

# 周报，2020年12月28日

屈原斌  
首都师范大学  
ybqu@cnu.edu.cn

## 1 上周计划

1. [\*\*\*] 修改开题报告和ppt。
2. [\*\*\*] 补充实验，分析数据。

## 2 上周计划执行情况

1. [✓×] 继续修改。
2. [✓]

## 3 本周部分重点工作详述

### 3.1 $DataSet_1$ 实验

- 补充方案三（对所有测试数据表示取平均）实验结果，见图1
  - 方案三的结果优于方案一、二
  - HABiLstm.W(+64)结果F1-score最优

|                       |     | 最优Threshold | precision | recall | F1-score |
|-----------------------|-----|-------------|-----------|--------|----------|
| HABiLstm.W            | 离题  |             | 0.1465    | 0.1187 | 0.1311   |
|                       | 不离题 | 41          | 0.9752    | 0.9804 | 0.9778   |
| Bert.CLS(微调9-11层)     | 离题  |             | 0.127     | 0.142  | 0.1341   |
|                       | 不离题 | 14          | 0.9756    | 0.9724 | 0.974    |
| HABiLstm.W(+64)       | 离题  |             | 0.171     | 0.144  | 0.1563   |
|                       | 不离题 | 30          | 0.9759    | 0.9803 | 0.9781   |
| Bert.CLS(微调9-11层,+64) | 离题  |             | 0.1789    | 0.1265 | 0.1482   |
|                       | 不离题 | 7           | 0.9755    | 0.9836 | 0.9795   |

Table 1: 方案三指标更新

### 3.2 实验结果分析

- 模型：HBiLSTM\_W
- 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
- 取Threshold=1的结果进行分析（离题作文Recall最高）

|     | precision | recall | f1-score | support |
|-----|-----------|--------|----------|---------|
| 离题  | 0.0411    | 0.4351 | 0.0751   | 1542    |
| 不离题 | 0.9781    | 0.7126 | 0.8245   | 54493   |

Table 2: Threshold=1时指标

- **题目：**阅读下面的文字，按要求作文。时光飞逝，同学们告别童年，走进少年时代。我们一边感受着青春的喜悦；同时，也体验着成长带来的诸多烦恼和困扰。但也许正是这些，才让你的青春年华更加充满魅力、色彩斑斓。要求：以“成长的烦恼”为题，立意自定，文体不限，写一篇不少于600字的文章。

- **正文：**  
 成长的烦恼  
 双腿瘫痪后，我的脾气变得暴怒无常。望着望着天上北归的雁阵”我会突然把面前的玻璃砸碎”听着听着李谷一甜美的歌声，我会猛地把手边的东西摔向四周的墙壁。母亲就悄悄地躲出去！在我看不见的地方偷偷地听着我的动静。当一切恢复沉寂，她又悄悄地进来，眼边红红的，看着我”听说北海的花儿都开了，我推着你去走走的也总是那么说。母亲喜欢走，可自从我的腿瘫痪后他待弄的那花都死了。“不”我不去坚住哭声说出的狼儿在一块儿，我狠命地捶打这两条腿”喊着“”我可活什么劲上好桑番话，教好儿活““，  
 可我却一直都不知道，也的病已经到了那梦里地。后来妹妹告诉我，我也常常肝疼得整宿整宿翻来覆去地睡不要觉。  
 那天我又独自坐在屋里，看着窗外的树才”唰唰啦啦”地飘落“母亲进来了，挡在窗前以北海的菊花开了一手推着你去看看吧。”她憔悴的脸上现出央求般神色。什么时候人比你要愿意，就明天”她说。我的回答已经让她喜出望外了。么好吧，就明天的我说。她高兴得一会儿坐下！一会儿站起。”那就赶紧准备的”哎呀！烦不烦！几步路，有什么好难站她也笑了、坐在身边，絮絮叨叨说着次看完一菊花，咱们去”仿膳，你小时候最爱吃那儿的豌豆黄儿。还记得那是我带你去上吗？你偏说那撒的花是毛毛只，跑着，一脚踩扁一个“心她忽然不说了。对于“跑”和“踩”“类的字眼儿她比我敏感。她又悄悄地出去了——  
 她出去后，就再也没回来。  
 邻居们把她抬上车时，她还在大口大口地吐着鲜血。没想到她已经病成那样。看着三轮车远去，也绝没有想到那竟是永远的诀别。  
 邻居小伙子背着我去看他的时候她正艰难呼吸着，像被那一手艰难的生活。别人告诉我，她昏迷前的最后一句话的那个有病的儿子和我那个未成年的女儿““，  
 又是秋天妹妹推我去北海看了菊花。黄色的花淡雅，白色的花高洁，紫色的花热烈而深沉、泼泼洒洒，秋风中正开得烂漫山我懂的亲没有说完话。妹妹也懂我们在一块一，好好儿活的小  
 - **真实标签：**完全离题  
 - **预测标签：**不离题  
 - **相似度排名：**1  
 - **分数：**14.0

