Kohama, H., Ban, Y., Hirakawa, T., Yamashita, T., Fujiyoshi, H., Itai, A., & Usami, H. (2023). Recommending learning actions using neural network. *International Conference on Computers in Education*. https://doi.org/10.58459/icce.2023.946

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

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.