

周报，2021年05月31日

屈原斌  
首都师范大学  
ybqu@cnu.edu.cn

## 1 上周计划

1. 通用场景离题检测实验更新
2. 标题-正文匹配实验更新

## 2 上周计划执行情况

1. [✓]
2. [×]

## 3 本周部分重点工作详述

### 3.1 英文数据集

- 数据集：
  - ICLE数据集，共10个主题，离题:不离题=51:733(1:14)
- 实验策略：
  - 统计正文中包含prompt中关键词的个数，按照分位数划分，分别调参
- 指标见表1
- 结论：
  - TFIDF分布与BERT生成模型结论相反
  - TFIDF上的结果表明对于不离题作文相对 `spearman` 相关系数较高

		离题作文数	R@10	R@20	R@50	R@all	P@1	P@5	P@10	spearman	ndcg	ndcg@10
Bert-Gen	≥中位数	11	0.6833	0.6833	0.6833	0.6833	0.3000	0.1600	0.0800	0.2165	0.9863	0.8991
	<中位数	40	0.6098	0.6169	0.6383	0.6526	0.7000	0.3400	0.2200	0.1739	0.9697	0.8432
	≥1/4分位数	6	0.4500	0.5000	0.5000	0.5000	0.0000	0.0800	0.0500	0.1834	0.9870	0.9300
	<1/4分位数	45	0.5181	0.5681	0.5748	0.5748	0.6000	0.3600	0.2200	0.0234	0.9665	0.8540
TFIDF	≥中位数	11	0.4000	0.4000	0.5333	0.6000	0.1000	0.0600	0.0400	0.0890	0.9808	0.9193
	<中位数	40	0.5417	0.7036	0.7583	0.7655	0.6000	0.3800	0.2300	0.1813	0.9652	0.8340
	≥1/4分位数	6	0.3500	0.3500	0.4000	0.4000	0.1000	0.0800	0.0400	0.1715	0.8872	0.8354
	<1/4分位数	45	0.4374	0.5640	0.6440	0.6774	0.5000	0.3600	0.2300	0.1475	0.9678	0.8364

Table 1: 英文数据集实验结果

### 3.2 中文数据集

- 数据集：
  - 共2100篇测试作文，离题:不离题=350:1750(1:5)
- 实验策略：
  - 直接在测试集上调参
- 指标见表2
- 结论：
  - 增大离题比例后，指标整体下降
  - TFIDF指标较分布式表示差（结果分析）

	<b>R@10</b>	<b>R@15</b>	<b>R@20</b>	<b>R@50</b>	<b>R@all</b>	<b>P@1</b>	<b>P@5</b>	<b>P@10</b>	<b>spearman</b>	<b>ndcg</b>	<b>ndcg@10</b>
<b>TFIDF</b>	0.1703	0.2503	0.3243	0.4781	0.4781	0.7000	0.7600	0.6200	0.0113	0.8771	0.5880
<b>Skip-gram</b>	0.0799	0.0986	0.1137	0.2364	0.6238	0.7000	0.4600	0.2700	-0.0005	0.9678	0.8037
<b>Doc2vec</b>	0.0870	0.1293	0.1715	0.3027	0.5293	0.7000	0.5200	0.3100	0.0647	0.9794	0.7532
<b>Bert.Classification</b>	0.1438	0.1599	0.1822	0.2980	0.7782	0.5000	0.5200	0.5000	0.1479	0.9799	0.8257
<b>Bert-Gen</b>	0.1677	0.1810	0.2030	0.2707	0.7145	0.5000	0.5200	0.5900	0.1219	0.9638	0.8527

Table 2: 中文数据集实验结果

## 4 下周计划

1. [\*\*\*] 实验结果分析
2. [\*\*\*] 标题-正文匹配测试集check，更新实验结果
3. [\*\*\*\*] 尝试更新一些HT表示（加入关键字）