

# 周报，2021年02月08日

屈原斌  
首都师范大学  
ybqu@cnu.edu.cn

## 1 上周计划

1. [\*\*\*] 实验更新

## 2 上周计划执行情况

1. [✓]

## 3 本周部分重点工作详述

### 3.1 实验更新

#### 3.1.1 中文数据集

- 分类模型更新，见表1
- 结论：
  - xlnet截断512长度后指标下降，但优于BERT

	Accuracy	Precision	Recall	F1-score
HABiLstm	0.7	0.714	0.691	0.694
HABiLstm(加入测试集)	0.686	0.694	0.675	0.678
BERT	0.572	0.588	0.565	0.568
BERT(加入测试集)	0.558	0.584	0.553	0.555
XLNet (截取512字)	0.686	0.706	0.68	0.685
XLNet (截取850字)	0.809	0.818	0.805	0.806

Table 1: 分类指标更新

- seq2seq实验更新
  - 数据集：乐乐课堂数据
  - 数据分布见表2：
  - 指标见表3：

	训练集	验证集	测试集
作文数	68428	8553	8554

Table 2: seq2seq模型数据分布

- 聚类效果，见表4：
  - 数据集：全部不离题作文，共59560篇
  - 方法：K-means聚类，评价指标：purity\_score

rouge_1	0.7379
rouge_2	0.6952
rouge_l	0.7384

Table 3: seq2seq模型指标

– 结论：

- \* 去除题目后获取篇章表示更好
- ? XLNet篇章表示指标最低(待分析)

		未去除题目	去除题目
未添加测试集主题	HABiLstm	0.52	0.579
	BERT	0.48	0.518
	XLnet	-	0.475
添加测试集主题	HABiLstm	0.695	0.742
	BERT	0.566	0.636

Table 4: 篇章表示聚类指标

#### ● 离题实验更新

- 方案一：基于排序方法（针对不同主题动态调节阈值），指标见表5
- 方案二：基于相似度（去除排序，直接根据相似度判断，若与主题相似度小于阈值，则判断为离题，阈值取**0.25**），指标见表6
- 结论：
  - \* 表示更准确的情况下，使用相似度直接判断指标优于排序方法

		离题			不离题		
		Precision	Recall	F1-score	Precision	Recall	F1-score
未添加测试集主题	HABiLstm	0.122	0.0759	0.0936	0.9763	0.9859	0.9811
	BERT	0.1096	0.0415	0.0602	0.9756	0.9913	0.9834
	XLNet	0.1655	0.0305	0.0515	0.9754	0.996	0.9856
测试集加入训练	HABiLstm	0.0968	0.1304	0.1111	0.9936	0.9911	0.9924
	BERT	0.1511	0.0661	0.092	0.9762	0.9904	0.9832

Table 5: 方案一指标（基于排序方法）

### 3.1.2 英文数据集

- 标注数据分布，见表7(共13个主题，830篇作文)
- 离题实验更新：
  - 方案一：基于排序方法（针对不同主题动态调节阈值）（**未完成**）
    - \* 参考题目：筛选其他题目下作文数大于20篇的题目，共70个题目

## 4 下周计划

1. [\*\*\*] 完成英文seq2seq生成模型
2. [\*\*\*] 完善英文数据实验

		离题			不离题		
		Precision	Recall	F1-score	Precision	Recall	F1-score
未添加测试集主题	<b>HABiLstm</b> <b>BERT</b>	0.0574	0.3982	0.1003	0.9816	0.8306	0.8998
		0.0449	0.3145	0.0786	0.0543	0.3145	0.0926
测试集加入训练	<b>HABiLstm</b> <b>BERT</b>	0.0652	0.3632	0.1106	0.9813	0.8652	0.9196
		0.0543	0.3145	0.0926	0.9797	0.8582	0.9149

Table 6: 方案二指标（基于相似度方法）

score	1	1.5	2	2.5	3	3.5	4
作文数	0	0	8	44	105	230	443

Table 7: 英文数据集标注数据分布