

ML HW7 Report

B09901066 謝承修

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? (the rules you applied must be different from the sample code)**

I add two rules to give a better performance, one is to check $\text{end_index} - \text{start_index} \leq 30$, it makes the answer more reasonable, and the other is to check whether $\text{end_index} > \text{start_index}$, which makes sense easily.

2. **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 : the pretrained model you used, performance of the pretrained model you used, the difference between BERT and the pretrained model you used (architecture,pretraining loss, etc.)**

I use *luhua/chinese_pretrain_mrc_macbert_large* to be the pretrained model, and the score on Kaggle is 0.78983, which is better than BERT (0.76643) when other parameters remain the same . The model uses more Chinese data (Chinese MRC macbert-large) than BERT, making the pretraining loss lower, and it has architecture with 無答案數據構造 etc. To know more information, there's a link to the document.

https://huggingface.co/luhua/chinese_pretrain_mrc_macbert_large