

周报

屈原斌

2020 年 6 月 22 日

1 上周任务

- 重新看了一下那篇论文；
- 写代码；

1.1 BLM 代码复现

- 数据处理部分有一些问题：
 - ptb数据集中有大量的[UNK]这个要怎么处理？
 - token掩盖是随机的，所以给定固定生成的顺序这个要怎么去设定？（可不可以不用给定，就直接按照softmax之后的top1作为下一个要生成的空白）
 - 生成新空白的分类器标签要如何处理？

2 本周计划

- 先看一下mlm代码，看一下随机mask过程和损失计算；
- 继续写代码；