ML HW7 Report

物理三 潘阜承 B08202036

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?**

First, to deal with the problem that end position might be smaller than the start position, I add a rule (end_index >= start_index) so that only those cases with end position larger than start position will be taken into consideration. Furthermore, to restrict the length of the answer, I add another rule (end_index – start_index <= 30) to filter out answers that are too long.

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 its properties.**

For this part, I chose MacBERT as my pretrained model. MacBERT reaches accuracy of 0.805 for the validation set in our homework. Also, after ensemble, it reaches accuracy of 0.84832 in the public test set. MacBERT is an improved BERT which replace the original MLM task into "MLM as correction" task, which the details can be found in the reference below. What's more, MacBERT uses LAMB as the optimizer instead of AdamW which is used in the original BERT. Other than those features, the neural architecture of MacBERT is basically the same as BERT.

Reference:

1. https://arxiv.org/pdf/2004.13922.pdf