

周报，2020年11月09日

屈原斌  
首都师范大学  
ybqu@cnu.edu.cn

## 1 上周计划

1. [\*\*\*] 标注数据分析。
2. [\*\*\*] 重新构建 $DataSet_2$ 数据集。
3. [\*\*\*] Baseline实验。

## 2 上周计划执行情况

1. [✓×] 一致率结果统计完成，标注数据还没有看。
2. [✓] 数据构建完成。
3. [✓×] 完成部分Baseline实验。

## 3 本周部分重点工作详述

### 3.1 标注数据分析

- 第一部分数据：（5题目，共616篇作文）
- 分析：（见Table 1,2,3,4）
  - 离题作文占比60%左右；
  - 一致率较第二次试标降低；
  - 标题部分的标注转写错误标签差别较大；

Table 1: 第一部分数据双评结果（离题标注）

	score=1	score=2	score=3	score=4	总计
标注1	63	96	257	200	<b>616</b>
标注2	121	175	102	218	<b>616</b>

Table 2: 第一部分数据双评结果（标题和题目匹配标注）

	是	否	无	转写错误
标注1	340	42	232	2
标注2	329	28	230	29

- 第二部分数据：（30题目 × 50篇作文，共1500篇作文）
-

Table 3: 第一部分数据双评结果（标题和正文匹配标注）

	是	否	无	转写错误
标注1	341	41	232	2
标注2	335	22	230	29

Table 4: 第一部分数据双评结果（一致率）

	四类一致率 1,2,3,4	三类一致率 1,(2/3),4	二类一致率 (1/2/3),4	跨一档率	跨二档率	跨三档率	标题和题目匹配	标题和正文匹配
作文数	252	340	414	245	95	24	565	538
百分比(%)	40.9%	55.3%	67.2%	39.8%	15.4%	3.9%	91.7%	87.3%

- 分析：（见Table 5,6,7,8）
  - 离题作文占比70%左右；
  - 一致率较第二次试标降低；
  - 标题部分的标注转写错误标签差别较大，存在部分标注标签错误问题；

Table 5: 第二部分数据双评结果（离题标注）

	score=1	score=2	score=3	score=4	总计
标注1	335	346	390	429	<b>1500</b>
标注2	530	355	336	279	<b>1500</b>

### 3.2 $DataSet_2$ 数据集构建

- 数据量：30题目×50篇作文
- 样本数分布：见Table 9,10
- 数据构建方法：
  - 正样本：从当前题目下一类卷+二类卷抽取20篇；
  - 负样本：
    - \* 从其他题目下一类卷+二类卷抽取20篇；
    - \* 抽取其他题目下作文对段落或句子随机打乱；
    - \* 添加优美散文；

Table 9: 单个题目下作文分布

	作文数	总计
正样本	20	20
负样本	其他题目抽取	20
	随机打乱段落句子顺序	8
	优美散文	2

Table 10: 所有题目下作文分布

	作文数
正样本	600
负样本	900

### 3.3 Baseline实验

- 数据集： $DataSet_2$
- 模型：BertModel / HBiLstm / HBiLstm\*
- 注：\*表示使用字向量。

Table 6: 第二部分数据双评结果（标题和题目匹配标注）

	是	否	无	转写错误
标注1	869	118	480	33
标注2	824	112	468	95

Table 7: 第二部分数据双评结果（标题和正文匹配标注）

	是	否	无	转写错误
标注1	838	146	478	38
标注2	548	388	468	96

- 实验设置：
  - topk=3;
  - 阈值(Threshold)=1;
- 实验结果：
  - BertModel
    - \* 实验结果见Table 11;
  - HBiLstm
    - \* 实验结果见Table 12;
  - HBiLstm\*
    - \* 实验结果见Table 13;
- 分析：
  - HBiLstm 训练过后的结果优于Bert的结果；
  - 模型对优美散文和随机打乱的识别效果较差；
  -

#### 4 下周计划

1. [\*\*\*] 完成标注数据分析。
2. [\*\*\*] 完善baseline实验。
3. [\*\*\*] 分析实验结果。

Table 8: 第二部分数据双评结果（一致率）

	四类一致率 1,2,3,4	三类一致率 1,(2/3),4	二类一致率 (1/2/3),4	跨一档率	跨二档率	跨三档率	标题和题目匹配	标题和正文匹配
作文数	660	852	1126	588	231	21	1329	1052
百分比(%)	44.0%	56.8%	75.1%	39.2%	15.4%	1.4%	88.6%	70.1%

Table 11: BertModel实验结果（取aver）

	A	P	R	F1
离题作文	-	0.68	0.97	0.80
不离题作文	-	0.88	0.32	0.47
总体(macro avg)	0.71	0.78	0.64	0.63
其他题目下作文	-	0.59	0.97	0.73
随机打乱	-	0.36	0.97	0.53
优美散文	-	0.12	0.95	0.22

<sup>a</sup> 下面三个prf值均为离题作文的

Table 12: HBiLstm实验结果（取aver）

	A	P	R	F1
离题作文	-	0.68	0.97	0.80
不离题作文	-	0.88	0.32	0.47
总体(macro avg)	0.71	0.78	0.64	0.63
其他题目下作文	-	0.70	0.96	0.81
随机打乱	-	0.50	0.96	0.66
优美散文	-	0.20	0.98	0.33

<sup>a</sup> 下面三个prf值均为离题作文的

Table 13: HBiLstm\*实验结果（取aver）

	A	P	R	F1
离题作文	-	0.80	0.96	0.87
不离题作文	-	0.91	0.64	0.75
总体(macro avg)	0.83	0.86	0.80	0.81
其他题目下作文	-	0.72	0.96	0.82
随机打乱	-	0.52	0.97	0.67
优美散文	-	0.20	0.97	0.33

<sup>a</sup> 下面三个prf值均为离题作文的