

第十届“泰迪杯” 数据挖掘挑战赛

优秀 作品

作品名称：疫情背景下的周边游需求图谱分析

荣获奖项：特等奖并获泰迪杯

作品单位：河海大学

作品成员：陈挺 徐涯昕 朱燕燕

指导老师：施国良

封面为后期添加，原作品没有此页。

疫情背景下的周边游需求图谱分析

摘要

互联网时代下，信息爆炸是我们面临的机遇与挑战，用户生成内容（UGC）和在线旅游（OTA）数据是反映旅游市场发展现状的重要信息来源，尤其是用户评论和游记攻略数据具有一手性，能够最真实地还原用户的旅行体验。但与此同时，相关数据较为零散和碎片化，如何从中获取旅游相关要素并挖掘要素间的关联成为一种迫切的需求。因此，本研究围绕用户生成内容和在线旅游数据，运用自然语言处理技术进行数据分析挖掘，并结合 Neo4j 图数据库和知识图谱的概念，对旅游要素间的隐含关联模式进行进一步的挖掘和呈现。主要研究内容和成果包括以下三部分：

一、微信公众号文章分类：首先基于赛题所给的文旅相关主题词筛选并进行近义词扩充，形成主题词库；然后，采用 TextRank、TF-IDF 分别提取语料关键词，与主题词库进行匹配，同时建立一定规则对标题和正文进行直接匹配，共得到 3 个标签，3 个标签一致的作为伪标注数据用于训练模型；最后，基于中文 Bert 预训练语言模型训练分类器，作为以上三种方法得到结果不一致时的判别依据；达到了较好的公众号文章分类效果。

二、旅游产品提取和热度分析：首先，基于 CLUENER 细粒度命名实体识别语料库筛选相关实体，采用 BiLSTM-CRF 训练模型，针对游记攻略、景区评论、酒店评论和餐饮评论 4 张表进行命名实体识别（NER），并结合 LAC 库设计了一套针对旅游产品的命名实体识别筛选和优化方法；其次，基于 TextCNN 分类模型、Single-Pass 聚类算法和地理编码进行同义地名合并，共整合提取到 631 个有效实体，累计 1014 种表述方式；然后，建立包括参与热度、反响热度和宣传热度三大方面指标的热度评价体系，并进行指标计算；最后，采用 AHP-TOPSIS 方法进行指标权重计算和排序，得到最终热度评价结果。

三、产品关联挖掘和本地旅游图谱构建：首先，在前一阶段的数据准备和实体抽取后进行关联模式的定义，根据本地经济发展程度、地理位置等因素定义了八种关联模式，其中包含三种隐含关系的模式；然后，对每种关联模式进行量化计算和标准化处理，将旅游产品和关联存入 Neo4j 图数据库中，生成疫情前后茂名市旅游产品关联图谱和本地旅游宏观概念图谱；基于产品关联模式、目的地热

度和评论热词变化进行对比分析，发现游客的旅游倾向从“久游饱览式”过渡到“深挖全面式”、餐饮业迎来较好发展趋势和营销风口、旅游产品的共现关系愈加明显；最后，基于以上发现撰写信件向茂名市旅游主管部门提出茂名旅游业发展政策建议。

关键词：旅游图谱；命名实体识别；热度评价；关联挖掘

“泰迪杯”挑战赛优秀作品

Abstract

In the Internet era, information explosion is an opportunity and challenge we are facing. User generated content (UGC) and online travel (OTA) data are important sources of information to reflect the development status of the tourism market. In particular, user comments and travel strategy data have one chirality, which can truly restore the user's travel experience. But at the same time, the relevant data are scattered and fragmented. How to obtain tourism related elements and mine the correlation between elements has become an urgent need. Therefore, this study focuses on user generated content and online tourism data, uses natural language processing technology for data analysis and mining, and combines the concepts of neo4j map database and knowledge map to further mine and present the implicit association patterns between tourism elements. The main research contents and achievements include the following three parts:

1. Wechat official account article classification: first, select and expand the synonyms based on the theme words related to culture and tourism given by the competition questions to form a theme thesaurus; Then, TextRank and TF-IDF are used to extract the corpus keywords and match them with the thesaurus. At the same time, certain rules are established to directly match the title and text. A total of three tags are obtained. The consistent three tags are used as pseudo tagging data for training the model; Finally, the classifier is trained based on the Chinese Bert pre training language model as the discrimination basis when the results obtained by the above three methods are inconsistent; It achieves a better classification effect of official account articles.

2. Tourism product extraction and popularity analysis: firstly, relevant entities are screened based on the cluener fine-grained Named Entity Recognition Corpus. Using the BiLSTM-CRF training model, named entity recognition (NER) is carried out for four tables: travel strategy, scenic spot comments, Hotel Comments and catering comments. Combined with the lac database, a set of named entity recognition screening and optimization method for tourism products is designed; Secondly,

synonymous place names are merged based on TextCNN classification model, single pass clustering algorithm and geocoding. A total of 631 effective entities are integrated and extracted, with a total of 1014 expressions; Then, establish a heat evaluation system including three indexes: participation heat, response heat and publicity heat, and calculate the indexes; Finally, AHP-TOPSIS method is used to calculate and sort the index weight, and the final heat evaluation result is obtained.

3. Product association mining and local tourism map construction: firstly, after the data preparation and entity extraction in the previous stage, the association mode is defined. According to the local economic development degree, geographical location and other factors, eight Association modes are defined, including three implicit relationship modes; Then, the quantitative calculation and standardized processing were carried out for each association model, and the tourism products and associations were stored in the neo4j map database to generate the association map of tourism products in Maoming City and the macro concept map of local tourism before and after the epidemic; Based on the comparative analysis of product correlation mode, target geothermal degree and the change of comment hot words, it is found that the tourism tendency of tourists has changed from "long-term sightseeing" to "deep and comprehensive". The catering industry has ushered in a better development trend, and the relationship between marketing outlets and tourism products is becoming more and more obvious; Finally, based on the above findings, write a letter to put forward policy suggestions for the development of Maoming tourism to the competent Tourism Department of Maoming City.

Key words: tourism map; Named entity identification; Heat evaluation; Association mining

目录

第一章 绪论.....	6
1.1 背景与意义.....	6
1.2 工作及思路.....	6
第二章 针对问题一的解决方案.....	7
2.1 问题分析及思路.....	7
2.2 文本分类全流程.....	8
2.2.1 文旅词库扩充和数据预处理.....	8
2.2.2 TextRank、TF-IDF 关键词提取与判别.....	8
2.2.3 标题、正文规则判别.....	9
2.2.4 模型训练数据准备.....	9
2.2.5 Bert 分类模型训练.....	9
第三章 针对问题二的解决方案.....	10
3.1 问题分析及思路.....	10
3.2 旅游产品提取流程.....	10
3.2.1 数据清洗.....	10
3.2.2 数据准备.....	11
3.2.3 基于 BiLSTM-CRF 和 LAC 的命名实体识别.....	11
3.2.4 基于 TextCNN 的命名实体分类.....	12
3.2.5 基于 Single-Pass 算法的命名实体聚类.....	14
3.2.6 同义实体合并和无效实体剔除.....	15
3.2.7 实体编码与回溯.....	16
3.3 基于 AHP-TOPSIS 的多维度热度评价模型.....	16
3.3.1 热度评价体系构建.....	16
3.3.2 AHP 确定指标权重.....	17
3.3.3 TOPSIS 排序过程.....	18
3.4 可视化分析.....	20
第四章 针对问题三的解决方案.....	20
4.1 问题分析及思路.....	21
4.2 本地旅游产品关联图谱构建及分析.....	21
4.2.1 关联模式定义.....	21
4.2.2 针对隐含关系的抓取.....	22
4.3.4 关联度的计算和量化.....	24
4.3.5 关联图谱存储.....	25
4.3.6 关联图谱分析.....	25
4.3 本地旅游产品宏观图谱构建及分析.....	26
第五章 针对问题四的解决方案.....	27
5.1 问题分析及思路.....	27
5.2 旅游图谱对比分析.....	28
5.3 建议信.....	30

第一章 绪论

1.1 背景与意义

在国民经济全面复苏的背景下，居民旅游需求释放。中国宏观经济在后疫情时代率先复苏，2020年GDP增长2.3%，是全球唯一实现经济正增长的主要经济体，2021年同比增长8.1%¹，在全球主要经济体中名列前茅，诸多国际组织看好未来中国经济发展。根据中国旅游研究院，旅游行业正在步入“相对景气”到“景气”的通道。中国作为全球最大的商旅市场，疫情的确带来负面影响，但疫情缓解令商旅需求迅速反弹，2021年中国企业差旅支出预算情况明显好转。

新冠疫情常态化防控成为背景，短程旅游迎来暴涨风口。在出境游停滞，跨省游存在较高风险时，更多游客会选择短程旅游。在该背景下分析游客消费需求行为的变化，对于旅游公司的产品销售、策略制定和市场开拓具有长远的作用。同时随着旅游板块的核心投资逻辑“修复+成长”成为共识，疫情边际影响减弱和政策的调整，越来越多的业务得到“修复”[1]。

旅游发展抓住数字经济的潮流，充分利用用户生成内容（UGC）数据。从文本中提取相关的旅游要素，挖掘其相关性和隐含的关系，转化为新的“成长”动能。当地旅游部门可以利用本研究结果有侧重点投资和宣传旅游产品，游客在选择时也将更具有目的性，充分利用线下本土化资源，即利用绘制本地旅游图谱成为全新的增量挑战。在帮助厘清当地旅游现状的同时，把握其热度和特点形成二次旅游经济拉动[2]。

