



Behavior Classification of Exposition Visitors

Pattern Recognition Final Project

Team member: 張博閔 謝霖泳 劉亦傑 鍾凱恩

Introduction

We proposed 4 different model architectures to solve this problem, including **Transformer**, **LSTM/RNN**, **Random Forest** and **CatBoost**. For data preprocessing, we tried some different approaches. First, we padded 0s in front of/behind the original sequence. Second, we grouped the data according to the total # of stations visited. We want to compare the accuracy w/w.o. these preprocessing methods under the model that performs the best.

Framework

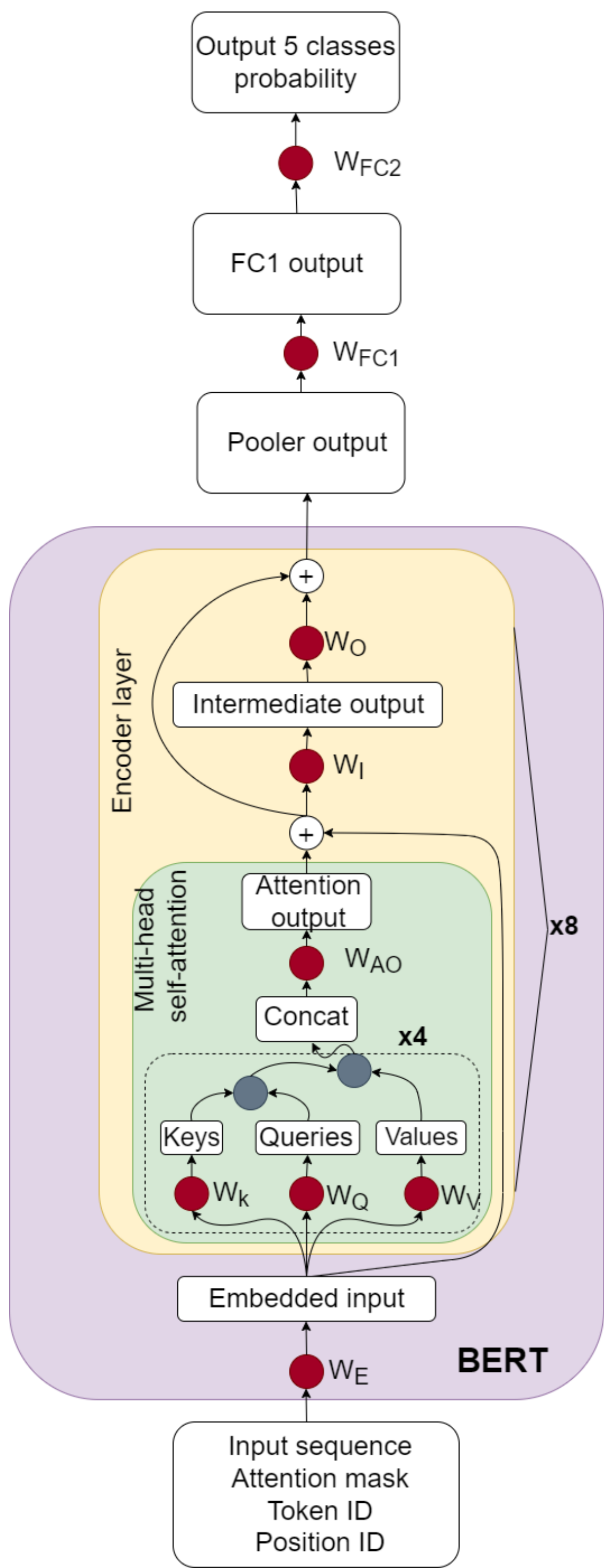


Fig. 1 The proposed method

Description

Our architecture based on BERT, which is a seq2seq model. Since we don't need to generate sequence and it's a classification problem, only the first half of transformer (i.e. encoder) is used to extract features. Originally, BERT mostly applied on NLP, so we didn't use any pretrained weights. Instead, we trained it from scratch.

Position encoding enables the model to figure out the order of input sequence, and multi-head self-attention mechanism enhances the robustness of our model. After tuning the parameters, we can acquire the result below.

Results

	Aidea score	Validation acc.
LSTM	0.1498253	0.9598
RNN	0.1368455	0.9411
Random Forest	0.1781107	0.9534
Catboost (w/ grouping)	0.1276041	0.9611
Catboost	0.1115650	0.9625
BERT (w/ grouping)	0.0991397	0.9842
BERT	0.0468931	0.9871
BERT+XLNET	0.0439630	-

排名	隊伍名稱	成績	上傳時間	次數
1	cbchen	0.0326282	2022/06/12 21:09:48	16
2	love0416much	0.0407222	2022/06/12 15:39:22	7
3	liujack	0.0439630	2022/06/13 04:26:36	20

Table. 1&2 Our best result

Summary

After our experiments, we found that accuracy is better without grouping. Besides, if we pad 0s before original sequence and select transformer with BERT architecture as our model, it performs the best, which reached the top 3 in Aidea platform.

The reason why BERT performs the best among all our methods is **multi-head self-attention**. Multi-head property enables the model to learn more from different perspectives in feature level. Besides, self-attention mechanism gives the model a wider view of the sequence than LSTM or RNN does.

Combined with both advantages mentioned above, BERT performs ultimately well in this pattern recognition project.