1.(2%) 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)

Ans:

I select the top 30 highest probability of start and end positions, find a combination such that the total probability is the highest, end index is bigger than start index and start index is within the paragraph.

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. pretrained model: luhua/chinese_pretrain_mrc_macbert_large
ii. performance of the pretrained model: single: around 0.83, ensemble3: 0.85
iii. MacBERT is an improved BERT with novel MLM as correction pre-training task. Instead of masking with [MASK] token, which never appears in the fine-tuning stage, MacBERT use similar words for the masking purpose. there is no differences in the main neural architecture between original bert and macbert

Reference:

https://github.com/basketballandlearn/MRC_Competition_Dureader
https://huggingface.co/hfl/chinese-macbert-large