RESEARCH-ARTICLE

GLU-Transformer for Predicting Withdrawal in VLE

HAOXIN LUO

QIANYU YU

SHAOHAO XIE

HAOHAI WU

KAIBIN XU

Published: 15 November 2024

ICBAR 2024: 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management
November 15 - 17, 2024
Guangdong, Chengdu, China

ICBAR '24: Proceedings of the 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management (November 2024)
https://doi.org/10.1145/3718751.3718887
ISBN: 9798400709753

GLU-Transformer for Predicting Withdrawal in VLE

Haoxin Luo
Shantou Polytechnic
Shantou, Guangdong, China
hxluo@stpt.edu.cn

Qianyu Yu*
Shantou Polytechnic
Shantou, Guangdong, China
qyyu@stpt.edu.cn

Shaohao Xie
Shantou Polytechnic
Shantou, Guangdong, China
shxie@stpt.edu.cn

Haohai Wu
Shantou Polytechnic
Shantou, Guangdong, China
hhwu@stpt.edu.cn

Kaibin Xu
Shantou Polytechnic
Shantou, Guangdong, China
kbxu@stpt.edu.cn

Abstract

Virtual learning is a new and effective mode of education, where predicting students' early departure in advance helps make a timely strategy to keep students. By analyzing the click behavior of students in the open-source Open University Learning Analytics dataset, we propose GLU-Transformer to predict the probability of students leaving early. Experiments show that the model proposed in this study has stable and good performance. By means of using the data on students' first 25 weeks of activity in OULA dataset for prediction, the model achieves 96.54% accuracy, 97.68% precision and 92.57%recall.

CCS Concepts

• **Computing methodologies** → Artificial intelligence; • **Applied computing** → Education; E-learning.

Keywords

deep learning, transformer, virtual learning environment (VLE)

ACM Reference Format:
Haoxin Luo, Qianyu Yu, Shaohao Xie, Haohai Wu, and Kaibin Xu. 2024. GLU-Transformer for Predicting Withdrawal in VLE. In *2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR 2024), June 28–30, 2024, Chengdu, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3718751.3718887

1 Introduction

With the research attention of the researchers and the educators, virtual learning environments have become an effective educational model in the field of modern education, especially since the coronavirus pandemic, the educational model of virtual learning has received more attention. Although virtual learning provides students with a more flexible and convenient way to learn, how to further improve the intelligence level of the virtual learning environment and make it better adapt to the individual needs of students is still challenging in the current education field. As a cutting-edge artificial intelligence technology, deep learning provides a new opportunity for the intelligent development of virtual learning environment [1, 2].

By simulating the operation mechanism of the human brain neural network, deep learning can automatically extract and learn complex features in the data, so as to achieve intelligent decision-making and prediction. In the virtual learning environment, deep learning can be applied to a lot of aspects, such as learning resource recommendation, learning path optimization, learning emotion recognition, etc., in order to provide more personalized and accurate learning support for students [3].

This study aims to investigate the utilization of deep learning within virtual learning environments, leveraging the accessible Open University Learning Analytics Dataset (OULAD) to examine student behavior. Specifically, it focuses on integrating students' clickstream data into a unified vector and exploring the influence of this virtual environment on their educational performance [4].

Here are the contributions from our research:

- Put forward a deep neural network model called GLU-Transformer to forecast the possibility that students may abandon a course in a virtual learning environment(VLE);
- Compare and review previous studies using deep learning techniques to predict outcomes in VLE where students are likely to abandon a course;
- In the case of predicting early student dropout, try to use a variety of methods for improving the accuracy of GLU-Transformer. According to the testing, the model was able to correctly predict whether a student would abandon a course by week 25 of the semester, with 96.54% accuracy and 97.68% precision.

This paper is based on the OULAD dataset, which records students' demographics, performance, learning behaviors and other information in an anonymized form, and needs to go through preprocessing steps to convert the data into actionable and effective information. Therefore, we preprocess the weekly student clickstream information on the basis of the open source of Hassan et al. [5] to form a vector of students' behavioral characteristics, and then send it to our model for training, and adjust the model after testing and feedback data until getting the best results.