1.2 工作及思路

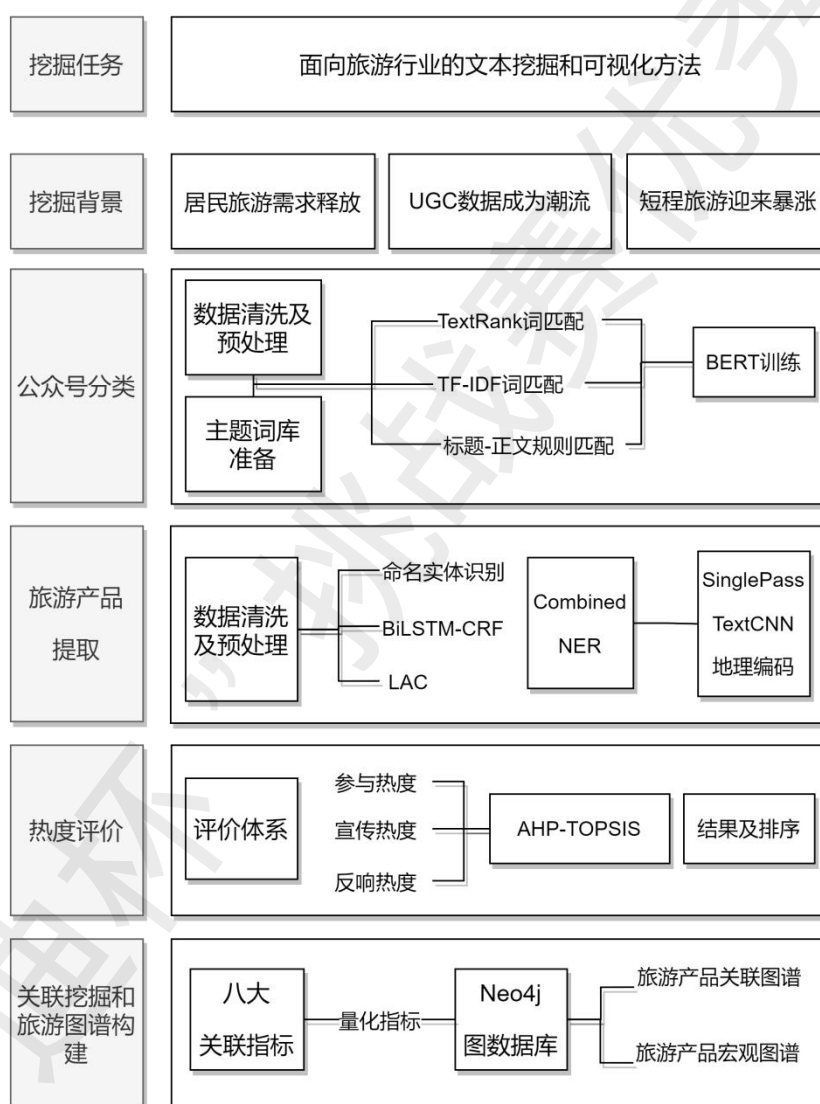
旅游产品是指旅游经营者为了满足旅游者在旅游活动中的各种需要，凭借各种旅游设备设施和环境条件向旅游市场提供的全部服务要素之总和。旅游供给是在一定时期内、一定市场上，旅游经营者在一定价格水平下向旅游市场提供的旅游产品数量。旅游产品包含旅游客体（即旅游吸引物）和旅游介体（指交通、通讯、住宿、餐饮、购物等企业）为旅游活动提供的支持性服务。

面对大数据环境下的资源管理和集成困境，知识图谱²应运而生，然而领域知识图谱的构建是具有行业依赖性的，各行各业有不同的概念体系、数据源以及复

1 数据来源：世界旅游联盟与浩华联合发布的《2021年中国国内旅游市场景气报告》

2 是结构化的语义知识库，用于以符号形式描述物理世界中的概念及其相互关系

杂语言体系，必须根据相应的领域进行设计。对于旅游知识图谱而言，游客是旅游业兴旺发展的基础，而游客所有关注的重点是游、食、住与购，构建时也应重点考虑该方面的知识抽取；对于该题目提出的文本挖掘任务和相关的可视化方法。本文基于三种现实挖掘背景需求，针对数据中的公众号分类任务进行三重匹配进行 BERT 训练，实现产品提取、热度评价、关联挖掘和图谱的可视化实现等多重任务，最后利用图数据库实现交互式的旅游图谱并给当地管理部门提出三大具有建设性的建议，具体思路如下。



第二章 针对问题一的解决方案

2.1 问题分析及思路

题目要求需要构建一个文本分类模型，根据微信公众号推送的文章进行分类，主要根据其主题与文旅的相关性，分成“相关”和“不相关”。其中大量的

文旅相关主题已经呈现，例如：旅游、假期、活动、酒店等，更多的文旅主题需要团队进行挖掘。

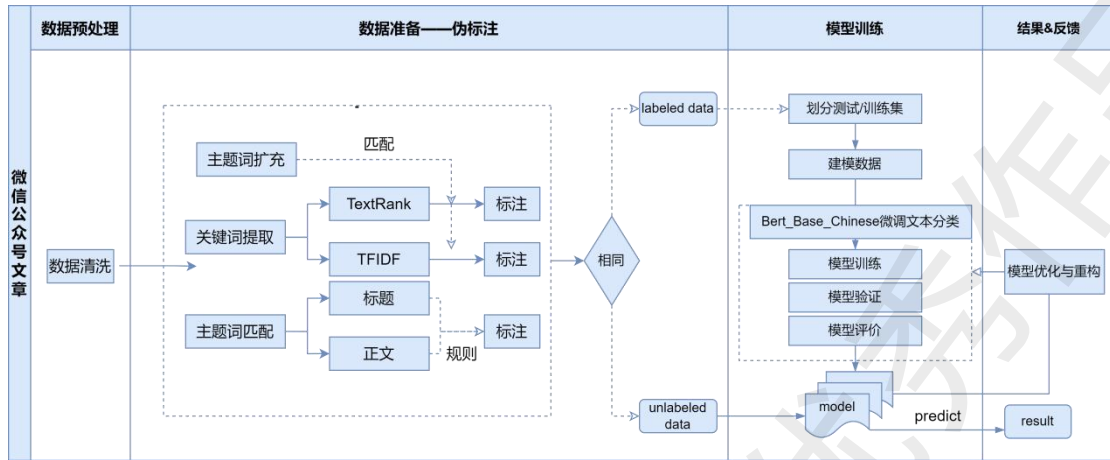


图 2 新闻文本分类流程图

本章基于题目所给的文旅相关主题词，进行近义词扩充形成主题词库；采用 TextRank、TF-IDF 分别提取语料关键词，与主题词库进行匹配，同时采用规则对标题和正文进行直接匹配，共得到 3 个标签，3 个标签一致的作为伪标注数据用于训练模型。基于中文 Bert 预训练语言模型训练分类器，用于以上三种方法得到结果不一致时的判别依据。

2.2 文本分类全流程

2.2.1 文旅词库扩充和数据预处理

基于题目所给的文旅相关主题词进行扩充筛选，对于筛选后的每个主题词，使用 Synonyms 库生成 5 个同义词，扩充形成文旅相关主题词库，共包括 143 个主题词。此外，经探索数据中不存在完全重复的数据，筛去空格和特殊字符后采用 Jieba 模块进行分词，并筛选保留词性为名词、动词的词语。

2.2.2 TextRank、TF-IDF 关键词提取与判别

TextRank 算法是一种基于图的用于关键词抽取和文档摘要的排序算法，由谷歌的网页重要性排序算法 PageRank 算法改进而来，它利用一篇文档内部的词语间的共现信息(语义)便可以抽取关键词，它能够从一个给定的文本中抽取该文本的关键词、关键词组，并使用抽取式的自动文摘方法抽取该文本的关键句。其基本思想是将文档看作一个词的网络，该网络中的链接表示词与词之间的语义关系，具体公式如下，其中 $WS(V_i)$ 表示句子 i 的权重，右侧求和表示每个相邻句子对本句子的贡献程度， d 是阻尼系数，一般为 0.85。

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{W_{ji}}{\sum_{V_k \in Out(V_j)} W_{jk}} WS(V_j)$$

采用 TextRank、TF-IDF 提取正文 15 个关键词，并将标题分词筛选词性后加入到关键词中，若出现主题词库中的词，则认为其与文旅相关，否则不相关；采用两种方法得到两种是否与文旅相关的标签。

2.2.3 标题、正文规则判别

采用规则直接匹配标题和全文，如果标题中出现主题词或正文出现 20 个以上主题词，认为其与文旅相关，否则不相关，得到一种是否与文旅相关的标签。

2.2.4 模型训练数据准备

通过以上 3 种方式得到 3 个标签，3 个标签均一致的认为分类正确，作为伪标注数据用于模型训练；3 个标签不同的认为分类存疑，使用训练的模型进行预测。

2.2.5 Bert 分类模型训练

由于对于文本分类这个场景来说其输入只有一个序列，所以在构建数据集的时候并不需要构造 Segment Embedding 的输入，直接默认使用全为 0 即可。基于生成的伪标注语料，重新使用 TextRank、TF-IDF 各提取 50 个关键词，并将标题分词筛选词性后加入到关键词中与标题结合去重后作为模型输入，使用中文 Bert 预训练语言模型进行训练，得到一个公众号文章关键词到文旅相关性的分类模型，并用于分类存疑的数据。

