Machine Learning 2023 Spring Homework 7 Report

生醫電資所 R11945071 莊聰賢

1. After your model predicts the probability of answer span start/end position, what rules did you apply to determine the final start/end position?

Since in some case, the evaluation results of start\_index will bigger than end\_index, I write an if-else statement in my script to let the function ignore this kind of prediction.

1. Try another type of pretrained model which can be found in huggingface’s Model Hub (e.g. BERT -> BERT-wwm-ext, or BERT -> RoBERTa ), and describe
   1. the pretrained model you used

chinese-macbert-large

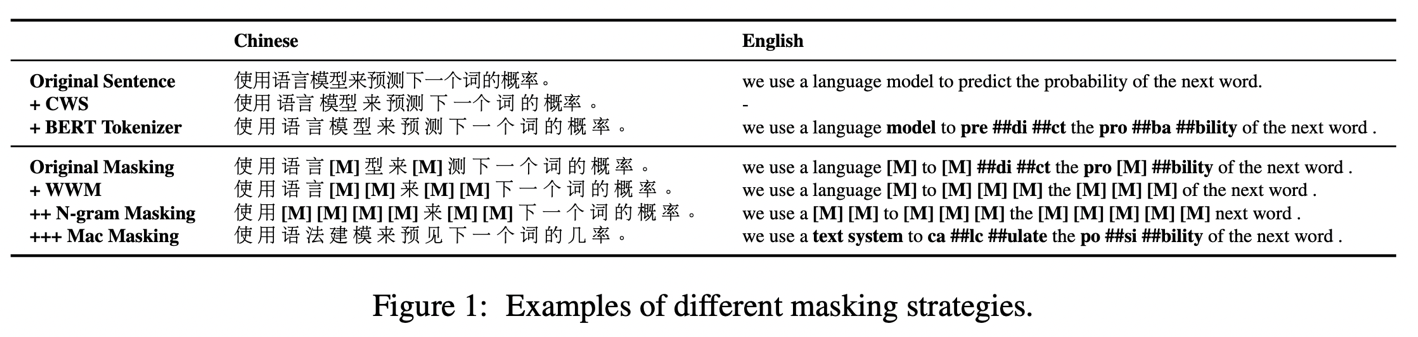
* 1. performance of the pretrained model you used

Kaggle public score: 0.82533

* 1. the difference between BERT and the pretrained model you used (architecture, pretraining loss, etc.)

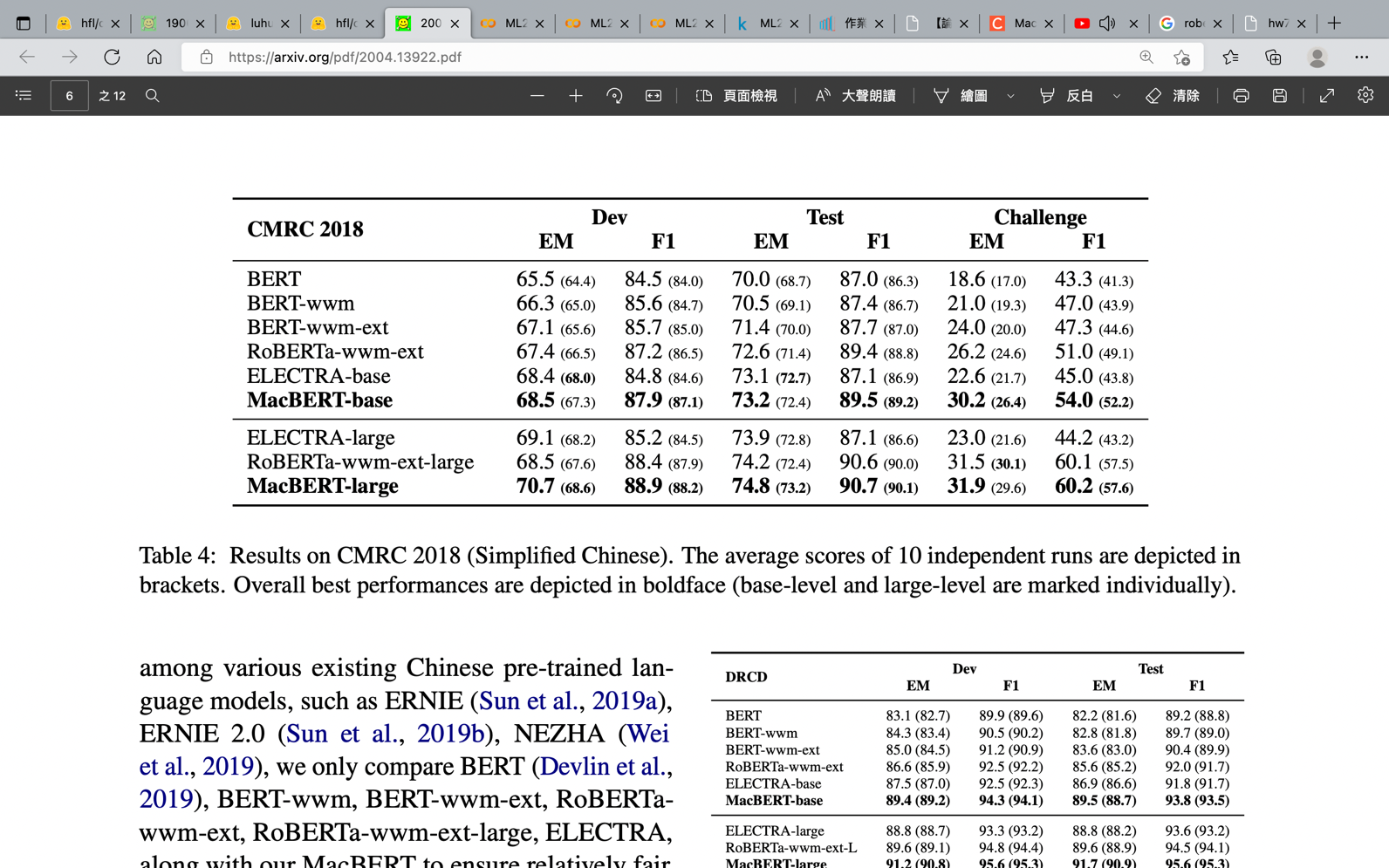
Following are some techniques used when pretraining MacBERT (MLM as correction BERT):

