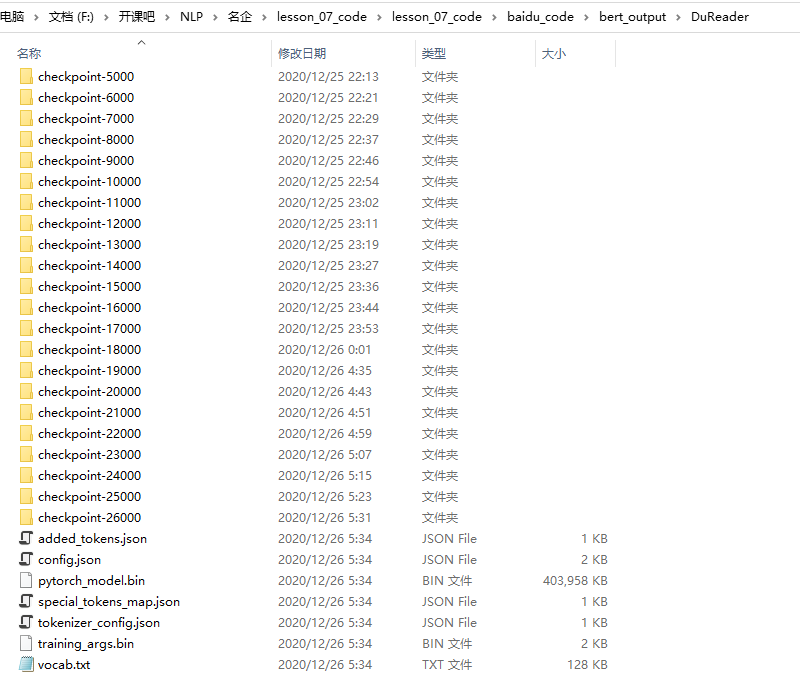
1跑通结果存储于



简答题

1 bi-DAF输入层的构建方式

Character Embedding 指的是将context和question经过char级别的embedding。

word embedding 指的是将context和question经过word级别的embedding。

contextual embedding 是将word的Embedding 和char的Embedding

合并起来经过一个Highway网络，相当于一个门控思想，使一部分的输入直接通过，其余一部分通过一个转换网络，主要公式如下

IMG_256

2 解释BERT中的概念：

Bert为了增加词向量模型泛化任务，充分描述词级别的关系特征，实现真正的双向encoding mask LM 它是一个预训练词级别的任务，通过对句子中一部分词的遮罩，来预测词的任务,一般采用15%的token随机mask,但是又为了与下游任务中没有mask匹配，所以按照一定的比例在需要预测的词位置上输入原词或者随机某个词。一般80%用[mask],10%用随机采用的一个单词替换，10%不做替换。

positional encoding 是因为计算attention的时候，没有考虑位置信息，不同位置相同词的attention计算起来是一样的，这显然不符合语法关系。所以加入位置编码来实现相同词的不同位置计算出来的权重是不同，从而它们的重要性也是不同。

IMG_256

IMG_256