赛题介绍



INSPUC浪潮

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
 - 量击点•
 - 评论详情(可选)
 - 发布时间

		_		
新闻标题	新闻内容	点击量	发布时间	
土忠州孙还涛驹听省意观巾建设工作	5月18日上午,自变吊变、巾变书记土忠州和巾3	2686	2018/5/18	
CCTV2中国财经报道:智能科技改变生活爱城市网APP上线	在第二届世界智能大会上,"爱城市网"APP以其的	475	2018/5/18	
第二届世界智能大会主论坛嘉宾发言摘编:人工智能引领产业转型升级	第二届世界智能大会 br/> <td>612</td> <td>2018/5/18</td>	612	2018/5/18	
CCTV2中国财经报道:智能科技改变生活爱城市网APP上线	在第二届世界智能大会上,"爱城市网"APP以其情	1406	2018/5/18	
"天津爱城市网"亮相世界智能大会,一站式城市服务受关注	5月16日至19日,第二届世界智能大会在天津召开	769	2018/5/18	
济南市发布大雾橙色预警信号,出行注意安全	济南市气象局2018年05月17日06时00分发布 <stro< td=""><td>5418</td><td>2018/5/18</td></stro<>	5418	2018/5/18	
助力"政务一网通"改革,天津爱城市网正式上线发布	为深入贯彻落实习近平总书记在全国网络	1170	2018/5/18	
暴雨马上来!济南发布强降雨防范紧急通知!还有冰雹+雷电+9级大风	<div style="text-align:center;"><span style="font-s</td><td>17256</td><td>2018/5/17</td></tr><tr><td>相约智能科技盛宴,请收下这份玩转浪潮展区攻略</td><td><div style=" text-align:center;"=""><span style="font-s</td><td>883</td><td>2018/5/17</td></tr><tr><td>市民" td="" 随手拍",污染无处藏!"爱城市网"app上线"污染随手拍"应用<=""><td>拿起相机,你就是拍客;关爱家园,你就是环保</td><td>2445</td><td>2018/5/17</td></div>	拿起相机,你就是拍客;关爱家园,你就是环保	2445	2018/5/17
"爱城市网"APP上线"污染随手拍"应用	为落实好《济南市土壤污染防治工作方案》中"你	735	2018/5/17	
感恩母亲节 这个母亲节,教会妈妈用智能手机吧	你可曾记得——小的时候她总是将最好吃的、最	4908	2018/5/17	
【免费送电影票啦】这个母亲节带妈妈一起看电影	中国人含蓄的传统,让我们不太轻易表达自己,	9355	2018/5/17	
有奖互动 你和最好的朋友认识多久了?	爱城市网的小伙伴们,又到了我们每日互动的	5192	2018/5/17	
"中国数据开放第一人"孙丕恕,为何对这个便民应用"情有独钟"?	5月7日下午,一如其他日子一样,爱城市网运营	10613	2018/5/17	
有奖互动 你有没有小时候打死都不吃,长大后却很喜欢吃的食物?	<div style="text-align:center;">爱城市网的小伙伴</div>	5301	2018/5/16	
有奖互动 你等公交车最久的一次,等了多长时间?	各位小伙伴大家好~又到了每日有奖互动的轻松的	5531	2018/5/16	
保护土壤,爱城市网邀您"污染随手拍"	看似平凡的土壤,是我们生命赖以生存的基础。	794	2018/5/16	
只做了一件事,远离50%的交通事故,@驾驶人一定要看!	追尾事故是最常见的事故形态之一,轻则掉点车	1676	2018/5/16	
山东医养结合出新规!医疗、养老"跨界"无障碍	长期以来,"养老的不看病,看病的不养老"是很		2018/5/16	
假期第三天!来了解最新高速出行信息	截止05月01日07时00分: %nbsp;	510	2018/5/16	
五一假期首日高速路况全攻略,肯定有你想看的	轰轰烈烈的五一假期开始啦!外出游玩、回家党			
2010年工	工作拥宣清股小安东泽仁的。	7534	2010/5/16	
→ 基础表 评论表 🕂 🕀	: 4			

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