

周报

屈原斌

2020 年 9 月 14 日

1 上周任务

- BERT分类任务；
- 初中作文数据处理；

2 任务详情

2.1 BERT分类实验

- BERT分类实验训练数据量太大了，目前还是没有实验结果，验证集上目前的准确率达到0.828；

2.2 初中作文数据处理

- 分数信息统计：
 - 分数为百分制，主要统计了[0,60), [60,70), [70, 80), [80, 90), [90, 100]分数段的作文数，具体数据如表一，主要集中在70-90分之间；
 - 同时也统计了60分以下的作文数，具体数据如表二，0分作文占比10%左右；
 - 分数信息比作文信息数量多，目前还没有具体进行筛选；

表 1: 各分数段作文数统计

分数段	[0, 60)	[60, 70)	[70, 80)	[80, 90)	[90, 100]	总计
作文数	89438	179491	514334	357603	42731	1183497

表 2: 0-60分数段作文数统计

分数段	0	(0, 10)	[10, 20)	[20, 30)	[30, 40)	[40, 50)	[50, 60)	总计
作文数	9068	2950	4590	5006	8379	15524	43921	89438

3 本周计划

- 把BERT部分的实验跑完；
- 处理数据；