

周报

屈原斌

2020 年 8 月 3 日

1 上周任务

- 字向量模型训练;
- 离题检测 - 相似度计算;

1.1 字向量模型

- 使用word2vec训练, 词表大小: 6646;

1.2 相似度计算

- 目前的工作: 在每套试题中抽取100份作文, 根据分类模型的隐藏层输出分别对每套题下的100份作文(来自于同一个题目)之间计算相似度, 得到每套试题的一个平均相似度;
- 实验结果: 目前的计算结果不是很理想, 最大的值才达到0.49;
- 下一步工作: 计算每套试题之间的平均相似度, 对比单套题目的平均相似度, 观察最终结果是否有区分度;

2 本周计划

- 使用字向量做一下分类实验;
- 完成相似度计算后续实验内容;