*Corresponding author

ICBAR 2024, Chengdu, China

ACM ISBN 979-8-4007-0975-3/2024/06
https://doi.org/10.1145/3718751.3718887

The paper is structured as follows: Section 2 discusses the research literature on the application of AI in virtual learning environments in detail. Section 3 describes the dataset and the methodology used to implement the research in this paper. Section 4 discusses the experimental results of the predictive model. In the final section, there is a summary about the paper.

2 RELATED WORK

In 2015, Khobragade et al. employed the Naive Bayes and decision tree algorithms to forecast the likelihood of student failure in a particular course. Additionally, they utilized survey and reporting data to assess the risk of student dropout. The comprehensive dataset encompassed student scores, family background information, social characteristics, and previous academic performance. Upon comparison, the Naive Bayes algorithm had higher accuracy and overall performance compared to decision trees, achieving 87% accuracy [6].

In 2016, Wei-Xiang et al. presented a combined approach that integrates Machine Learning Feature Selection (MLFS) with Support Vector Machine (SVM) to identify and validate key features that significantly influence students' academic performance in an elementary school in Taiwan. When compared to other machine learning classification algorithms, this proposed method demonstrates its effectiveness and achieves a higher accuracy rate of 92.39% [7].

Rizvi et al. conducted a comprehensive study exploring the dynamic effects of six demographic characteristics on online learning outcomes. By analyzing a diverse sample of 8581 UK learners enrolling in four different online courses offered by the Open University, the researchers uncovered significant correlations between regional factors, community poverty levels, and prior educational backgrounds with the success of online learning [8].

Fei et al. recognized that students' behavioral traits are continuously captured over a period of time, showing that dropout prediction is essentially the prediction problem of a time series, so they proposed a recurrent neural network (RNN) model dropout prediction method with long short-term memory (LSTM) cells, in which they interpret the generated data as a time series problem by calculating the learner's log data on a weekly basis, analyzing student behavior and then predicting dropout rate. The recurrent neural network (RNN) model, equipped with long short-term memory (LSTM) cells, significantly outperforms both the baseline methods and their other proposed techniques by a substantial margin [9].

Hassan et al. [5] harnessed deep learning techniques to anticipate students' dropout risks, specifically leveraging deep long short-term memory (LSTM) models to forecast early dropout tendencies through insightful data on students' engagements with the online educational platform. When compared to traditional machine learning approaches, such as logistic regression and artificial neural networks, the implemented LSTM model demonstrated superior proficiency in predicting early dropout cases. This model surpassed other methodologies in terms of learning accuracy, precision, and recall, achieving remarkable figures of 97.25% accuracy, 92.79% precision, and 85.92% recall. This method has the potential to be widely applied in the field of education to help educational institutions identify and figure out students' problems in time and improve the quality of education.

Chen et al. [10] proposed an intelligent framework for explainable student performance prediction (ESPP) for explainable student performance prediction to provide interpretability of predicted results. The framework first leverages the weekly student activity dataset of time series, and uses a mixed data sampling method to work out the problem of dataset imbalance in the virtual learning environment.

Al-Tameemi et al. [11] brought forward a deep neural network-based prediction model for students' academic performance based on educational resources such as OULAD, which uses the interaction data in the students' virtual learning environment (VLE), such as the number of clicks, assessment scores, and so on to measure the correlation with students' final grades and consequently make an accurate prediction on students' academic performance. Ahmed et al. [12] thoroughly utilized the OULAD dataset to investigate and introduce a deep learning framework aimed at predicting student performance in a virtual learning setting. Following dataset preprocessing, their proposed deep neural network model exhibited impressive results, achieving a high accuracy rate of approximately 91.29% and a low loss value of around 0.18.

3 METHODOLOGY

3.1 Data description

The dataset utilized in this paper is the Open University Learning Analytics Dataset (OULAD), encompassing data on students from the Open University during 2013 and 2014. This comprehensive dataset includes demographic details, login patterns, and assessment behaviors of 32,593 students across nine months of courses. Notably, OULAD is freely accessible on https://analyse.kmi.open.ac.uk/open_dataset. Certified by the Open Data Institute [4], OULAD documents student performance in diverse subjects ranging from social sciences to Science, Technology, Engineering, and Mathematics (STEM). The courses are structured into seven distinct modules, each taught at least twice a year with varying intervals, ensuring a diverse and enriching educational experience. The performance of the students was divided into four levels, and after statistics, it was found that 9% were excellent students, 38% passing students, 22% failing students, and 31% drop-out students, as shown in Table 1. In the raw data, the number of clickstreams records the interaction between students and the online environment. Student interactive data includes a variety of activities across heterogeneous modules, where each activity represents a student's learning behavior in an online learning environment, including accessing course materials, submitting assessments, participating in discussion forums, participating in video discussions, etc.