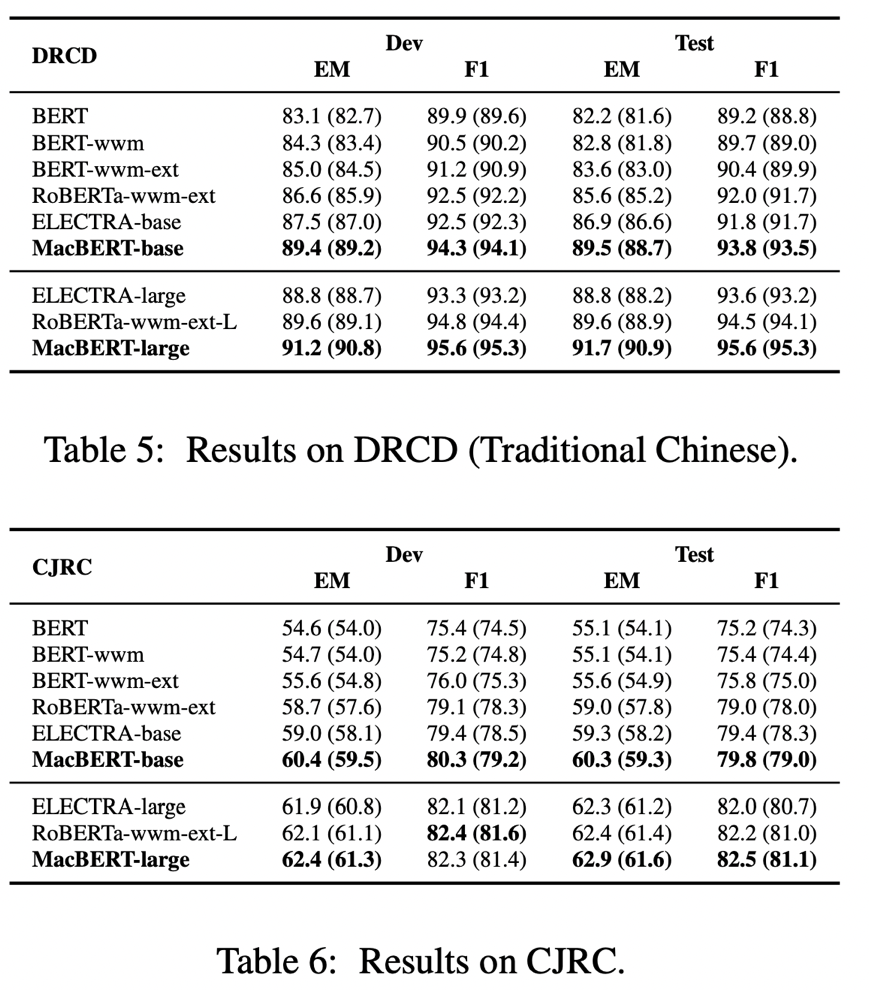
* + - In MLM (masked language model) task:
      1. Using N-gram masking strategy when applying whole word masking
      2. To eliminate the gap between pretraining task and fine-tuning task, they used similar words generated based on word2vec to be the mask
      3. 15% of input words will be masked, 80% of them will be changed into similar words, 10% of them will become random words, and 10% of them will keep original.



* + - In NSP (next sentence prediction) task:
      1. Using sentence-order prediction introduced by ALBERT.

Following three tables are the results of MacBERT in MRC (machine reading comprehension), and in most of cases, MacBERT did get the best results.





1. References:
   1. Cui, Yiming, et al. "Revisiting pre-trained models for Chinese natural language processing." *arXiv preprint arXiv:2004.13922* (2020).