财经新闻分析

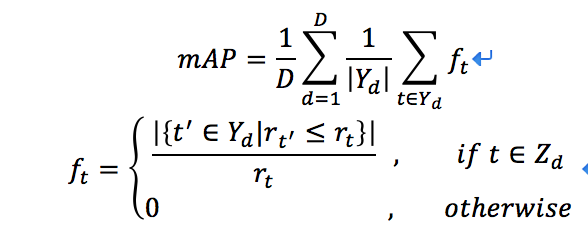
**赛题背景**

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，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：



其中D表示测试集中新闻的总数量，Yd表示新闻d的n条真实相似新闻集合，无序，Yd={Yd1,Yd2,Yd3,……,Ydn }；Zd表示选手提交m条(赛题中m=20）相似新闻的有序集合Zd={Zd1,Zd2,Zd3,……,Zdm }；Zd中各元素的Rank值分别1,2,3,……,m,记为ri。对于集合K，|K|表示K中元素的个数,即|Yd |=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. 文件格式txt。
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... 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。