HW7　— R10945015 黃資荃

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)

Since the answers are not always near the middle of the window, I use np.random.randint to randomly choose the start/end position.

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 : chinese-roberta-wwm-ext-large.
   - performance of the model you used : 0.77.
   - the difference between BERT and the pretrained model you used (architecture, pretraining loss, etc.)

   1、相較於 BERT 使用固定的 Mask, RoBERTa 在 training 時會改變 Mask 的位置, 提高輸入的隨機性
   2、調整 optimizer Adam 的參數 (ex.β、ε)
   3、RoBERTa 模型使用更多的訓練資料 (160G), 比起 BERT 的16G 可以提升模型準確度