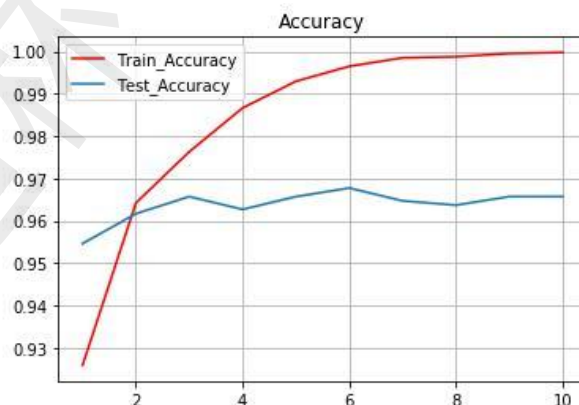


图 3 Bert 模型训练准确率变化



图 4 Bert 模型训练损失变化

观察图形不难发现，Bert 预训练微调文本分类模型在新闻文本分类数据集上训练效果良好，损失函数不断降低，多次实验发现 10 个 epoch 之后模型的准确率不再提高，因此最终实验将 epoch 值取 10，模型在测试集的的精确率达到 98.23%，在测试集上精确率为 96.57%，使用预训练微调的方式弥补了训练数据的不足问题，分类效果也远超大部分文本分类模型。

第三章 针对问题二的解决方案

3.1 问题分析及思路

为充分挖掘茂名市的旅游发展现状和商业服务，本文需要通过四个数据表格提取旅游产品，涵盖酒店、餐饮、文旅等，并通过构建热度指标体系，采用相关数学模型计算单个旅游产品的热度值并按年度进行排名和分析。

为解决以上问题，本章首先通过去重等阶段进行数据清洗，其次利用爬取茂名旅游景点作为补充词库，通过 BiLSTM-CRF 模型和 LAC 分词工具实现高精度的命名实体识别集合。为避免将不同的类别实体聚为一类，采用 TextCNN 进行分类；分类后的结果用 Single-Pass 算法的命名实体聚类方法进行聚类，形成以餐饮、酒店、景区为代表的实体集，在同义实体合并和无效实体筛除阶段后，进行回溯和编码，最终实现旅游产品提取的全流程。

3.2 旅游产品提取流程

3.2.1 数据清洗

对非茂名旅游产品、存在友情链接重复内容、去停用词和繁体文字进行处理。数据清洗的步骤包括文本去重、压缩去词、短句删除等。在分析景区数据时，

将存在部分不属于茂名的景区评论（例如：中国南极长城站、广东海洋大学等）整行删除；游记攻略中存在的友情链接等大量重复内容，将严重影响产品提取和热度评价效果，由于结构较为统一，采用正则表达式过滤，删去网址和网址前的标题；此外，数据基于川大机器学习实验室停用词库，结合本题实际应用场景进行了补充；最后，由于数据存在部分繁体内容，使用 python 中的 OpenCC 库将繁体内容转化为简体。

3.2.2 数据准备

本文从马蜂窝网站爬取茂名主要热门旅游景点（去重处理后共 200 条），结合题目景区评论表中的景区名称³，最终形成“茂名旅游景点词库”，加入 LAC 分词工具的用户词典，特用于提升针对茂名的分词和旅游产品识别效果。此外，本文基于部分餐饮、景区、酒店名称的结构特征，形成“分类特征词库”，主要用于提升命名实体分类的准确性。

3.2.3 基于 BiLSTM-CRF 和 LAC 的命名实体识别

命名实体识别（Named Entity Recognition, NER）是指从一段自然语言文本中找出相关实体，标注其位置及类型，实际上是序列标注问题。条件随机场（Conditional Random Fields, CRF）[3]CRF 模型统计全局概率，在归一化时考虑数据在全局的分布，而不是仅仅在局部进行归一化。在传统机器学习中，CRF 被看作是命名实体识别的主流模型，优点在于在对一个位置进行标注的过程中 CRF 可以利用内部及上下文特征信息。BiLSTM-CRF 直观显示了模型结构与优势，双向 LSTM 模型能够同时捕捉正向信息和反向信息，CRF 则同时考虑到标注序列的合理性。模型主要由 Embedding 层，双向 LSTM 层和 CRF 层构成，具体结构如下图。

本文使用 BiLSTM-CRF 模型基于 CLUENER 细粒度命名实体识别语料库⁴，筛选其中的“旅游”和“地址”类别标注实体，训练针对旅游领域构建的 NER 模型。

³ 备注：如果与马蜂窝网站景点存在重合，取其中较长的作为标准名。

⁴ 语料库链接：<https://github.com/CLUEbenchmark/CLUENER2020>

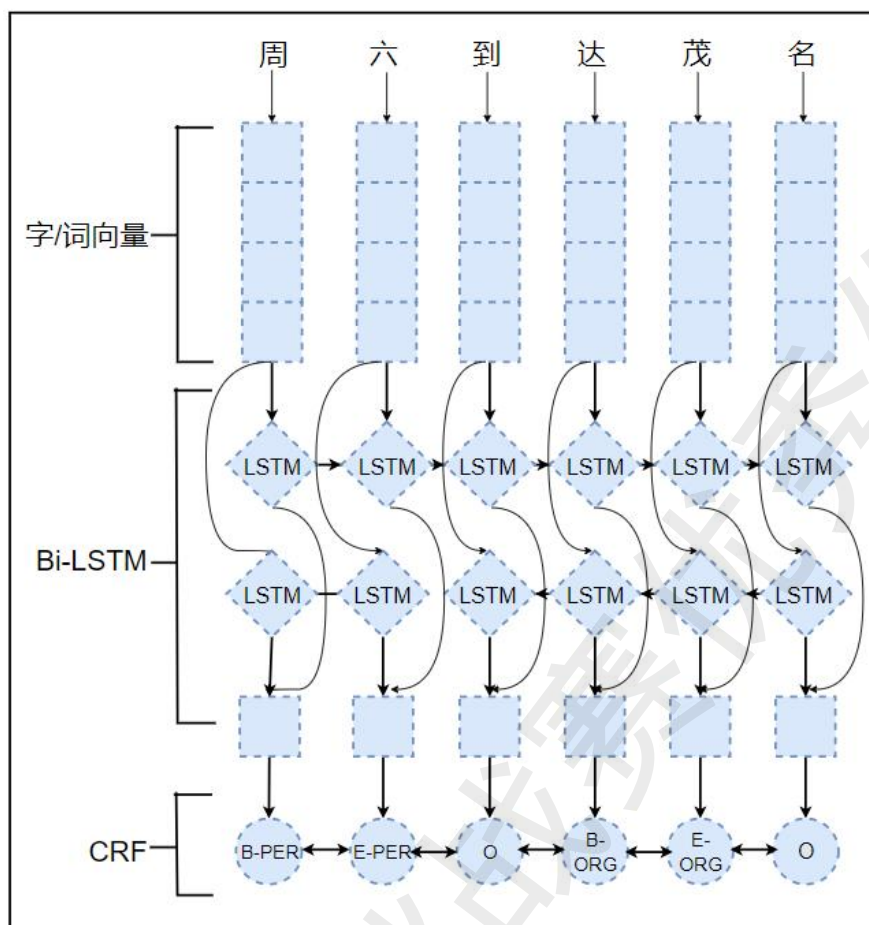


图 5 BiLSTM-CRF 结构图

3.2.4 基于 TextCNN 的命名实体分类

经过标志词匹配后，将在 3.2.3 中提取出的旅游产品，为避免将不同的类别实体聚为一类，本文根据数据特点和当地景区特点将命名实体分为三类：餐饮、酒店、景区。采用的方法是命名实体分类中的标志词匹配和 TextCNN。

TextCNN 的基本思想是利用 CNN 的结构进行分类，该模型将词向量组成的句子利用不同大小的窗口信息，提取相关特征。从结构上 TextCNN 分为输入层、卷积层和池化层，如下图所示。

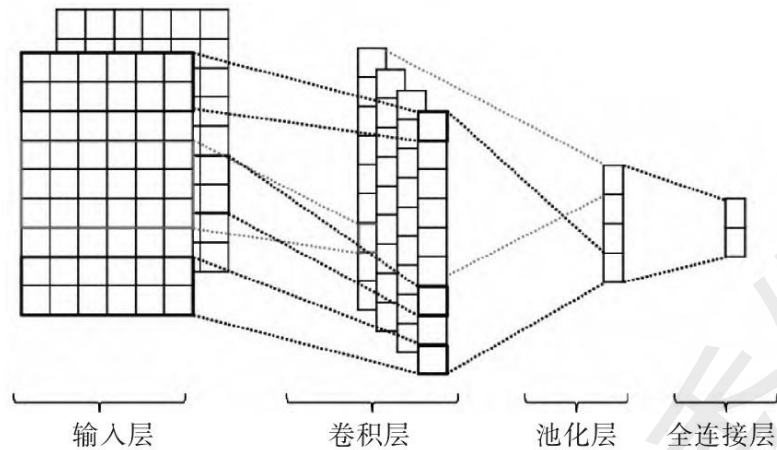


图 6 TextCNN 模型结构

TextCNN 的输入层为一个文本的字向量组成的矩阵，由横向为每个字向量，同时为了保持字向量长度一致，进行了补 0 操作。卷积层即在纵向和横向两个方向保持卷积核与词向量宽度相同；池化层采用最大池化，即从滑动窗口中产生的特征向量筛选出最大特征[4]。

