CANYU ZHANG

Missing Words Prediction For Long Sentence

Document damage is a common concern faced by researchers. The original paragraph might be meaningless since some key words are missing. Even though the missing word prediction has drawn more and more attention, current words sill can address this problem ideally. For example, current word prediction models always performs poorly on the long paragraph, since it's hard for the model to learning the meaning from too many words. In this paper, we focus on solving this problems. We modify the the transformer based network to make the prediction for missing words in long paragraphs. To justify the effectiveness of proposed method, we generate some long paragraph to text the performance of modified models. Proposed method generates a promising result.

Additional Key Words and Phrases: Word Prediction, Long Paragraph, NLP

ACM Reference Format:
Canyu Zhang. 2022. Missing Words Prediction For Long Sentence. 1, 1 (November 2022), 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Word missing is a common problem happens in natural language processing area, which has drawn more and more interests. Though a lot of methods have been proposed for this tasks, they are far from practical usage. Most existing models focus on predicting missing words in short sentences, the prediction ability will be worse on long sentences. Since learning the feature for long sentences is more challenging. But the long sentences will be more common in real life. In order to solve this limitation, we choose to modify the previous transformer based models for long sentence missing parts prediction. Our model can consider the whole sentences as a whole, beating previous models like RNN and LSTM. Since there are few long sentence datasets, we generate some data for training and evaluation. To justify the performance of proposed method, we conduct thorough experiments on proposed models and baselines.

2 RELATED WORK

Transformer. Transformer can analysis the different parts from sentences or images with the help of self-attention system. It can consider the whole input as whole, capturing more global information compared with previous models RNN and LSTM. It is firstly proposed for machine translation, and recently it has been used for image classification, e.g., ViT succeeds in many other image processing tasks, with superior performance over classical CNN-based models. Transformer network has also been widely used in 3D point cloud processing. For instance, Point Transformer constructs self-attention networks for 3D point cloud semantic segmentation and classification. In this paper, we also choose transformer based model to solve the word prediction task.

Author's address: Canyu Zhang, UofSC.

XXXX-XXXX/2022/11-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

3 METHOD

Our model is based on previous NLP model BERT, which is a bidirectional transformer using a combination of masked language modeling objective and next sentence prediction. It is pretrained on a large corpus comprising the Toronto Book Corpus and Wikipedia. The base model is similar to the original transformer. And we will fine-tune the pretrained BERT model for our task since training a transformer from the beginning is difficult. The input is the masked sentences, following the previous work, we use symbol [CLS] to show the beginning of a sentence, [SEP] to separate different sentences and [MASK] to indicate the masked part. The model will output the predictions for the masked words.

4 EXPERIMENTS

We train our model using Adam as the optimizer (1= 0.9, 2 = 0.999) with the learning rate 0.0001. Training epoch is 150 epochs and the batch size is set to 32. All experiments are implemented on the two NVIDIA Tesla V100 GPUs. For evaluation, we use Exact Match (EM) and F1 score to evaluate the performance of different models.

5 EVALUATION

The results from our experiments are shown in Table 1.

Metric	EM	F1
Mask Ratio		
1%		
5%		
10%		

Table 1. Results from our model under different settings.