Report Questions

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)

   model 會選擇開始機率與結束機率相加最大為預測結果，由於 model 預測結果可能發生 start_index 比 end_index 後面的情形，因此加上 start_index 必須大於 end_index 的條件:

   if 0 < end_index - start_index :
   if prob > max_prob:
   max_prob = prob
   …

   確保在有答案的情況下有最大的機率。

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

     BERT -> BERT-wwm-ext

- performance of the pretrained model you used

The performance of **BERT** in kaggle is **0.58168**
The performance of **BERT-wwm-ext** in kaggle is **0.61113**.

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

BERT-wwm-ext 更改了原預訓練階段的訓練樣本生成策略:

Whole Word Masking (wwm): 英文最小的 token 是一個單詞，而中文最小的 token 是字，但以中文來說，詞通常會包含較多的資訊，若以字為單位可能導致模型訓練不佳。因此相對於 BERT 使用 WordPiece 的方法，wwm 將 Mask 替換成一個完整的詞而不是字，讓 BERT 更好的應用在中文任務。

並且 BERT-wwm-ext 還增加了訓練數據集同時也增加了訓練步數:

1. 預訓練數據集做了增加，次數達到 5.4B；
2. 訓練步數增大，訓練第一階段 1M 步，訓練第二階段 400K 步

```
Epoch 2 | Step 2900 | loss = 1.412, acc = 0.514
Epoch 2 | Step 3000 | loss = 1.244, acc = 0.522
Epoch 2 | Step 3100 | loss = 1.284, acc = 0.537
Epoch 2 | Step 3200 | loss = 1.332, acc = 0.550
Epoch 2 | Step 3300 | loss = 1.325, acc = 0.524
Epoch 2 | Step 3400 | loss = 1.339, acc = 0.532
Epoch 2 | Step 3500 | loss = 1.352, acc = 0.506
Epoch 2 | Step 3600 | loss = 1.368, acc = 0.516
Epoch 2 | Step 3700 | loss = 1.348, acc = 0.536
Epoch 2 | Step 3800 | loss = 1.353, acc = 0.541
Epoch 2 | Step 3900 | loss = 1.215, acc = 0.554
Evaluating Dev Set ...
  0%|          | 0/4131 [00:00<?, ?it/s]
Validation | Epoch 2 | acc = 0.569
```

BERT pretrained model

```
Epoch 2 | Step 2900 | loss = 1.215, acc = 0.550
Epoch 2 | Step 3000 | loss = 1.056, acc = 0.590
Epoch 2 | Step 3100 | loss = 1.128, acc = 0.579
Epoch 2 | Step 3200 | loss = 1.176, acc = 0.561
Epoch 2 | Step 3300 | loss = 1.119, acc = 0.566
Epoch 2 | Step 3400 | loss = 1.205, acc = 0.558
Epoch 2 | Step 3500 | loss = 1.234, acc = 0.549
Epoch 2 | Step 3600 | loss = 1.208, acc = 0.564
Epoch 2 | Step 3700 | loss = 1.190, acc = 0.585
Epoch 2 | Step 3800 | loss = 1.160, acc = 0.575
Epoch 2 | Step 3900 | loss = 1.078, acc = 0.601
Evaluating Dev Set ...
  0%|          | 0/4131 [00:00<?, ?it/s]
Validation | Epoch 2 | acc = 0.601
```

BERT-wwm-ext pretrained model