

活动首页

答题与数据

提交记录

排行榜

招商银行FinTech精英训练营线上复赛

财经新闻分析

返回全部题目列表

赛题背景

财经新闻作为重要却海量的投资数据，无时无刻不在影响着投资者们的投资决策，为了更好地提示客户当下新闻事件对应的投资机会和投资风险，本课以研发“历史事件连连看”为目的，旨在根据当前新闻内容从历史事件中搜索出相似新闻报道，后期可以结合事件与行情，辅助客户采取相应投资策略。

该赛题是让参赛者为每一条测试集数据寻找其最相似的TOP 20条新闻（不包含测试新闻本身），我们会根据参赛者提交的结果和实际的数据进行对比，采用mAP值作为评价指标。

评价指标

该赛题是让参赛者为每一条测试集数据寻找其最相似的TOP 20条新闻，我们会根据参赛者提交的结果和实际的数据进行对比，采用mAP值作为评价指标，评分公式如下：

$$mAP = \frac{1}{D} \sum_{d=1}^D \frac{1}{|Y_d|} \sum_{t \in Y_d} f_t$$
$$f_t = \begin{cases} \frac{|\{t' \in Y_d | r_{t'} \leq r_t\}|}{r_t} & , \quad \text{if } t \in Z_d \\ 0 & , \quad \text{otherwise} \end{cases}$$

其中D表示测试集中新闻的总数量，Y_d表示新闻d的n条真实相似新闻集合，无序，Y_d={Y_{d1},Y_{d2},Y_{d3},.....,Y_{dn} }；Z_d表示选手提交m条(赛题中m=20)相似新闻的有序集合Z_d={Z_{d1},Z_{d2},Z_{d3},.....,Z_{dm} }；Z_d中各元素的Rank值分别1,2,3,.....,m,记为r_i。对于集合K，|K|表示K中元素的个数,即|Y_d |=n。

数据说明

训练集数据train_data.csv：

id 训练集的新闻编号

title 训练集新闻（标题）

测试集数据test_data.csv：

id 测试集的新闻编号

title 测试集新闻（标题）

提交格式

1.采用UTF-8无BOM编码的txt文件提交。

2.提交的文件中至少包含两列，source_id和target_id，标注在文件对应列的第一行（如果有其他列也需标注列的名称）。

3.每行包含一条数据，数据用制表符“\t”隔开。

4.必须按照给定的验证数据集进行预测，不能有遗漏的数据或多出的数据。

格式：

source_id target_id similarity(可选) source_title(可选) target_title(可选)

新闻1 新闻x1

新闻1 新闻x2

...

↑


👁

👤

💬

✍

📱



扫一扫下载牛客APP

新闻1 新闻x20
新闻2 新闻y1
...

特别提醒 本赛题数据仅允许用于本次竞赛，禁止参赛者用于其他用途，禁止使用外部数据。

郑重声明 我们发现有的同学共享答案或共享数据，触发了我们的报警，这是不被允许的。请各位同学遵守复赛规则，一定要下载自己的数据，运行自己的算法，按要求提交自己的结果。复赛试题有一定难度，但相信对于优秀的你们来说，研究一段时间后一定能交出一份满意的答卷。

文件	文件格式	
test_data.csv	csv(3.801KB)	高速下载 普通下载
train_data.csv	csv(99.612MB)	高速下载 普通下载
sample.txt	txt(11.43KB)	高速下载 普通下载

上传预测结果数据，格式仅限txt文件，大小不得超过10M

运行

牛客网，程序员必备求职神器

QQ群：169195721

微 信：www_nowcoder_com

微 博：牛客网

关注

关注



扫一扫，把题目装进口袋

[关于我们](#) [加入我们](#) [意见反馈](#) [企业服务](#)
[网站合作](#) [免责声明](#) [友情链接](#)

京ICP备14055008号-4

京公网安备 11011402010078号

牛客网©2017 All Rights Reserved