

2021年06月30日进度汇报

屈原斌
首都师范大学
ybqqu@cnu.edu.cn

1 今日进度

1. 标题-正文匹配实验

2 工作详述

2.1 标题-正文匹配实验

- 数据集构造

- 数据来源: 乐乐课堂 + 笔神作文
- 构建方法, 见图1
- 数据分布, 见表1

- 实验方案:

- BERT-NSP分类
- Attention Sum Reader分类
 - * 字向量替换为词向量

- 指标: (未完成)

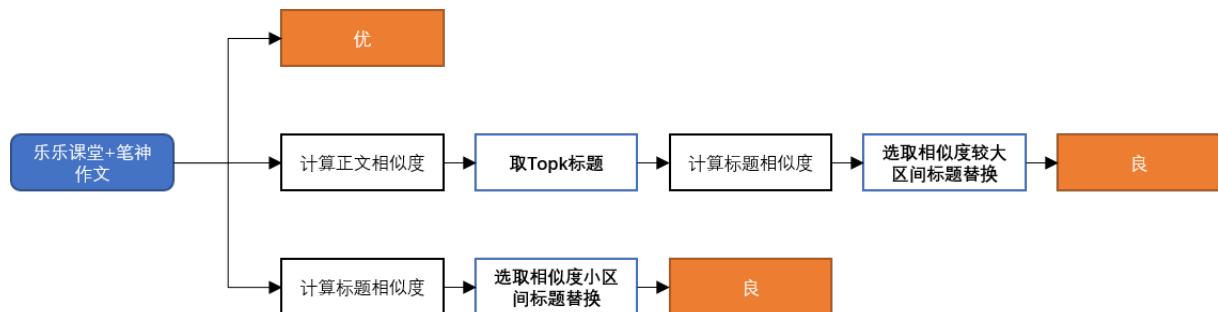


Figure 1: 数据集构造方法

	优	良	中	合计
训练集	10w 20w	5w 10w	5w 10w	20w 40w
验证集	500	50	50	600
测试集	2000	200	200	2400

Table 1: 构建数据分布