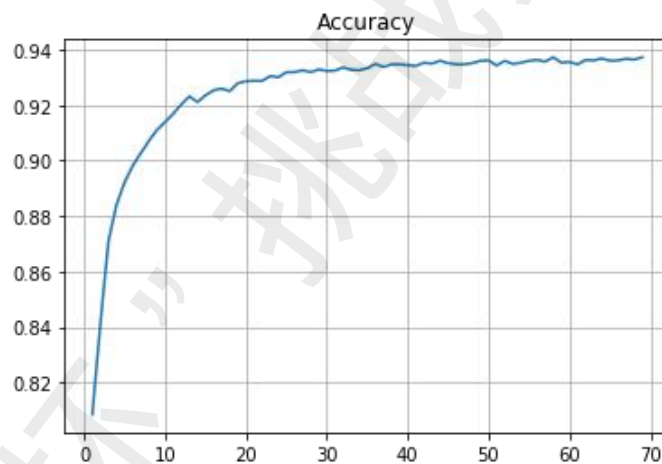


图 7 TextCNN 模型训练准确率变化

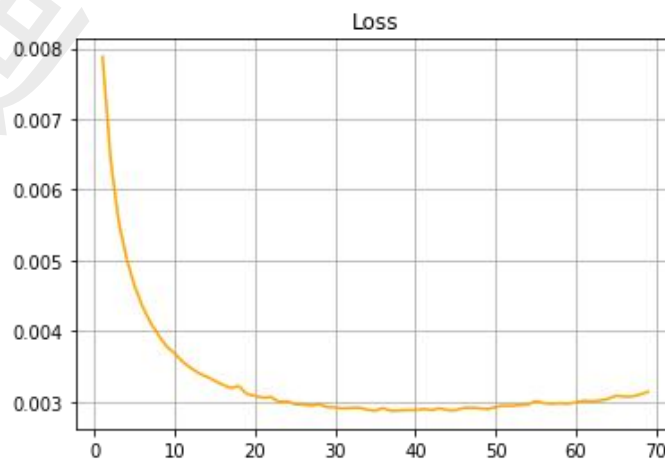


图 8 TextCNN 模型训练损失变化

3.2.5 基于 Single-Pass 算法的命名实体聚类

Single-Pass clustering 是一种简洁且高效的文本聚类算法，由于同实体存在不同表述方式、错别字、识别不够完整等情况，在命名实体分类的基础上，按分类对全部实体进行基于文本相似度的聚类。由于命名实体通常较短，而 Single-Pass 是一种短文本聚类的算法[5]，适合进行命名实体的聚类。

Single-Pass 聚类顺序处理输入的单个文本，每次处理一篇便增量跟更新聚类。预先设定一个相似度阈值 T_c ，如果新文本与已有文档相似度超过 T_c ，即可以归为一簇；否则根据该文本再创建一个新文档，通过调整 T_c 可以控制聚类的粒度。本文具体操作如下图：

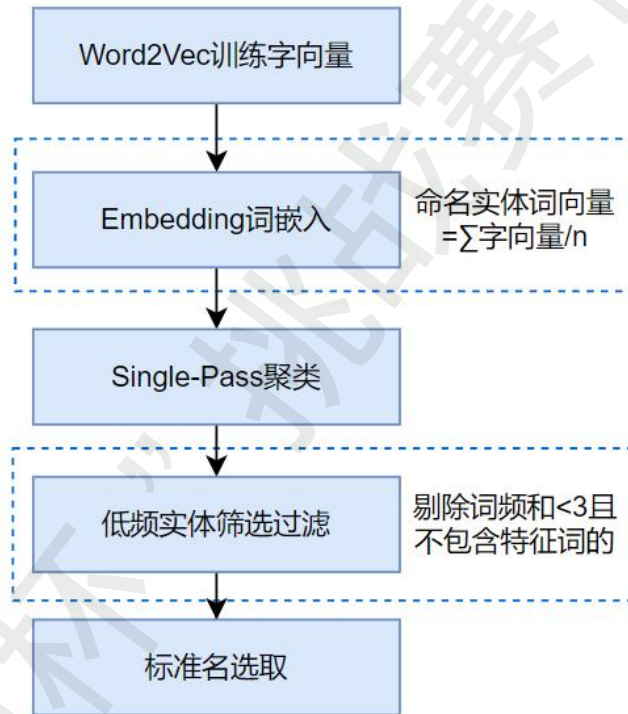


图 9 基于 Single-Pass 算法的命名实体聚类

- (1) Word2Vec 训练 200 维的字向量。
- (2) Embedding 词嵌入。这里因为是字向量就是把每个字转换为向量，各字的向量求平均之后得到整个命名实体的向量，两个向量的余弦相似度可以衡量这两个命名实体的相似程度；
- (3) Single-Pass 聚类：对每一个新的命名实体，计算它与已经成组的命名实体词中每个词的余弦相似度，公式如下：

$$TFIDF \text{计算公式: } TFIDF = \underbrace{TF}_{\text{词频}} \times \underbrace{IDF}_{\text{逆文档频率}} = \frac{n_{ij}}{\sum_k n_{kj}} \times \log\left(\frac{|D|}{1+|D_i|}\right)$$

图 2.2 TF-IDF 计算方式

其最大值高于阈值则分入该组；计算完没有满足条件的组则独立成组。

(4) 低频实体筛选过滤：对于聚类后整组词频之和少于 3 条且不包含实体特征词的，认为不位于茂名或识别效果不佳，予以剔除；

(5) 标准名选取：首先匹配“茂名景点词库”中包含的实体名作为标准名，其次采用组内数量最多的实体名作为标准名，最后采用长度最大的实体名作为标准名；

最后聚类结果形成以 0-餐饮、1-景区、2-酒店为类别的实体集合。

3.2.6 同义实体合并和无效实体剔除

调用百度地图 API 获取地理位置信息⁵，在茂名市范围内检索该实体并获取经纬度，具体可参见表 1 中的地理坐标。对于经纬度坐标相同的实体可以认为是同一实体并进行合并；若补充地理结构多次检索后仍不存在，则可以认为该实体不位于茂名或者指代过于宽泛，则剔除该实体。

表 1 关于提取的旅游产品实体相关属性

产品名称	产品类型	频数	地理坐标	所有表述方式
茂名浪漫海岸温德姆酒店望海餐厅	0-餐饮	5	(21.468733, 111.15005)	['茂名浪漫海岸温德姆酒店望海餐厅', '望海餐厅']
熊猫餐厅	0-餐饮	1	(21.903363, 110.864454)	['熊猫餐厅']
茶餐厅	0-餐饮	1	(21.664978, 110.942047)	['茶餐厅']
碇洲岛	1-景区	22	(21.669232, 110.912828)	['汾洲岛', '同心公园', '碇洲岛', '碇洲岛', '探胜岛', '湄洲岛']
放鸡岛	1-景区	77	(21.387867, 111.197295)	['潜梦岛', '搜放鸡岛', '放鸡岛海洋公园', '放雞島', '放鸡岛']
仙人洞	1-景区	21	(22.262326, 111.233761)	['网红瀑布', '仙人洞村', '仙人洞']

⁵ 具体参考网址：<https://lbsyun.baidu.com/>

柏曼酒店	2-酒店	2	(21.66949, 110.940213)	['柏曼酒店', '柏曼酒店']
IU酒店	2-酒店	2	(21.662472, 110.924481)	['IU酒店', 'iu酒店']
茂名凯里亚德酒店	2-酒店	5	(21.657806, 110.91179)	['茂名凯里亚德酒店', '凯里亚德酒店']

3.2.7 实体编码与回溯

该阶段共提取到 **631** 个有效实体，累计 **1014** 种表述方式，例如：酒店中的实体“维也纳酒店”，具有另外四种表述方式：“維也納酒店”“维也纳3好酒店”“维也纳酒店”“宿维也纳酒店”。对实体进行编码，同一实体的不同表述方式共用同一编码。

最后对实体回溯来源，建立 **631** 个有效实体和游记评论关联共 **10515** 篇次，其中不同实体算多次、同一篇多次提及同一实体不再进行重复计算⁶。将题目所需格式整理至 result2-1.csv 中，具体可参见表 2。

表 2 旅游产品与语料的关联——以游记攻略为例

语料 ID	产品 ID	产品名称
游记攻略-1001	ID10	巽寮湾
游记攻略-1001	ID11	汾洲岛
游记攻略-1001	ID12	放鸡岛
游记攻略-1001	ID42	湾舟岛
游记攻略-1003	ID12	放鸡岛
游记攻略-1003	ID13	仙人洞
游记攻略-1003	ID14	高州市博物馆
游记攻略-1003	ID15	御水古温泉
游记攻略-1003	ID16	广东石油化工学院
游记攻略-1003	ID17	茂名博物馆

3.3 基于 AHP-TOPSIS 的多维度热度评价模型

3.3.1 热度评价体系构建

评价体系构建是热度评价的核心环节，其设计应涵盖游客在选择旅游产品时的各类决策考量[6]。本文参考大量旅游顶刊和咨询研报的指标选取，结合本题旅游产品的特征最终确定相关评价指标，如下表 3 所示，最终分为三个层次：顶

⁶ 若考虑多种表述情况，则建立关联 20434 篇次。

层为目标层，即旅游产品的多维度热度评价体系；中间层为准则层，包括 3 个一级指标，该层次需要考虑游客接受旅游产品宣传或者参与的全过程，参与热度是指产品被购买或者被体验的热度评价，反响热度强调游客对产品的体验评价，宣传热度是指旅游产品热度间接传达到游客的效率；底层为指标层，包括 7 个二级指标，该层次为热度评价的具体指标，具体解释见表 4。

