

周报，2021年1月4日

屈原斌
首都师范大学
ybqu@cnu.edu.cn

1 上周计划

1. [***] 更新离题检测实验。

2 上周计划执行情况

1. [✓×]

3 本周部分重点工作详述

3.1 $DataSet_1$ 数据集

- 去掉部分离题数据，只使用1、4档作文
- 数据分布见表1：

	作文数
完全离题 (score=1)	544
切题 (score=4)	29373
合计	29917

Table 1: $DataSet_1$ 数据分布

3.2 实验结果

- 方案：对所有测试数据表示取平均作为题目表示，计算相似度
- 实验结果见表2：

		最优Threshold	precision	recall	F1-score
HABiLstm_W	离题	75	0.2494	0.1783	0.2079
	不离题		0.9849	0.9901	0.9875
HABiLstm_W(+64)	离题	41	0.253	0.2684	0.2605
	不离题		0.9864	0.9856	0.986
Doc2vec	离题	100	0.0624	0.4154	0.1084
	不离题		0.9879	0.8843	0.9332
TF-IDF	离题	7	0.0374	0.4430	0.0689
	不离题		0.9871	0.7886	0.8768

Table 2: 指标更新

* 注：TF-IDF根据35题目抽取关键字（主题词抽取不是很准确，没有进行词扩展）

- 实验结果分析：
 - Doc2vec、TF-IDF上指标较差
 - HABiLstm_W(+64)指标最优

	accuracy	precision	recall	f1
Bert	0.4503	0.3456	0.3692	0.4356

Table 3: $DataSet_1$ 数据分布

3.3 补充实验

- 实验：利用LSTM筛选后的句子训练BERT模型
- 实验结果见表3：
- 实验结果分析：
？准确率达到0.4左右基本不升高

4 存在问题

- 问题1 35个测试题目下范文的相似度排名都是最大的，无法根据范文排名设置阈值，尝试计算每个题目的最优阈值；
- 问题2 根据段落训练分类模型指标不上升；

5 下周计划

1. [***] 修改调阈值方法。
2. [***] 复现论文中其他方法。