

## HW7

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)

處理 [PAD], [SEP], [UNK] 出現在答案中的情形

[PAD], [SEP] 以空白取代，[UNK] 使他變回原本的 paragraph

處理 start, end 的 index

先找 start index 在從 start index 後面找 end index，可以避免 end index < start index 的情況

限制 end index-start index 的值

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

Pretrained model	hfl/chinese-roberta-wwm-ext-large
Performance	0.81776 (in kaggle)
Difference	1.預訓練階段採用 wwm 策略進行 mask，動態調整 Masking 機制 2.取消 Next Sentence Prediction ( NSP ) loss 3.訓練時間更長，batch size 更大，訓練資料更多； 4.訓練序列更長；