Figure 1: 案例1（离题作文识别为不离题作文）

- 分析：
  - 离题作文分类错误（共871篇，分析主要哪个离题分档的作文比较难分）
    - \* 统计结果见表2
    - \* 分析案例，见图2
      - 完全离题作文较好识别，部分离题作文较难识别
      - 案例1的作文和当前题目下部分作文较为相似，都是写父母，身边人的事

|          | 完全离题 (score=1) | 部分离题 (score=2) | 部分离题 (score=3) | 总计     |
|----------|----------------|----------------|----------------|--------|
| 分类错误作文数  | 207            | 317            | 347            | 871    |
| 分类错误作文占比 | 38%            | 63.90%         | 69.10%         | 56.50% |
| 总作文数     | 544            | 496            | 502            | 1542   |

Table 3: 离题作文分类错误数据统计

- 不离题作文分类错误（共15662篇，分析主要哪类卷的作文比较难分）
  - \* 统计结果见表2
  - \* 分析案例，见图3
    - 一类卷未召回的比例较大

|          | 四五类卷-[0,60) | 三类卷-[60,80) | 二类卷-[80,90) | 一类卷-[90,100] | 总计     |
|----------|-------------|-------------|-------------|--------------|--------|
| 分类错误作文数  | 174         | 9413        | 4971        | 1104         | 15662  |
| 分类错误作文占比 | 30.30%      | 28.20%      | 28.10%      | 38.10%       | 28.70% |
| 总作文数     | 574         | 33348       | 17679       | 2892         | 54493  |

Table 4: 不离题作文分类错误数据统计

## 4 下周计划

1. [\*\*\*] 修改开题报告。
2. [\*\*\*] 更新传统方法实验。

- 题目：作文题目：你，就这样留在了我的记忆里要求：①写一篇记叙文。②字数在600字左右。③文中不要出现真实的地名、校名和人名等透露自己身份的信息。

- 正文：

在我离年级的时候，我和她路一起去姥姥家。因为那时候我还会骑电动车，所以我就骑着自行车去的，弟总坐在后面“我们俩就这样去姥姥家了”我们骑着自行车，边骑边笑我和弟弟已经好久没这样过了”到细柳家我们才想来我也记着了”幸好老道把我的手机拿来了要不然我们这咋办呀。于是，我的感觉给爸爸打电话，内容是“喂。爸爸我姐姐的钱包忘拿了，你帮我们送过来吧。要不然你们别去了，现在就回来吧。我不！我就要去姥姥家，最终。老爸还是给我们送钱来了、理你们现在在哪儿。z，在小看校门口“你快来。

爸爸到小学校门口时，我便高兴极了。对我们说了这样一句话：如果你们再丢三落四的，就别去了一知道了，你女儿我是谁呀”不会再忘了0

拿来了钱我和弟弟正想着怎么蒸呢“却又突然想来，热些不能吃过于稠越的东西，所以我如来这打某去理个超市买点东西。超市里的东西实在着太多了，我们品了一些果子是这些都是带婆的站的。又给弟弟兵工“复巧克力“

否完东西后，我们就走了”骑着自行车吃着巧克力这种感觉真好呀”才刚出发不到但分钟，弟弟就吵着说我”可我又没有东西给他吃。就在这时弟弟说了一句话”前面的筐子里不有一大包饼干吗，你想干吗。那不是给你买的是给姥姥买的。姐您就给我一个“藏吗，姐，就一个。只有一个行“给你吧。说好了，只有一个我—姐你真小气！

啊！终走到了，可是累死我了“我们在姥姥家吃完饭。玩到下午3点钟这回家的。本来续续最想留我们在她家过的，可是我们的作业还没有写点“所以才在的，要不然我才不走呢

回到家后。爸爸不由分说的的骂了我一顿，然后让我握手机拿出来“我一看我的口袋里没有，弟弟那也没直”这才知道原来中机去了“

紧接着，爸爸从他的感寂“袋里拿出了我的手机。我对爸爸说”我的手机怎么在你这人是母为“你在超市黑天东西，走的时候掉了，这么机的动静。你竟然没有听见前后来还是一位好心人给你送来的，他看到你手机里有德德，所以给我打电话，让我去迈的一

我真想知道他叫什么！可惜我不知道他叫什么“

- 真实标签：不离题

- 预测标签：完全离题

- 相似度排名：11

- 分数：79.2

- 相似度最大题目：【大作文】初中三年，你遇到过很多有才华的老师，遇到过很多有个性的同学，也看过不少的书籍，认识了古今中外的很多名人……当然，通过细致的观察，你也发现了生活中很多只有一面之缘的人却给你留下了深刻的印象；通过用心体会，你对父母也有了全新的认识……请以“他，\_\_\_\_\_”或“她，\_\_\_\_\_”为题，写一篇不少于600字的文章。要求：①将题目填写完整。②自定立意，自选文体（诗歌除外）；③要有真情实感；④不得出现真实的人名、校名和地名。⑤书写规范、标点正确、卷面整洁。

Figure 2: 案例2（不离题作文识别为离题作文）