3.2 Data preprocessing

First, the OULA dataset connects forms such as student information sheets, assessment forms, as well as registration forms which are based on id_student attributes to achieve information beyond assessment activities. We do this by merging all the tables and removing some of the same duplicate attributes. At the same time, Sum_click attribute is added to count the number of times students interact with the VLE material each day. Through the statistics of

Table 1: of the class labels.

Label	Number of Students
Withdrawn	10156
Fail	7052
Pass	12361
Distinction	3024
Total	32593

the number of student clickstreams, it is found that the number of student clicks has a strong correlation with whether students pass the course or leave the course. This finding was first made by Hassan in [5], for which the number of clicks per campaign was aggregated on a weekly basis from the original OULA dataset, which was calculated from 20 different campaigns available in the OULA dataset, as shown in Figure 1. In the open-source dataset of Hassan et al., multiple tables from OULAD's raw data have been aggregated based on student anonymous ID numbers into a table that aggregates the number of clicks on all activities for each student over 38 weeks. Although OULAD also contains other information of students, such as gender, age, region, and other demographic information. However, in this paper, we mainly focus on the relationship between the click prevalence of 20 different activities of each student in VLE and early exit. So we preprocessed the data based on the open-source dataset of Hassan et al., and set the labels of the top and passing students as passing while those of the early leavers as early exit. Then the click information of each student's 20 different activities is used as the model training to apply to our model.

3.3 Deep learning for Student Performance Prediction

Transformer was an attempt-to-learn model architecture used to deal with natural language processing (NLP) and sequence-to-sequence tasks proposed by Vaswani et al. [13] in 2017. The Transformer architecture introduces a self-attention mechanism which is a key innovation allowing it to perform excellently when processing sequence data. Self-Attention is one of Transformer's innovative core concepts, enabling the model to take all positions in the input sequence into considerations at the same time, rather than handling stepwise like RNNs or CNNs. The self-attention mechanism allows the model to give different attention weights according to different parts of the input sequence, thus neatly capturing semantic relationships.

The internal self-attention operation is depicted as follows: three matrices are generated from a single input matrix E_n and its own weights: Query Q, key K, and value V, such that,

$$Q_i^l = E_n W_i^q, \quad K_i^l = E_n W_i^k, \quad V_i^l = E_n W_i^v \tag{1}$$

From (1), with three matrices as input, the attention operation can be performed with,

$$Z_i^l = softmax\left(\frac{Q_i^l K_i^{l^T}}{\sqrt{d}}\right) V_i^l \tag{2}$$

where d is the dimension of Q_i and K_i; Equation 2) weights the matrix value V_i based on Q_i and K_i for all input sequences of E_n. In our model, we use a multi-head attention layer. In contrast to them in a single attention head, in a multi-head attention layer, the queries, keys and values are swapped linearly by different linear projections, and then these values are concatenated and projected again to get the final output value, such that,

$$Z_n^l = concat\left(Z_1^l, Z_2^l, \ldots, Z_h^l\right) W^l \tag{3}$$

Where W^l is the weight to concatenate $Z_{\rangle}^l$ such that it learns the cross-relation of all individual heads.

Transformer has a good performance in time series prediction. Considering that the dataset we researched is to predict the probability of whether students leave early, which is a binary classification problem, and that the sequence and dimension of the dataset are relatively small, we improve Transformer comparing with the datasets in other fields. In the work, two layer transformer encoders with 512 hidden nodes are used, and GLULayer is added after the second layer encoders in place of the traditional activation function for the purpose of excelling the convergence speed when training the model and obtaining better performance. At last, the linear layer and the Softmax layer are used to binarize the output sequence.