表 3 评价体系

目标层								旅游产品热度评价体系							
一级指标		参与热度			反响热度			宣传热度							
二级指标		评价次数	活跃天数	周围文旅产品密度	游客印象	回购热度	公众号文章	游记攻略							

表 4 指标详细解释

评价指标		计算方法	
评价次数	频数，越高越好，数据四个表格累计次数（其中游记评价按出发时间计算）		
活跃天数	频数，越高越好，数据中四个表格累计天数加和		
周围文旅密度	越高越好，百度地图获取经纬度坐标计算 5km 距离内旅游产品个数		
游客印象	越高越好（接近 1），对每个评论/游记中描述该旅游产品的部分进行情感分析 ⁷		
回购热度	越高越好，旅游产品提到“下次再来”“住了多次”“第二次来”等回购标识的评论数		
公众号文章	越高越好，公众号文章提到的次数		
游记攻略	越高越好，游记攻略中提到的次数（以发布时间为准）		

3.3.2 AHP 确定指标权重

AHP 法是由美国 T.L.s aaty 提出，其优点是可将模糊或难以量化的问题分解为若干层次和多个指标实现定量与定性的结合，通过数学模型运算，最终确定各指标间权重[7]，具体步骤如下：

(1) 建立指标体系，具体见 3.3.1。

(2) 构造对比矩阵，判断矩阵权重。根据根据专家意见，比较各指标间的重要性，得出判断矩阵。判断矩阵采用 9 度法打分，“1”表示 i 比 j 同样重要；

⁷ 情感分析：

“3”表示*i*比*j*稍微重要；其余同理。判断矩阵权重计算公式如下：

$$\bar{W} = \sqrt{\prod_{j=1}^n (j - 12 \dots 1)} \quad (1)$$

$$W = \frac{\bar{W}_1}{\sum_{i=1}^n \bar{W}_1} \quad (2)$$

本章计算权重大小值为[1.893,0.976,0.851,0.851,0.81,0.81,0.81]^T。

(3) 计算最大特征根。公式如下：

$$\lambda_{\max}^2 = \sum_{i=1}^n (AW)_i \quad (3)$$

(4) 检验一致性。当 $CR < 0.1$ ，说明权重系数分配合理，可见表 5。即最终得到最底层即方案层各决策方案相对于总目标的权重，并给出这一组合权重所依据整个递阶层次结构所有判断的总 CR，最终得出决策结果。一致性检验公式如下：

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (4)$$

$$CR = \frac{CI}{RI} \quad (5)$$

表 5 随机一致性 RI 表格

随机一致性 RI 表格														
n 阶	3	4	5	6	7	8	9	10	11	12	13	14	15	16
RI 值	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.52	1.54	1.56	1.58	1.59	1.594
n 阶	17	18	19	20	21	22	23	24	25	26	27	28	29	30
RI 值	1.60	1.61	1.62	1.62	1.63	1.64	1.64	1.64	1.65	1.65	1.66	1.66	1.66	1.672
	64	33	07	92	58	03	62	97	56	87	31	7	93	4

3.3.3 TOPSIS 排序过程

TOPSIS 法基于标准化后的数据寻找正、负理想解，并通过计算相对贴进度来评价各方案的优劣。基于 AHP 法计算得到各指标权重，并采用 TOPSIS 法对热度指标进行进一步评价，具体步骤如下

(1) 规范化矩阵 R。根据层次总排序形成方案层与目标层矩阵并进行规范化，得到规范化矩阵 R:

$$R = |r_{ij}| \quad (6)$$

(2) 将方案属性权重向量 (w) 和规范化矩阵 R 结合, 构造加权规范化决策矩阵 V。

$$w = \begin{bmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_n \end{bmatrix} \quad (7)$$

$$V = |V_{ij}| = R \times w \quad (8)$$

$$V = \begin{bmatrix} W_1 \Gamma_{11} & W_2 r_{12} & \cdots & W_n \Gamma_{1n} \\ W_1 \Gamma_{21} & W_2 r_{22} & \cdots & W_n \Gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_1 \Gamma_{m1} & W_2 \Gamma_{m2} & \cdots & W_n r_{mn} \end{bmatrix} \quad (9)$$

(3) 根据矩阵 V, 选择每个方案各个关键指标中得分最多的作为正理想解 (V+), 得分最少的负理想解 (V-), 并计算各方案与正负理想解之间的距离, 计算公式如下:

$$D_i^+ = \sqrt{\sum_{j=1}^n (V_{ij} - V^{+2j})^2} \quad i = 1, 2, \dots, m \quad (10)$$

$$D_i^- = \sqrt{\sum_{j=1}^n (V_{ij} - V^{-2j})^2} \quad i = 1, 2, \dots, m \quad (11)$$

(4) 确定相对接近度, 并对评价对象的优劣进行排序, 计算公式如下:

$$C_i = D_i^- / (D_i^+ + D_i^-) \quad i = 1, 2, \dots, m \quad (12)$$

表 6 AHP-TOPSIS 热度排序结果展示随机一致性 RI 表格

产品 ID	产品类型	产品名称	产品热度	年份
ID23	0-餐饮	清香面包店 (车田街店)	96.14	2018
ID30	0-餐饮	小乔紫菜卷 (小乔店)	80.08	2018
ID27	0-餐饮	果之度 (为民店)	79.86	2018
ID29	0-餐饮	优之品西点	78	2018
ID10	0-餐饮	甜在心扉千层蛋糕 (茂名店)	77.52	2018

ID18	0-餐饮	盛香烧鹅（东方市场店）	69.52	2018
ID25	0-餐饮	优乐堡（开发区店）	66.69	2018

3.4 可视化分析

为进一步形象呈现本文抽取的茂名市旅游产品地理分布，我们采用某平台的开源交互程序⁸，对景区、酒店和餐饮⁹三类旅游产品进行地理位置的可视化，分别用蓝色、绿色、红色进行区别，具体可见下图。进一步分析该图可得，旅游产品总体较均匀分布在茂名市，但其中大多景区集中在茂名市的南部，此分布图也将有助于当地政府管理部门对资源管理和调度有更清晰的认知。

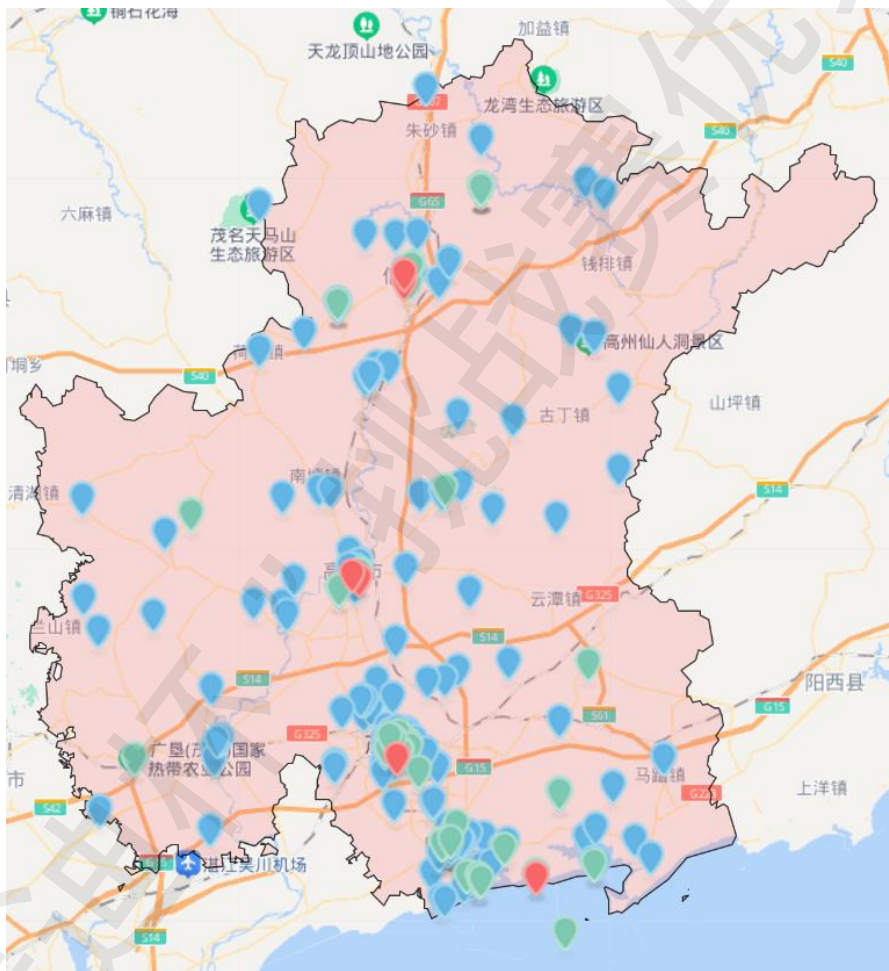


图 10 茂名市旅游产品分布

第四章 针对问题三的解决方案

⁸ 本文的交互项目链接：<https://zt.changjing.com.cn/map/viewer?mid=atgnCaIQxUj5Ck0SVGifDw>

⁹ 为更清晰呈现产品差异，餐饮类产品没有进行全部展示。

4.1 问题分析及思路

本章需要针对问题 2 提取出的旅游产品进行关联分析，不仅需要充分考虑到旅游产品之间的强关联模式，更要充分从地理位置、评价特征等多维度挖掘产品间隐含关系[8]，并且将关联度进行量化。在此基础上进行旅游图谱的构建，并在宏观和微观层面进行多视角的可视化呈现与分析。

