

周报

屈原斌

2020 年 8 月 17 日

1 上周任务

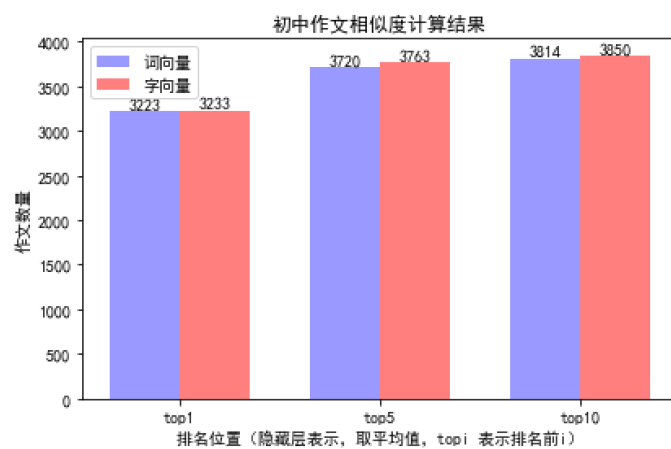
- 修改相似度计算代码

1.1 相似度计算

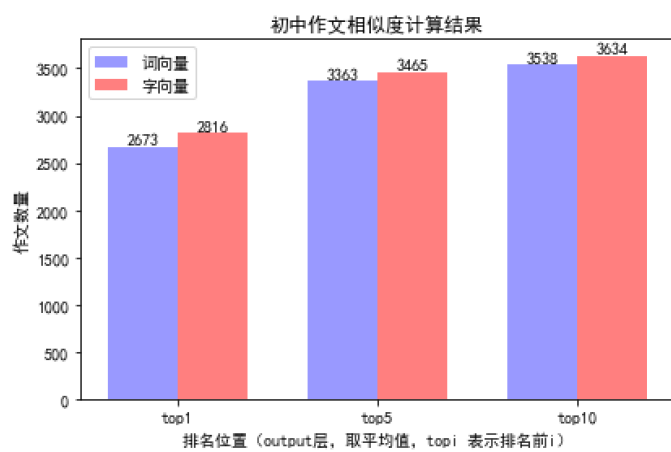
- 实验设置：
 - 范文集：每个主题抽取10篇，共4140篇；
 - 测试集：每个主题抽取10篇，共4140篇；
 - 策略：
 - * 隐藏层表示：取最后一层表示（output层）/ 取分类前一层表示；
 - * 相似度计算：全部求和取平均 / 取最大值 / 取topk求和平均；
- 实验结果（见图1，图2，图3）：
 - 大部分测试作文（3/4）可以正确区分；
 - 使用字向量结果优于词向量；
 - 使用output层表示效果略差；

2 本周计划

- 在小学作文进行测试；
- 尝试其他策略；

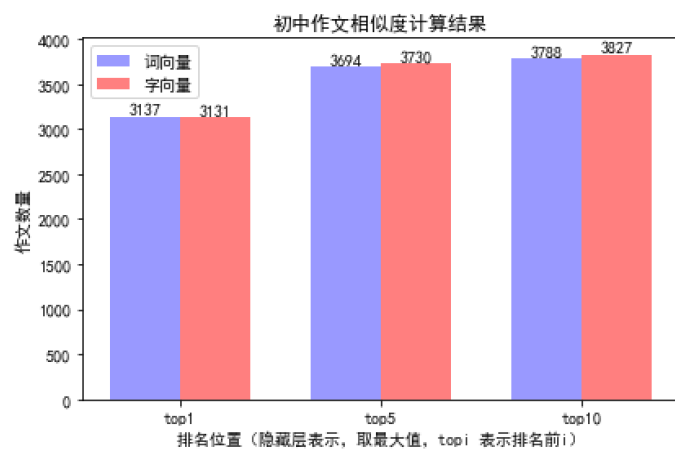


(a)

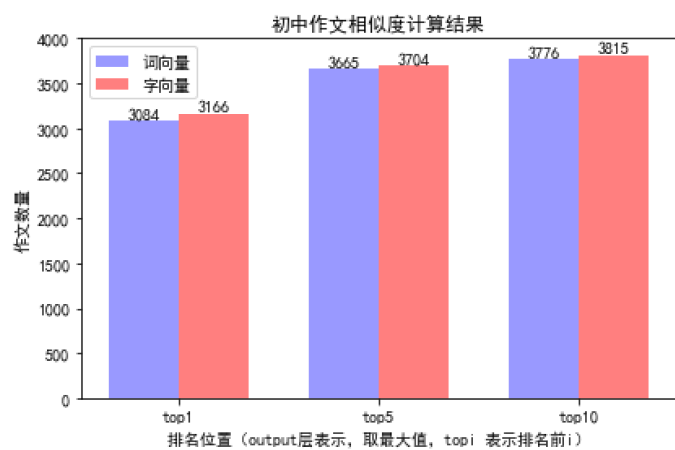


(b)

图 1: 取平均值相似度计算结果

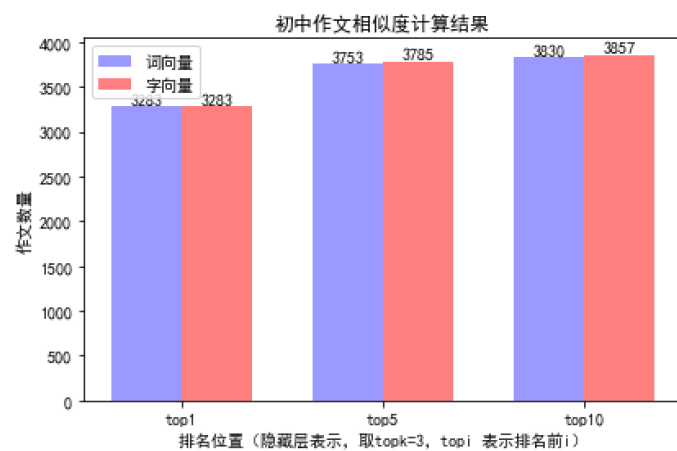


(a)

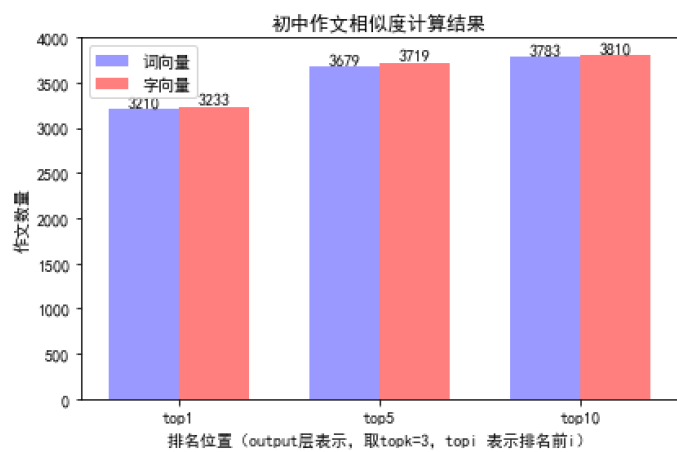


(b)

图 2: 取最大值相似度计算结果



(a)



(b)

图 3: 取topk=3相似度计算结果