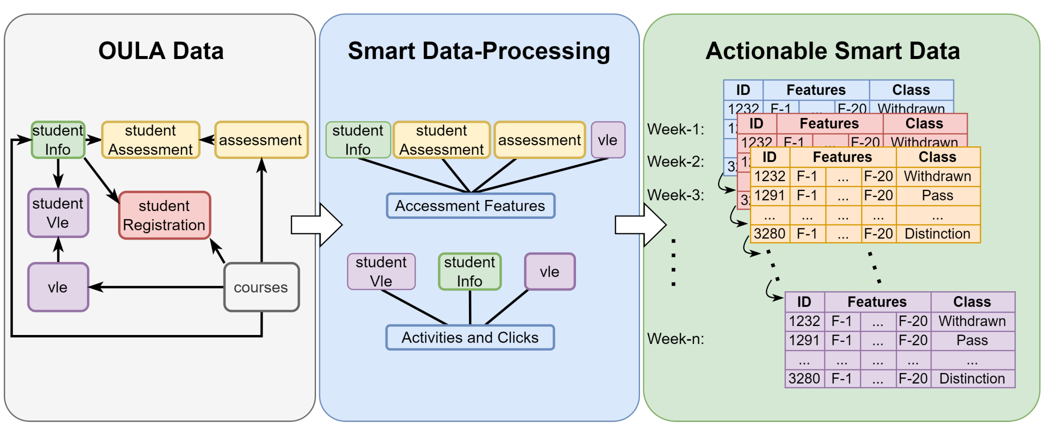

Figure 1: Steps from raw data to actionable smart data.

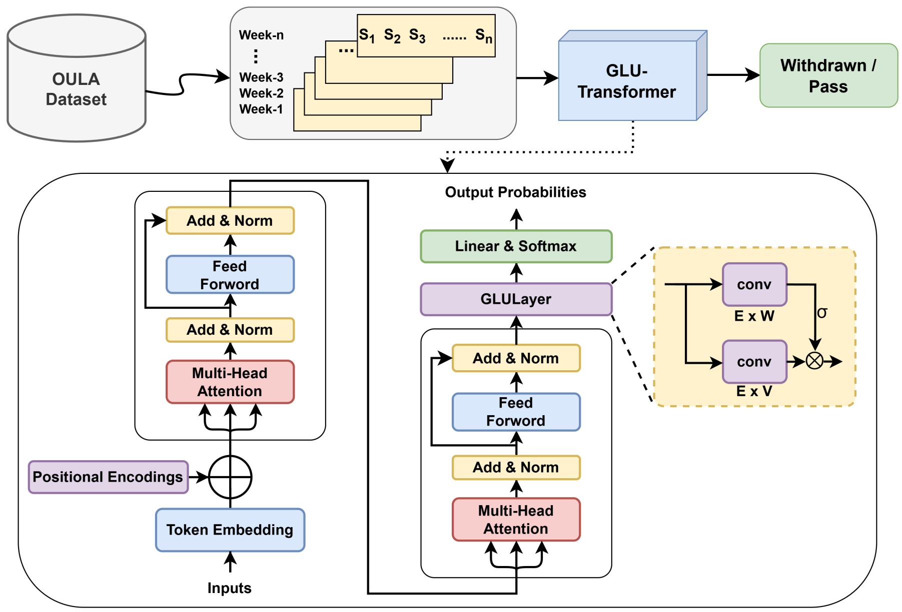

Figure 2: The proposed framework of GLU-Transformer for VLE, consisting of two phases: (1) Data Pre-processing: Count the number of clicks on different activities for each student each week; (2) Performance analysis and prediction.

GLU is the structure of a gated linear unit proposed by Yann et al. [14] in 2017, which is an activation function that dynamically regulates the flow of information by learning gating mechanisms, such as that,

$$h_l(X) = (X*W) \otimes \sigma(X*V) \tag{4}$$

Where σ is the sigmoid function, $\otimes$ is the element-wise product, X is the input data, and W, V are the parameter matrices to be learned.

Our model framework is mainly divided into two stages, as shown in Figure 2. To begin with, obtain data from the VLE system. And then preprocess the data, and count the number of clicks on different activities of each student per week, until forming an input sequence. Finally, the formed sequence was input into the GLU-Transformer for prediction.