故本章针对问题三，首先在前一阶段的数据准备和实体抽取后进行关联模式的定义，根据本地经济发展程度、地理位置等诸多客观及其衍生的主观因素定义了八种关联模式，其中包含三种隐含关系的模式；最后利用 Neo4j 导入节点的数量大、速度快、支持查询算法等特质，将结点和关系均存入 Neo4j 数据库中，最后进行宏观和微观的分析。

4.2 本地旅游产品关联图谱构建及分析

4.2.1 关联模式定义

针对关系抽取任务，结合游客需求和具体应用场景，从 result2-1 中旅游类产品中 0-餐饮、1-景区、2-酒店三大类出发，总结概括出八种关联模式，其中餐饮类旅游产品的关系共 5 种，景区 5 种，酒店 6 种，具体参考下表 7。

表 7 八种关联模式

关联模式	量化方法（标准化前）	备注
共现关系	A、B 出现次数大于等于 3 时， $P(AB)/P(A)+P(AB)/P(B)$	三类产品均适用
近邻关系	1/地理距离	地理距离<500m 的记为 500m（下同）
辐射关系	景区热度/(周围同类产品密度·地理距离 ²) ^{1/3}	限景区与酒店/餐饮之间
竞争关系	1/(经济距离·地理距离 ²) ^{1/3}	经济距离= A 餐饮酒店热度-B 餐饮酒店热度 ，限酒店/餐饮同类之间
导流关系	(A 景区热度+B 景区热度)/AB 景区地理距离	限景区之间
餐饮风格相似	名称、标题、评论内容关键词相似度	限餐饮之间
酒店风格相似	名称、房型设置、评论内容关键词相似度	限酒店之间
历史文化关联	历史文化、高层概念语料中共现次数	基于茂名百度百科历史文化、风景名胜构建语料，限景区之间

由于旅游类产品的关联模式强烈依赖于本地经济发展程度、地理位置等诸多

客观及其衍生的主观因素，故关联模式指标选取和计算需充分考虑茂名市当地情况。具体模式关系解读如下：

(1) **共现关系**。“共现”一词指的是有关两事物的描述在同一样本中出现的现象。旅游产品间的共现指的是在当前样本中出现了其他旅游产品的描述[10]，也就是说当前产品和其它产品之间有着一种指向性关系。在本文中，计算共现度可以反映两个旅游产品间联系的紧密程度，由产品 A 对产品 B、产品 B 对产品 A 的相对共现度加和得到。

(2) **近邻关系**。“近邻”是指两个旅游产品地理位置上的关系，距离位置在一定距离内的临近会促进双方产品被消费[11]。本文中的地理位置计算根据经纬度坐标计算两点球面距离，其中距离较近 (<500m) 的产品统一记作 500m，下同。

(3) **辐射关系**。景区对周围的餐饮和酒店等旅游产品热度均有一定的辐射带动作用[12]，辐射作用的大小由景区热度、被带动产品周围同类产品密度、景区和被带动产品间的地理距离决定，其中：景区热度越大，辐射带动能力越强；被带动产品周围同类产品密度越小，同类竞争越小，辐射带动能力越大；景区与被带动产品距离越近，辐射带动能力越强。本文将通过计算表 7 中的相关公式进行辐射作用的量化，并且进一步分析产品间的竞争力和变化趋势等。

(4) **竞争关系**。旅游领域大部分研究竞争力指标模型的，大多采用灰色关联分析法，该方法适用于一个系统发展变化态势的定量描述和比较。对于本文要构建的竞争关系，主要建立在餐饮和酒店相似度较高基础上的研究，主要影响因素包括地理距离和经济距离。地理距离越小竞争越大；经济距离指餐饮或酒店的定位、规模差异，用第二题得到的热度进行量化，经济距离越小竞争越大[13]。

(5) **导流关系**。景区之间由于热度和距离等因素，势必会存在某现象：A 店火，带动 B 店，本章将此现象定义为导流关联模式，计算公式主要是利用影响因素中的景区热度、景区距离，即等于 AB 热度之和/AB 距离。

4.2.2 针对隐含关系的抓取

现有的实体关系抽取研究大多集中在通用领域，针对特定领域研究较为匮乏，尤其在旅游领域[14]。网络中的旅游类文本，大多长度较长并且存在大量与三元组无关表述。例如，评论“十里银滩位于亚婆角度假区内，这里海岸线绵长，

沙质细腻而洁白。”该评论可能涉及主体较多且包含冗余文字较多。另外，对旅游类文本进行实体抽取应侧重于挖掘文本内部的语义联系，涉及更丰富的模型结构，尤其注意三元组重叠情况。故针对隐含关系的抓取，本文采用挖掘餐饮风格、酒店风格 and 历史文化关联三种形式，其中历史文化关联是指景点的历史溯源至同一典故或人物等情况。

(6) **餐饮风格相似**。在判断餐饮类旅游产品风格是否类似，根据名称、标题、评论内容关键词相似度，同餐饮评论标题、内容聚合，使用 TextRank 算法提取前 20 的关键词，可见图 11，与名称合并，然后通过余弦相似度计算不同产品的文本相似度，最终判断餐饮风格是否类似。

	餐饮名称	content	产品ID
0	CAKE情迷黑森林	蛋糕 榴莲 草莓 芒果 不错 切件 新鲜 奶油 口味 味道 感觉 性价比 有点 水果 夹心 ...	ID20
1	Hello炸鸡 (方兴店)	炸鸡 奶茶 同款 不错 味道 热情 蜂蜜 套餐 老板娘 芥末 口味 环境 老板 奶酪 柚子 ...	ID66
2	LINLEE 手打柠檬茶 (文创店)	柠檬茶 邻里 环境 店员 不错 口味 套餐 奶茶 哥哥 整体 美团 店里 港式 味道 性价比...	ID65
3	三两粉 (茂名东汇城店)	套餐 单人 经典 不错 环境 口味 长沙 性价比 味道 分量 服务员 整体 小菜 总体 态度...	ID67
4	丰年面包店 (茂南区店)	英寸 水果 蛋糕 套餐 圆形	ID31
...
68	鲜滋味蛋糕 季念生日蛋糕 (茂南区店)	榴莲	ID17
69	鲜饮时光 (高州店)	包间 寿司 原味 单人 味道 不错 老板 紫菜 口味 用餐 感觉 价格 套餐 舒服 性价比 ...	ID15
70	麦壳西点 (开发区店)	蛋糕 英寸 便当 芒果 老板 新鲜 不错 水果 味道 奶油 圆形 性价比 套餐 价格 口感 ...	ID40
71	麦当劳 (明湖店)	圆筒 金粉 不错 薯条 套餐 口味 甜品 环境 服务员 味道 原味 性价比 尴尬 旋风 整体...	ID77
72	黑色经典长沙臭豆腐 (为民路店)	臭豆腐 套餐 长沙 单人 建议 不错 肠胃 感觉 朋友 南市区 味道 椰果 咸味 香脆 花生...	ID52

图 11 餐饮风格相似部分结果呈现图

(7) **酒店风格相似**。判断酒店风格是否类似，思路同理，用名称、房型设置、评论内容关键词相似度，同酒店评论标题、内容聚合，使用 TextRank 算法提取前 20 的关键词，可见图 12，与名称、去重的房型名合并，计算不同产品的文本相似度。

酒店名称	入住房型	content	产品 ID
0 IU酒店(茂名人民南路油城大厦店)	[小U舒适大床房, 小U舒适大床房, 小U舒适双床房, 小U舒适双床房, 小U舒适大床房, ...]	小U舒适双床房 小U舒适大床房酒店 房间 干净 不错 前台 方便 热情 性价比 舒服 服务态...	ID564
1 Q加 茂名大禧假日酒店	[商务标准双人房, 商务标准双人房, 商务标准双人房, 豪华蜜月房, 商务标准双人房, 豪华...	豪华蜜月房 蜜月房 豪华双人房 商务标准双人房 豪华行政大床房 豪华商务酒店 不错 干净 ...	ID538
2 信宜丽晶酒店	[标准大床房]	标准大床房维也纳 空调 不错	ID614
3 信宜天鹅湖宾馆	[标准大床房]	标准大床房	ID617
4 兰欧酒店(信宜绍秀体育馆店)	[兰欧豪华大床房]	兰欧豪华大床房态度 美女 房间 前台 很棒 支持 很好 兰欧 宽敞	ID557
...
74 高州简尚精品酒店	[特价房]	特价房	ID582
75 高州顺得商务酒店	[标准双人房]	标准双人房吃饭 逛街 购物 方便 车位 停车 热闹	ID608
76 麗枫酒店(茂名水东店)	[豪华双床房, 豪华大床房, 豪华大床房, 雅致大床房, 雅致大床房, 雅致大床房, 雅致大...	行政大床房 雅致双床房 豪华大床房 雅致大床房 豪华双床房房间 方便 交通 前台 舒适 还有...	ID496
77 麗枫酒店(茂名电白万达广场店)	[豪华大床房, 豪华大床房, 豪华大床房, 豪华大床房, 豪华大床房, 豪华大床房, 豪...	行政大床房 家庭房 豪华双人房 豪华大床房 浪漫影音优享房 豪华双床房酒店 房间 方便 不错...	ID625
78 麗枫酒店(茂名高铁站店)	[标准大床房, 豪华大床房]	标准大床房 豪华大床房环境 服务态度 干净 方面	ID496

