

2021年03月07日进度汇报

屈原斌
首都师范大学
ybqu@cnu.edu.cn

1 今日进度

1. 完成英文BERT分类实验更新
2. 对比BERT取平均/CLS表示的离题指标
3. 中文BERT生成模型实验（中英文的都未跑完）
4. 好未来AI接口调用

2 结论

- 英文BERT分类模型指标更新，见Table 1
- 离题实验：
 - 方案一：基于题目排序方法，指标见Table2、3
 - 方案一：基于相似度方法，指标见Table4、5
- 结论：
 - 加入测试集后离题指标下降
 - 方案一中表示取平均方法离题指标优于CLS表示，方案二CLS表示更优

	Accuracy	Precision	Recall	F1-score
BERT	0.9198	0.9079	0.8954	0.8914
BERT(加入测试主题)	0.8805	0.8549	0.8414	0.8344

Table 1: Bert分类模型指标

			离题			不离题		
			precision	recall	f1-score	precision	recall	f1-score
未加入测试集	bert(取CLS)	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
	bert(取平均值)	开发集	0.4666	0.5431	0.4995	0.5327	0.4562	0.4887
		测试集	0.4647	0.5427	0.5005	0.5318	0.4539	0.4896
加入测试集	bert(取CLS)	开发集	0.4693	0.1325	0.2046	0.5346	0.8692	0.6617
		测试集	0.4765	0.1367	0.2123	0.5354	0.8691	0.6626
	bert(取平均值)	开发集	0.4670	0.1131	0.1811	0.5347	0.8895	0.6675
		测试集	0.4778	0.1161	0.1867	0.5353	0.8894	0.6683

Table 2: 方案一指标更新（离题F1-score调阈值）

Accuracy				离题			不离题		
				precision	recall	f1-score	precision	recall	f1-score
未加入测试集	bert(取CLS)	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
	bert(取平均值)	开发集	0.5554	0.8024	0.1400	0.1723	0.5455	0.8970	0.6723
		测试集	0.5331	0.7021	0.1446	0.1648	0.5387	0.8782	0.6545
加入测试集	bert(取CLS)	开发集	0.5627	0.7899	0.0996	0.1662	0.5509	0.9631	0.7004
		测试集	0.5328	0.5998	0.0771	0.1270	0.5358	0.9315	0.6796
	bert(取平均值)	开发集	0.5530	0.8156	0.0643	0.1115	0.5446	0.9762	0.6988
		测试集	0.5386	0.6359	0.0552	0.0977	0.5379	0.9614	0.6895

Table 3: 方案一指标更新（Accuracy调阈值）

				离题			不离题		
				precision	recall	f1-score	precision	recall	f1-score
未加入测试集	bert(取CLS)	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	
	bert(取平均值)	开发集	0.4906	0.9408	0.6388	0.1440	0.1137	0.1271	
		测试集	0.4731	0.9266	0.6200	0.1147	0.0897	0.1006	
加入测试集	bert(取CLS)	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	
	bert(取平均值)	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 4: 方案二指标更新（离题F1-score调阈值）

Table 5: Add caption

Accuracy				离题			不离题		
				precision	recall	f1-score	precision	recall	f1-score
未加入测试集	bert(取CLS)	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
	bert(取平均值)	开发集	0.5855	0.4970	0.3595	0.3861	0.5921	0.7737	0.6551
		测试集	0.5452	0.4279	0.3398	0.3452	0.5668	0.7275	0.6149
加入测试集	bert(取CLS)	开发集	0.5723	0.8515	0.1219	0.1890	0.5565	0.9563	0.7033
		测试集	0.5407	0.6537	0.0880	0.1429	0.5405	0.9375	0.6847
	bert(取平均值)	开发集	0.5735	0.7761	0.1309	0.2091	0.5575	0.9524	0.7032
		测试集	0.5443	0.6659	0.0910	0.1508	0.5425	0.9412	0.6876

Table 6: 方案二指标更新（Accuracy调阈值）