

年广东省茂名市的 OTA 和 UGC 数据，期待参赛者采用自然语言处理等数据挖掘方法通过建立本地旅游图谱的方式来分析新冠疫情时期该市周边游的发展。

二、解决问题

1、微信公众号文章分类

构建文本分类模型，对附件 1 提供的微信公众号的推送文章根据其内容与文旅的相关性分为“相关”和“不相关”两类，并将分类结果以表 1 的形式保存为文件“result1.csv”。与文旅相关性较强的主题有旅游、活动、节庆、特产、交通、酒店、景区、景点、文创、文化、乡村旅游、民宿、假日、假期、游客、采摘、赏花、春游、踏青、康养、公园、滨海游、度假、农家乐、剧本杀、旅行、徒步、工业旅游、线路、自驾游、团队游、攻略、游记、包车、玻璃栈道、游艇、高尔夫、温泉等等。

表 1 微信公众号文章分类

文章 ID	分类标签
0001	相关
.....

2、周边游产品热度分析

从附件提供的 OTA、UGC 数据中提取包括景区、酒店、网红景点、民宿、特色餐饮、乡村旅游、文创等旅游产品的实例和其他有用信息，将提取出的旅游产品和所依托的语料以的形式保存为文件“result2-1.csv”。建立旅游产品的多维度热度评价模型，对提取出的旅游产品按年度进行热度分析，并排名。将结果以表 3 的形式保存为文件“result2-2.csv”。

表 2 旅游产品提取表

语料 ID	产品 ID	产品名称
酒店评论-0088	ID1	黄妈妈私房菜
.....

表 3 旅游产品的热度

产品 ID	产品类型	产品名称	产品热度	年份
ID1	特色餐饮	黄妈妈私房菜	0.86	2020
.....

3、本地旅游图谱构建与分析

依据提供的 OTA、UGC 数据，对问题 2 中提取出的旅游产品进行关联分析，找出以景区、酒店、餐饮等为核心的强关联模式，结果以表 的形式保存为文件“result3.csv”。在此

基础上构建本地旅游图谱并选择合适方法进行可视化分析。鼓励参赛队挖掘旅游产品间隐含的关联模式并进行解释。

表 4 旅游产品间的关联

产品 1ID	产品 2ID	关联度	关联类型(可选)
ID1	ID5	0.76	景区-餐饮
.....	

4、疫情前后旅游产品需求的变化分析

基于历史数据，使用本地旅游图谱作为分析工具，分析新冠疫情前后茂名市旅游产品的变化，并撰写一封不超过 2 页的信件向该地区旅游主管部门提出旅游行业发展的政策建议。

三、附件说明

附件 1 数据来源于互联网公开渠道，具体表结构字段如下：

表 5 游记攻略

游记 ID	城市	游记标题	发布时间	出行天数	人物	人均费用	正文
-------	----	------	------	------	----	------	----

表 6 酒店评论

酒店评论 ID	城市	酒店名称	评论内容	入住日期	入住房型
---------	----	------	------	------	------

表 7 景区评论

景区评论 ID	城市	景区名称	评论内容
---------	----	------	------

表 8 餐饮评论

餐饮评论 ID	城市	餐饮名称	评论等级	评论内容
---------	----	------	------	------

表 9 微信公众号新闻

文章 ID	文章标题	公众号文章内容
-------	------	---------