图 12 酒店风格相似部分结果呈现图

(8) 历史文化关联。历史文化关联这一关系没有具体量化指标时，主要是基于茂名百度百科历史文化、风景名胜构建语料，挖掘历史文化、高层概念语料中的隐含关联模式，通过在两个语料中共现的频次来衡量其历史关联指标度，具体可见下图 13，例如“洗太夫人诞”，同义词“长坡旧城”、“洗太庙”、“高州城”等在 ID96、ID99、ID182 中均有共现，故这三类旅游产品可以认为具备历史文化关联。

```

Out[26]: [{'洗太夫人诞': ['长坡旧城', '洗太庙', '高州城'],
          {'滨海旅游六点一线': ['放鸡岛', '博贺港', '中国第一滩旅游度假区', '水东湾', '浪漫海岸国际旅游度假区'],
          '五大特色龙头旅游': ['滨海旅游': ['浪漫海岸', '滨海旅游度假区', '月亮湾滨海度假区', '虎头山海滨旅游区', '滨海渔村'],
          '洗夫人文化旅游': ['洗夫人庙', '洗太广场', '洗太夫人故里文化旅游区'],
          '潘茂名养生旅游': ['潘茂名纪念公园', '潘茂名纪念馆'],
          '美丽乡村风情游': ['油甘窝村', '牙象村', '杏花村'],
          '山地运动休闲游': ['天马山生态旅游', '茂名森林公园', '沉香森林公园', '高凉岭森林公园', '北排森林公园']}]

In [27]: his_con_res = {k:([name_to_id[i] for i in v] if type(v)==list else {kk:[name_to_id[i] for i in vv if i in name_to_id.
his_con_res
4

Out[27]: {'洗太夫人诞': ['ID96', 'ID99', 'ID182'],
          '滨海旅游六点一线': ['ID90', 'ID164', 'ID274', 'ID117', 'ID116'],
          '五大特色龙头旅游': {'滨海旅游': ['ID116', 'ID145', 'ID456', 'ID189', 'ID328'],
          '洗夫人文化旅游': ['ID99', 'ID277', 'ID491'],
          '潘茂名养生旅游': ['ID314', 'ID136'],
          '美丽乡村风情游': ['ID446', 'ID371'],
          '山地运动休闲游': ['ID477', 'ID139', 'ID124', 'ID181', 'ID222']}

```

图 13 历史文化关联部分结果呈现图

4.3.4 关联度的计算和量化

在前一阶段的关联模式定义中，本文对八种关联模式进行了探索和定义，除去历史文化关联外，剩下的模式需要进行进一步的量化和标准化。根据表 7 中的计算方法进行模式的量化，之后利用 Z-Score 标准化方法和 Min-Max 标准化方法将指标数值局限至[0,1]，其中对部分集中在较小区间的数据取立方根。最终关联效果的部分展示如下表 8。

表 8 result3 部分数据展示

产品 1ID	产品 2ID	关联度	关联类型	产品名称
ID559	ID626	0.91	酒店风格相似	凯里亚德酒店(茂名高铁站店)-凯里亚德酒店(茂名电白万达广场店)
ID611	ID614	1	竞争关系	尚客优连锁酒店(信宜银湖路店)-信宜丽晶酒店
ID82	ID135	0.99	近邻关系	泰哥茶档-红旗街
ID94	ID96	0.87	共现关系	西城门-长坡旧城
ID274	ID5	0.97	辐射关系	中国第一滩旅游度假区-嘉嘉乐海鲜餐厅
ID116	ID458	0.87	导流关系	浪漫海岸国际旅游度假区-金沙湾海滨浴场
ID29	ID40	0.83	餐饮风格相似	优之品西点-麦壳西点(开发区店)

4.3.5 关联图谱存储

目前知识图谱的存储方式主要有关系型数据库、图数据库和基于 RDF 结构的存储方式。图数据库存储方式最大的优点在于,图结构即为知识图谱结构:图中的节点代表知识图谱的对象,图中的边代表知识图谱对象之间的关系[15]。Neo4j 图数据库以图形结构的形式存储信息,关联的数据本身就是它所包含的数据,因此它可以直接显示关联数据特征以及数据之间的关系。Neo4j 导入节点的数量大、速度快、支持查询算法,且相较于传统的关系型数据库,Neo4j 的遍历算法设计借助图结构自然伸展的特点,构建结果如图 14。

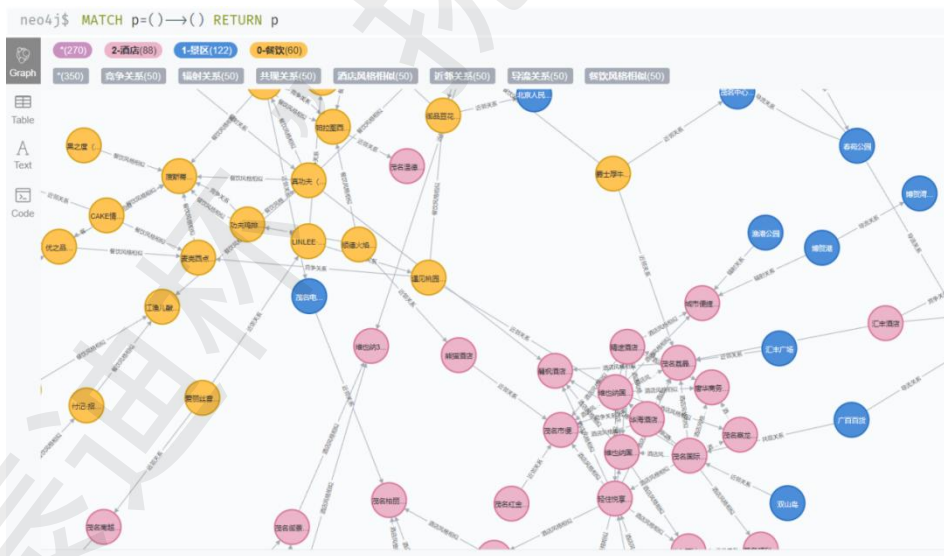


图 14 Neo4j 构建结果

4.3.6 关联图谱分析

关联图谱中主要包含餐饮(黄色)、景区(蓝色)和酒店(红色),由下图 15 局部图谱所示,共现关系出现频次最高,主要集中在景区;餐饮相对于其他

两者数量较少，且餐饮风格相似这一模式较为集中；酒店的竞争关系明显，例如鑫源酒店就与其他两个酒店只存在竞争关系。

此外，该关联模式中，两两结点中可能存在两者及以上的关联关系，例如维也纳国际酒店与华海酒店存在竞争关系、酒店风格相似等。辐射关系可以跨领域出现在餐饮和景区中，例如：茂名浪漫海餐厅与金沙湾海滨存在辐射关系。

该关联图谱分为疫情前和疫情后，下图只截取部分结点和关系进行展示，详细可见附录。

```
df_tra.loc[df_tra['year'].map(lambda x: 2018<=x<=2019), '出行天数'].mean()
4.940740740740741
```

```
df_tra.loc[df_tra['year'].map(lambda x: 2020<=x<=2021), '出行天数'].mean()
2.093220338983051
```

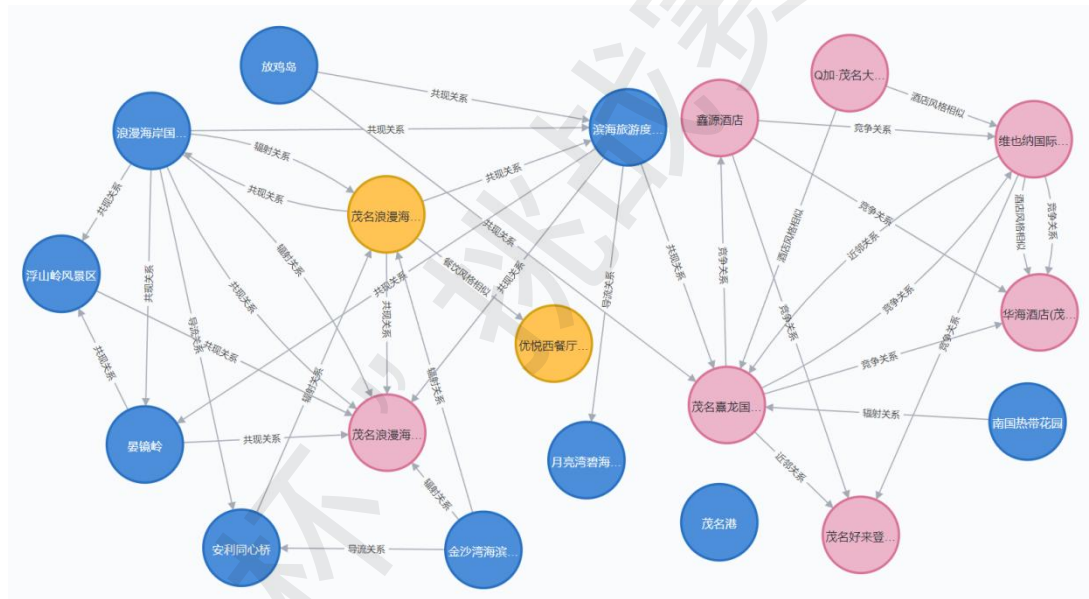


图 15 产品关联图谱（部分）

4.3 本地旅游产品宏观图谱构建及分析

为进一步对该地的旅游发展有宏观的认识和趋势了解，本文构建本地旅游宏观图谱，主要展示以茂名为中心，以餐饮、酒店、景区三大类辐射形成的产品图谱，其中详细展示了来源：住宿、攻略、景区评论等等。在该部分的展示中，仅展示热度最高的 5 个产品、词频最大的 8 个关键词；其中攻略，仅展示词频最大的 10 个关键词；此外，历史文化高层概念中，将这两年中共现的旅游产品都放

