

第二次作业

数据说明及提交要求

数据说明

- Train/test_news.tsv
 - Nid: 新闻id
 - Category: 新闻类别
 - SubCategory: 新闻子类别
 - Title: 新闻标题 (可为空)
 - Abstract: 新闻摘要 (可为空)

数据说明

- Train/test.tsv
 - Uid: 用户id
 - Date : 时间
 - History: 用户点击历史
 - Impression : 当前展示页面所包含的新闻, 以及用户的点击情况。在train和test中会有所区别。
 - 下图为train的Impression示例, 表示页面包含3条新闻, 用户点击了N49685

N49685-1 N28047-0 N29128-0

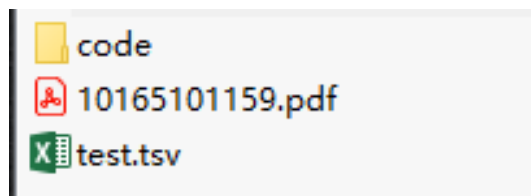
- 下图为test的Impression示例, 表示页面包含3条新闻。任务是预测用户会点击其中哪几条 (一条及以上) 。

N49685 N28047 N29128

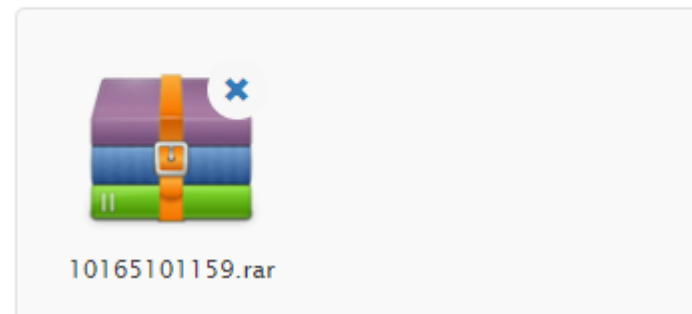
提交要求

- 提交文件

- 代码
- 报告
- 结果文件
- 按图示命名
- 的方法只能有一个test.tsv。如有多种方法请选择你认为最好，其余的在报告中说明即可。
- 压缩包用学号命名。



* 提交人姓名: 姜徐翔



- 预测结果在所给的test.tsv的最后插入新的列，列名为'Predict'，数据格式为字符串。如下图所示，假设预测结果为用户点击了第一个新闻，后两个都没点击，则test'的数值为'100'。