

# 周报

屈原斌

2020 年 8 月 24 日

## 1 上周任务

- 补充相似度计算实验；
- 找离题相关的论文；

### 1.1 补充实验

- 实验设置：
  - 数据：每个主题抽取10篇做范文，10篇做测试；
  - 方法：第一个主题的作文在第二个主题上进行测试，依次类推；
- 实验说明：
  - 结果统计排名在top1的作文数 $n_{top1}$ ，即区分错误的作文数和错误率 $E$ （ $E = n_{top1}/N$ ， $N$ 表示全部测试作文数目）；
  - 初中作文测试集合共4140篇；
  - 小学作文（四/五年级）测试集合各50篇；
- 实验结果：
  - 结果见表1和表2；
  - 结果分析：
    - \* 初中作文上错误率保持在0.1%左右；
    - \* 小学作文上错误率保持在10%左右；
    - \* 不同策略下的结果区分不明显；

表 1: 补充实验结果（取隐藏层表示）

		取平均值	取最大值	取topk=3
初中作文	词向量	0.0015(6)	0.0002(1)	0.0005(2)
	字向量	0.0007(3)	0.0007(3)	0.0005(2)
四年级作文	词向量	0.14(7)	0.08(4)	0.06(3)
	字向量	0.06(3)	0.12(6)	0.06(3)
五年级作文	词向量	0.00(0)	0.08(4)	0.02(1)
	字向量	0.02(1)	0.04(2)	0.02(1)

\*()内表示排名在top1的作文数

表 2: 补充实验结果（取output层表示）

		取平均值	取最大值	取topk=3
初中作文	词向量	0.0015(6)	0.00(0)	0.0002(1)
	字向量	0.0007(3)	0.00(0)	0.0005(2)
四年级作文	词向量	0.12(6)	0.12(6)	0.1(5)
	字向量	0.06(3)	0.08(4)	0.04(2)
五年级作文	词向量	0.06(3)	0.04(2)	0.02(1)
	字向量	0.04(2)	0.06(3)	0.02(1)

\*()内表示排名在top1的作文数

## 1.2 论文

- 目前找到了几篇论文，还没有找全；

## 2 本周计划

- 用Bert跑一下分类实验，做相似度测试；
- 看论文，看看有没有别的实验思路；
- 继续找论文，写综述报告；