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)
2. Add limit start\_index < end\_index
3. end\_index – start\_index < 40 in order to make sure that the answer is probable.
4. Try another type of pretrained model which can be found in [huggingface’s Model Hub](https://huggingface.co/models?search=chinese) (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.)

Model : Chinese-bert-wwm-ext

Performance: both pretraining loss and downstream task performance(which is this HW) is slightly better than BERT

Architecture difference: wwm means whole world masking, its improvement is to assure that a whole world but not only a single character in a word is masked.