

赛题介绍



赛题背景

在网络媒体飞速发展的今天，新闻数量爆炸式的增长，各类新闻应用每天都会产生各式各样的内容。但是对于读者来说，阅读过量而无序的新闻信息反而容易成为一种负担，因此**对新生成的新闻进行分类**是一项十分关键的技术。

比如在爱城市网APP的资讯板块中，每天都会更新很多的新闻，但是为了更好的满足用户需求，我们需要使用算法对新闻进行分类和推荐，在APP有限的版面中尽量能够使用户第一眼就得到他们最需要的信息。

赛题背景

文本分析建模能力

内容分类算法建模能力

提高对文本分析领域的认识和应用能力

激发算法设计和机器学习等领域的兴趣。

题目描述

本题中，参赛者需要**建立自己的新闻数据库**，设计实现基于该数据库的**新闻内容分类和推荐算法**，并做相应的前端效果展示。

为了帮助大家收集数据，我们将提供**20000条**基础新闻数据，包含的内容：新闻标题、新闻内容、点击次数、所有评论(可选)、发布时间。

这些数据由浪潮集团旗下“爱城市网”提供，参赛者也可以使用其他新闻网站的数据，比如编写网络爬虫在各大新闻网站抓取自己的新闻数据，但抓取的数据中也应包含以上信息（点击次数除外）。

详情

● 数据格式(excel)

- 新闻标题
- 新闻内容
- 点击量
- 评论详情(可选)
- 发布时间

[illegible]

详情

- 参赛者提交内容
 - 6月1日前提交至qinkangwei@inspur.com
 - 标题为 “山大实习挑战赛作品+队名”
 - 材料应体现队伍信息（包括但不限于成员姓名等）
 - 完成报告(pdf)
 - 分类、推荐算法的设计与实现（核心模块、伪代码）
 - 系统架构（前后端及相关技术）
 - 效果展示
 - 提交代码
 - 工程代码，如果工程特别大，可以只提供核心代码
- 评委评审后，得分较高的队伍进入复赛。

评分标准

- 完成新闻内容分类或者聚类算法（30分）
- 完成新闻推荐算法（30分）
 - 新闻推荐给用户或者特定人群
 - 参赛者可以假定用户画像标签集合已知（比如张三的用户画像标签为：济南、学生、山大、美术、体育等）
 - 给定一组标签后，系统可以搜索相应主题的新闻，并给出排名最高的若干新闻进行推荐。
- 前端的设计及表达效果（20分）
- 后端的设计及计算性能（20分）

注意事项

- 请尽量不要使用第三方算法API(比如ai.baidu.com)，如果使用第三方应用，算法部分将不会得高分。如果必须要使用的话，请注明相关的第三方应用，并将系统架构及页面展示等工程部分尽量做好以提高整体分数；
- 由于比赛时间较短，任务总量较重，请同学们组队完成；
- 前端设计优异，也会有一定的加分。

谢谢大家！



inspur 浪潮