

周报，2021年03月08日

屈原斌  
首都师范大学  
ybqu@cnu.edu.cn

## 1 上周计划

1. [\*\*\*] 英文分类实验更新
2. [\*\*\*] 生成实验更新
3. [\*\*\*] 英文离题实验更新

## 2 上周计划执行情况

1. [×] BERT模型未更新（服务器内存不够）
2. [×] BERT生成模型未跑完
2. [✓]

## 3 本周部分重点工作详述

### 3.1 英文数据集

#### 3.1.1 分类实验更新

- 数据集：ICLE数据集，处理数据后重新训练
- 指标更新，见表2
- 实验结论：
  - 加入测试主题后指标下降

	Accuracy	Precision	Recall	F1-score
<b>HABiLstm</b>	0.7139	0.6635	0.6663	0.6382
<b>HABiLstm(加入测试主题)</b>	0.6466	0.6146	0.6121	0.5947
<b>BERT</b>	0.9198	0.9079	0.8954	0.8914
<b>BERT(加入测试主题)</b>	-	-	-	-

Table 1: 英文数据分类指标

#### 3.1.2 生成模型更新

- 数据集：Xsum数据集，数据分布见表2
- 指标更新，见表3

	训练集	验证集	测试集
<b>Xsum</b>	214352	11333	11333

Table 2: 生成模型数据分布

	rouge_1	rouge_2	rouge_l
<b>BiLstm</b>	0.4346	0.2290	0.4022

Table 3: 生成模型指标更新

### 3.1.3 离题实验更新

- 数据集划分：五折交叉验证
- 指标更新：
  - 方案一：基于题目排序方法（共70个参考题目），见表4、5
  - 方案二：基于相似度方法，见表6、7
- 实验结论：
  - 添加测试集主题后指标下降
  - 根据Accuracy调阈值离题指标下降较多
  - 生成模型离题recall提升

			离题			不离题		
			precision	recall	f1-score	precision	recall	f1-score
未添加测试集	habilstm	开发集	0.4699	0.5769	0.5160	0.5432	0.4351	0.4809
		测试集	0.4732	0.5568	0.5111	0.5308	0.4472	0.4849
	bert	开发集	0.4676	0.6286	0.5346	0.5338	0.3737	0.4373
		测试集	0.4665	0.6280	0.5352	0.5340	0.3725	0.4387
添加测试集	habilstm	开发集	0.4981	0.3870	0.4343	0.5542	0.6632	0.6029
		测试集	0.5014	0.3875	0.4371	0.5537	0.6636	0.6036
	bert	开发集	-	-	-	-	-	-
		测试集	-	-	-	-	-	-
doc2vec		开发集	0.4722	0.0758	0.1303	0.5333	0.9226	0.6751
		测试集	0.4626	0.0774	0.1326	0.5333	0.9211	0.6755
seq2seq		开发集	0.4898	0.9287	0.6400	0.6591	0.1446	0.2099
		测试集	0.4687	0.8740	0.6058	0.5800	0.1343	0.1837

Table 4: 方案一指标更新（根据离题**F1-score**调阈值）

### 3.1.4 生成模型生成效果分析

- BiLstm生成模型
  - 不同长度文本生成指标见表8
  - Attention可视化例子见图1
  - 结论：
    - \* 模型对长文本生成效果较差
    - \* 模型主要关注到第一句话（提取关键词主要集中在第一句话）