4 EVALUATION AND RESULT

This section discusses the experiments executed on GLU-Transformer, and the specific process is as follows:

Step 1: Preprocess the OULA dataset to obtain the sequence data of the number of clicks of each student's different activities per week which are divided into a training dataset and a test dataset later.

Step 2: Input the training data into our model for training. Use AdamW optimizer to dynamically adjust the learning rate, and update the model weights.

Step 3: Using the test data into the trained model to achieve the test results.

Our model runs on a server with one GTX3060 GPU, with is implemented by PyTorch toolkit. Finally, we compare the test results with the results of other machine learning methods, so as to obtain conclusive results.

4.1 Experimental study

In this paper, the "pass" class is combined with the "distinction" class into a single class with 15,385 instances, while the "withdrawal" class has 10,156 instances. For the purpose of predicting the probability of student retreat as early as possible in the study, training tests are conducted with student data at 5 weeks, 10 weeks, 15 weeks, 20 weeks, and 25 weeks, respectively. The specific experimental results are shown in Figure 3.

As can be seen from the results depicted in Figure 3, the performance of our model proposed gradually improves with the number of weeks of input data increasing. When the input data is the number of clicks of the student for 5 weeks, the accuracy is only 80.07%. However, while the input data is the number of clicks of the student for 25 weeks, the accuracy can reach 96.54%. That is, by means of using input data of 25 weeks for prediction, there is the probability of 96.54% that a particular student drops out of a course. Similarly, the recall plot and precision plot also verify this view, indicating the good and stable performance of the model.

4.2 Comparison of deep learning model

To better evaluate the performance of the proposed model, in the study, the widely used LSTM, and LR models [5], are selected as baseline comparisons. Table 2 describes the comparison between the baseline models and the GLU-Transformer, recording the scores for 5 weeks, 10 weeks, 15 weeks, 20 weeks, and 25 weeks respectively. When the input data volume is small, the GLU-Transformer already performs better compared to the baseline methods. Whereas the input data is for 25 weeks, LSTM's recall and precision are both lower than those in the GLU-Transformer which is the method in this study, with the similar accuracy of them. Hassan et al. [5] obtains 92.79% precision and 85.92% recall

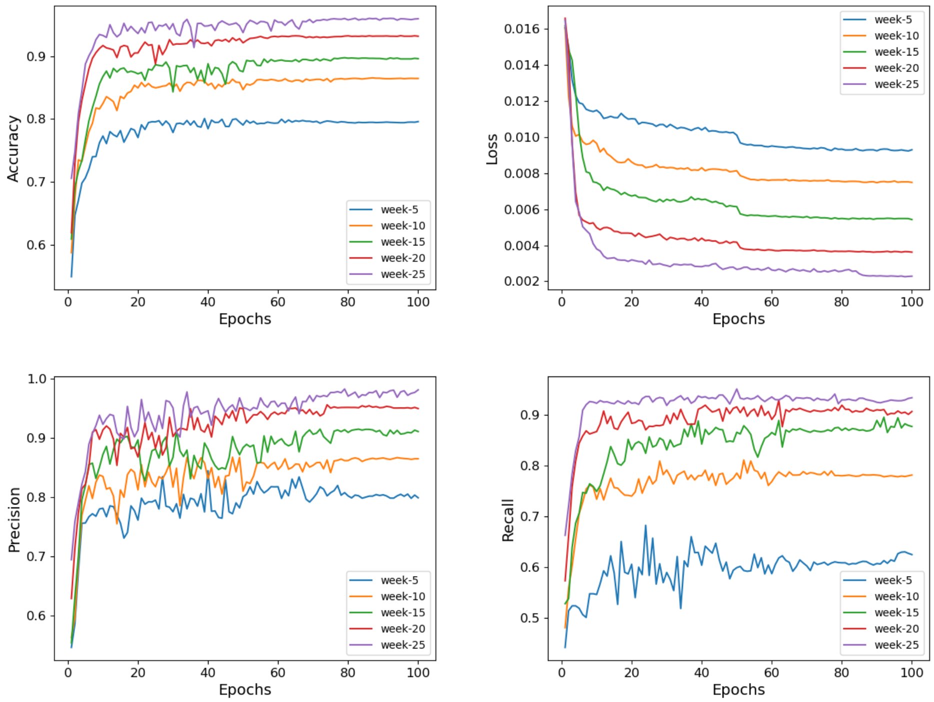

Figure 3: Week-wise learning accuracy, loss, validation precision and recall curves.

