

周报

屈原斌

2020 年 10 月 26 日

1 上周任务

- 确定数据标注工作；
- 看论文；

2 任务详情

2.1 数据标注

- 标注规范添加题目和题目要求的标签，对不符合要求的题目进行特殊标注；
- 先根据目前的审题结果抽取1个题目共50篇作文进行试标，后续根据试标结果再进一步确定标注工作；

3 本周计划

- 数据集构建：
 - 构建两套数据集：
 - * 抽取数据进行人工标注，构建比较合理的数据集，用于真实离题检测；
 - * 随机构建，抽取其他题目作文添加到当前题目构建负样本，用于防攻击、恶意提交；

- 抽取部分题目下的作文进行离题的分析：
 - 确定离题的准确定义；
 - 分析作文为什么离题?哪部分离题?生成总结文档；
- 确定后续具体方案和评价指标，确定预期的结果；