## 4 下周计划

1. [\*\*\*] 完成Bert生成模型
2. [\*\*\*] 完成层次Lstm生成模型
3. [\*\*\*] 更新/分析离题实验结果

Accuracy				离题			不离题		
				precision	recall	f1-score	precision	recall	f1-score
未添加测试集	habilstm	开发集	0.5506	0.6955	0.1308	0.2006	0.5446	0.9100	0.6806
		测试集	0.5349	0.5404	0.1193	0.1803	0.5391	0.8999	0.6727
	bert	开发集	0.5542	0.8030	0.1592	0.1885	0.5407	0.8742	0.6582
		测试集	0.5361	0.7138	0.1690	0.1833	0.5428	0.8634	0.6469
添加测试集	habilstm	开发集	0.5614	0.6312	0.1664	0.2267	0.5525	0.8911	0.6797
		测试集	0.5395	0.6066	0.1451	0.1979	0.5433	0.8868	0.6696
	bert	开发集	-	-	-	-	-	-	-
		测试集	-	-	-	-	-	-	-
doc2vec		开发集	0.5494	0.8372	0.0441	0.0807	0.5421	0.9879	0.6997
		测试集	0.5389	0.7366	0.0371	0.0684	0.5375	0.9779	0.6934
seq2seq		开发集	0.5651	0.5470	0.4616	0.4309	0.6395	0.6256	0.5455
		测试集	0.4961	0.4522	0.4164	0.3696	0.5320	0.5741	0.4861

Table 5: 方案一指标更新（根据Accuracy调阈值）

				离题			不离题		
				precision	recall	f1-score	precision	recall	f1-score
未添加测试集	habilstm	开发集	0.4845	0.9528	0.6408	0.7981	0.1103	0.1746	
		测试集	0.4782	0.9329	0.6309	0.5374	0.1073	0.1648	
	bert	开发集	0.4722	0.9803	0.6367	0.1462	0.0400	0.0628	
		测试集	0.4671	0.9677	0.6291	0.1089	0.0351	0.0530	
添加测试集	habilstm	开发集	0.4775	0.9719	0.6392	0.6391	0.0698	0.1200	
		测试集	0.4724	0.9588	0.6326	0.7155	0.0637	0.1079	
	bert	开发集	-	-	-	-	-	-	
		测试集	-	-	-	-	-	-	
doc2vec		开发集	0.4697	0.9892	0.6363	0.4857	0.0243	0.0454	
		测试集	0.4683	0.9883	0.6354	0.5000	0.0197	0.0375	
seq2seq		开发集	0.4663	1.0000	0.6352	0.0000	0.0000	0.0000	
		测试集	0.4663	1.0000	0.6359	0.0000	0.0000	0.0000	

Table 6: 方案二指标更新（根据离题F1-score调阈值）

Accuracy				离题			不离题		
				precision	recall	f1-score	precision	recall	f1-score
未添加测试集	habilstm	开发集	0.5855	0.6516	0.4381	0.4504	0.5955	0.6914	0.6275
		测试集	0.5449	0.5635	0.3964	0.4030	0.5689	0.6795	0.6047
	bert	开发集	0.5747	0.4841	0.4578	0.4233	0.5719	0.7348	0.6268
		测试集	0.5241	0.5926	0.2641	0.2977	0.5508	0.7786	0.6283
添加测试集	habilstm	开发集	0.5892	0.6068	0.3337	0.3931	0.5900	0.8058	0.6739
		测试集	0.5503	0.5557	0.3038	0.3558	0.5601	0.7660	0.6395
	bert	开发集	-	-	-	-	-	-	-
		测试集	-	-	-	-	-	-	-
doc2vec		开发集	0.5928	0.6856	0.3456	0.4125	0.5890	0.8076	0.6769
		测试集	0.5792	0.6018	0.3331	0.3910	0.5811	0.7948	0.6670
seq2seq		开发集	0.5530	0.4168	0.2786	0.2361	0.4486	0.7511	0.5588
		测试集	0.5051	0.2654	0.1535	0.1695	0.5109	0.8174	0.6202

Table 7: 方案二指标更新（根据Accuracy调阈值）

	作文数	rouge.1	rouge.2	rouge.l
长度≤400	6326	0.4593	0.2471	0.4231
长度>400	4994	0.4033	0.206	0.3758

Table 8: BiLstm模型不同文本长度指标



Figure 1: 生成模型Attention例子