Table 2: Comparison of GLU-Transformer with other baseline approaches.

	Weeks	Accuracy (%)	Precision (%)	Recall (%)
LR	5	72.15	65.12	64.01
	10	74.48	67.03	70.23
	15	76.98	69.54	74.13
	20	79.35	72.06	78.13
	25	88.16	83.25	80.03
LSTM	5	80.78	76.57	62.08
	10	84.15	76.89	67.65
	15	93.33	85.98	75.93
	20	95.30	90.07	78.93
	25	97.25	92.79	85.92
GLU-Transformer	5	80.07	81.73	62.06
	10	85.65	86.67	74.15
	15	89.76	91.35	82.96
	20	93.26	95.02	88.24
	25	96.54	**97.68**	**92.57**

by LSTM, while our method of GLU-Transformer achieves 97.68% precision and 92.57% recall.

Therefore, from the results given in Table 2, it can be concluded that GLU-Transformer has the superiority and robustness in predicting early withdrawal than other methods.

5 CONCLUSION AND FUTURE WORk

Making an early prediction of student dropout in virtual learning environments (VLE) allows for timely and effective strategies to retain online learners. There is a strong correlation between students' clickstream behavior towards different activities in VLE data and their likelihood of withdrawal. In this study, the prediction task of student dropout is categorized as a sequence classification task, and the GLU-Transformer model is proposed to predict student dropout. The model achieves 96.54% accuracy, 97.68% precision and 92.57%recall utilize data on students' first 25 weeks of activity. In the future, we aim to integrate the results of predicting student performance by GLU-Transformer, into the framework of educational systems.

References

[1] Hubalovsky, S., M. Hubalovska, and M.J.C.i.H.B. Musilek, Assessment of the influence of adaptive E-learning on learning effectiveness of primary school pupils. 2019. 92: p. 691-705.
[2] Nguyen Q, Huptych M, Rienties B Linking students' timing of engagement to learning design and academic performance. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. ACM; 2018:141–150.
[3] Yin W. An artificial intelligent virtual reality interactive model for distance education[J]. Journal of Mathematics, 2022, 2022: 1-7.
[4] Kuzilek, J., M. Hlosta, and Z.J.S.d. Zdrahal, Open university learning analytics dataset. 2017. 4(1): p. 1-8.
[5] Hassan S U, Waheed H, Aljohani N R, *et al.* Virtual learning environment to predict withdrawal by leveraging deep learning[J]. International Journal of Intelligent Systems, 2019, 34(8): 1935-1952.
[6] Khobragade L P, Mahadik P. Students' academic failure prediction using data mining[J]. International Journal of Advanced Research in Computer and Communication Engineering, 2015, 4(11): 290-298.
[7] W.-X. Liu and C.-H. Cheng, "A hybrid method based on mlfs approach to analyze students' academic achievement," pp. 1625–1630. 08, 2016.
[8] Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning; A decision tree based approach. Computers & Education, 137, 32–47.
[9] Fei M, Yeung D Y. Temporal models for predicting student dropout in massive open online courses[C]//2015 IEEE international conference on data mining workshop (ICDMW). IEEE, 2015: 256-263.
[10] Chen H C, Prasetyo E, Tseng S S, *et al.* Week-wise student performance early prediction in virtual learning environment using a deep explainable artificial intelligence[J]. Applied Sciences, 2022, 12(4): 1885.
[11] Al-Tameemi G, Xue J, Ajit S, *et al.* A deep neural network-based prediction model for students' academic performance[C]//2021 14th International Conference on Developments in eSystems Engineering (DeSE). IEEE, 2021: 364-369.
[12] Ahmed S, Helmy Y, Ouf S. A deep learning framework for predicting the student's performance in the virtual learning environment[C]//2022 5th International Conference on Computing and Informatics (ICCI). IEEE, 2022: 240-250.
[13] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
[14] Dauphin Y N, Fan A, Auli M, *et al.* Language modeling with gated convolutional networks[C]//International conference on machine learning. PMLR, 2017: 933-941.