本章从旅游图谱在疫情前和疫情后的时间横向对比，和旅游产品在不同年度的热度纵向对比入手，分别进行分析。并且针对茂名市当地情况为管理部门撰写建议信，其中主要为提升旅游的综合效益，并且整合资源，打通旅游生态，助力消费升级；从平台视角解决 C 端游客、B 端商家和 G 端政府的共性和个性化需求。

5.2 旅游图谱对比分析

分析宏观图谱可得餐饮的热点产品有较大变化，说明疫情前后对餐饮业的打击影响较大；对于服务的相关评价也有较大变化[17]。

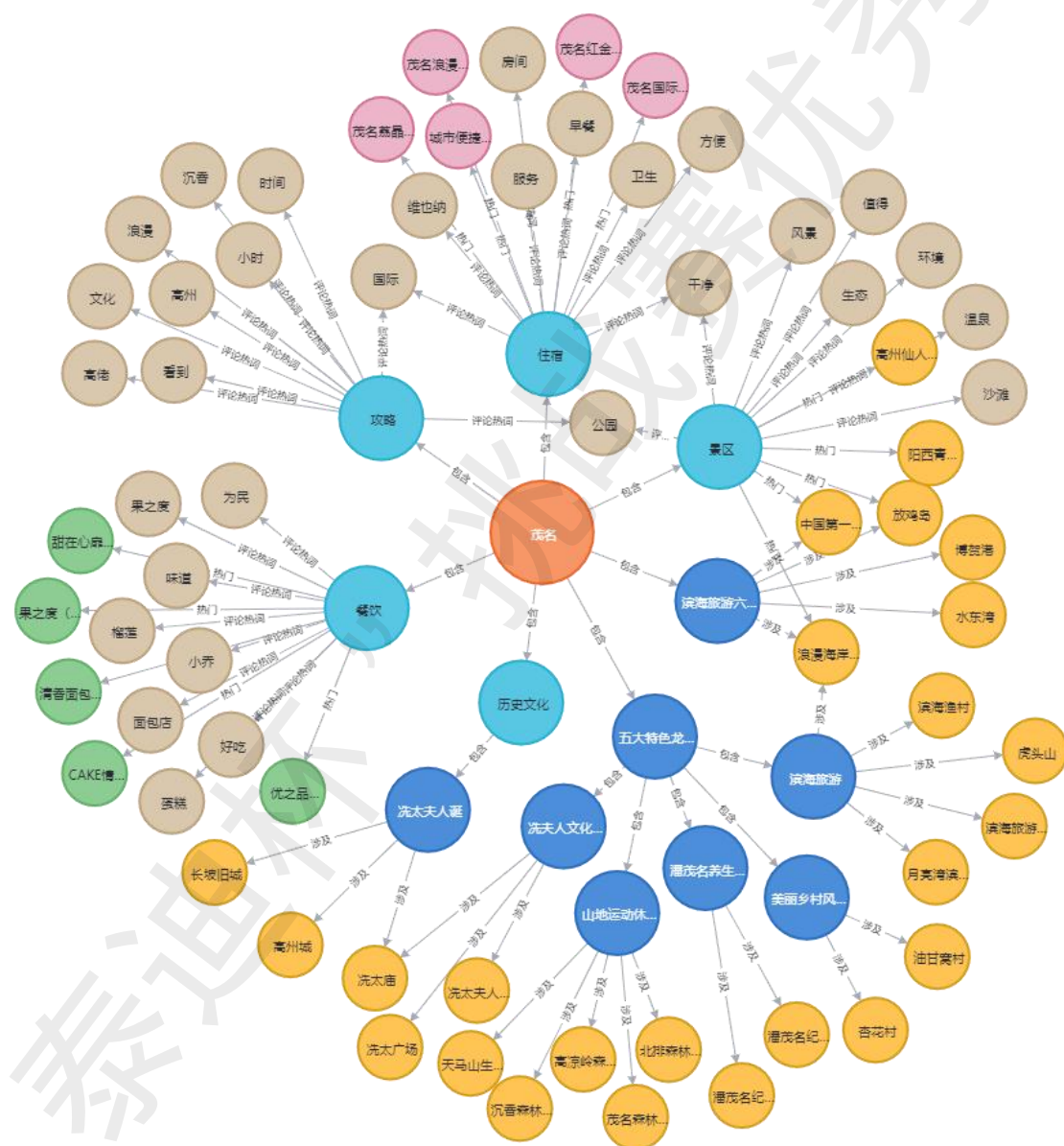


图 17 疫情前的宏观图谱

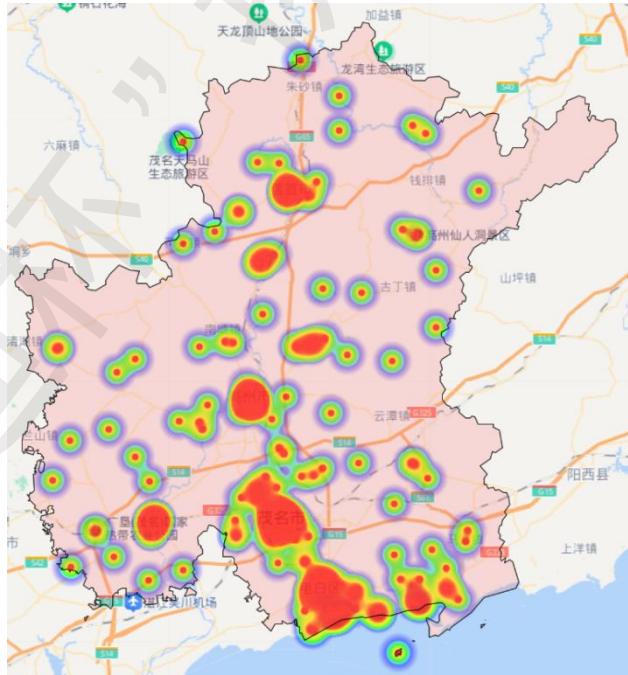
5.3 建议信

尊敬的管理部门人员：

您好！目前旅游业已然是现代服务业的重要组成部分，已经成为国民经济的战略性支柱产业，茂名市作为广东省农业大市，从黄金海岸线到北部大山岭，具备得天独厚的地理环境和丰富的乡村旅游资源，南海有渔村，北部有山村，中部有田园风光。优秀的

由于疫情原因，全球各地的旅游业都受到一定的影响，在这种情况下，本团队基于帮助改善旅游行业发展现状的初衷，通过挖掘茂名市旅游产品，分析其在疫情前和疫情后的热度对比、数量对比等多维度对比分析，形成了以产品关联模式和本市宏观整体分布等知识图谱，用清晰简洁的可视化表达形式，希望能给茂名当地管理人员提供一些发展思路。

首先本团队依靠提炼的旅游产品（疫情后）将其地理位置分布映射到高德地图中，形成如下图所示的产品分布热力图，其中越接近红色的区域代表产品聚集密度较高，该分布结果图显示：旅游产品总体较均匀分布在茂名市，但其中大多景区集中在茂名市的南部，你们可以将更多的治理资源和监管督查根据下图进行分布和调度。



此外，根据我们对宏观和微观两个层面制作的旅游图谱分析，疫情前后旅游产品的趋势如下：

(1) 游客的旅游倾向从“久游饱览式”过渡到“深挖全面式”。不确定的疫情因素导致大部分游客没有办法在某地游玩过久，相比于之前的旅游路线，游客会更加倾向于打卡式，将某地的所有热门经典均进行打卡。但是疫情后，大部分游客采用自驾游或者一至两天的游玩时间，故需要在某一两个景点花费整个旅程时间。图谱中的具体表现为：疫情前旅游景点更为分散，疫情后的暴涨式热门景点不完全集中。故在这种趋势下，需要当地管理部分具备较高质量和更高效率的统筹规划，对人流和设施资源进行规划。此外，为防止当地某旅游企业进行垄断式的引流增收，管理人员也应该对实时人流分布进行监管。为全面服务游客，管理部门还应该注重游客全场景和全过程的体验，开发全流程的智慧旅游平台，帮助游客在某一两个景点中也能获取“饱览式”的体验结果。

(2) 餐饮业迎来较好发展趋势和营销风口。从宏观的旅游产品图谱来看，疫情后的餐饮产品数量明显增多，并且大多集中在某一两个景区周围；该趋势说明游客对于旅游中的美食体验要求增加，并且更愿意尝试多样的美食类型和餐厅风格；在产品关联的旅游图谱中，餐厅风格类似的产品更多频次出现，说明游客对于同类型的餐厅也有更深的挖掘欲望。在此基础上，当地管理部门首先应该重视食品安全健康督查，防止出现任何食品安全问题；此外，由于旅游业向来有肆意定价的倾向，故管理部门需要进一步完善营销和定价制度，市场监督管理局也应对当地（尤其是新开）的餐厅进行更为严格的审查。

(3) 旅游产品的共现关系愈加明显。随着互联网和大数据平台的发展，尤其是各类短视频平台和社交 APP 的崛起，运用信息技术手段进行营销，打造本地旅游风格、突出文化特色已经成为了政府和商家全新的合作模式。通过抖音等 APP 发布打卡视频、宣传视频，从而引流至具体商家即将是旅游行业的新趋势。在此基础上，完善合理的营销策略至关重要，管理部门也可以在各大社交平台上开通官方账号进行统一宣传和通知；此外，利用官方账号进行预约、虚拟导览等业务也十分必要，从旅游路线到酒店预定，疫情背景下的透明化和预约进度制将助力当地旅游的快速发展。

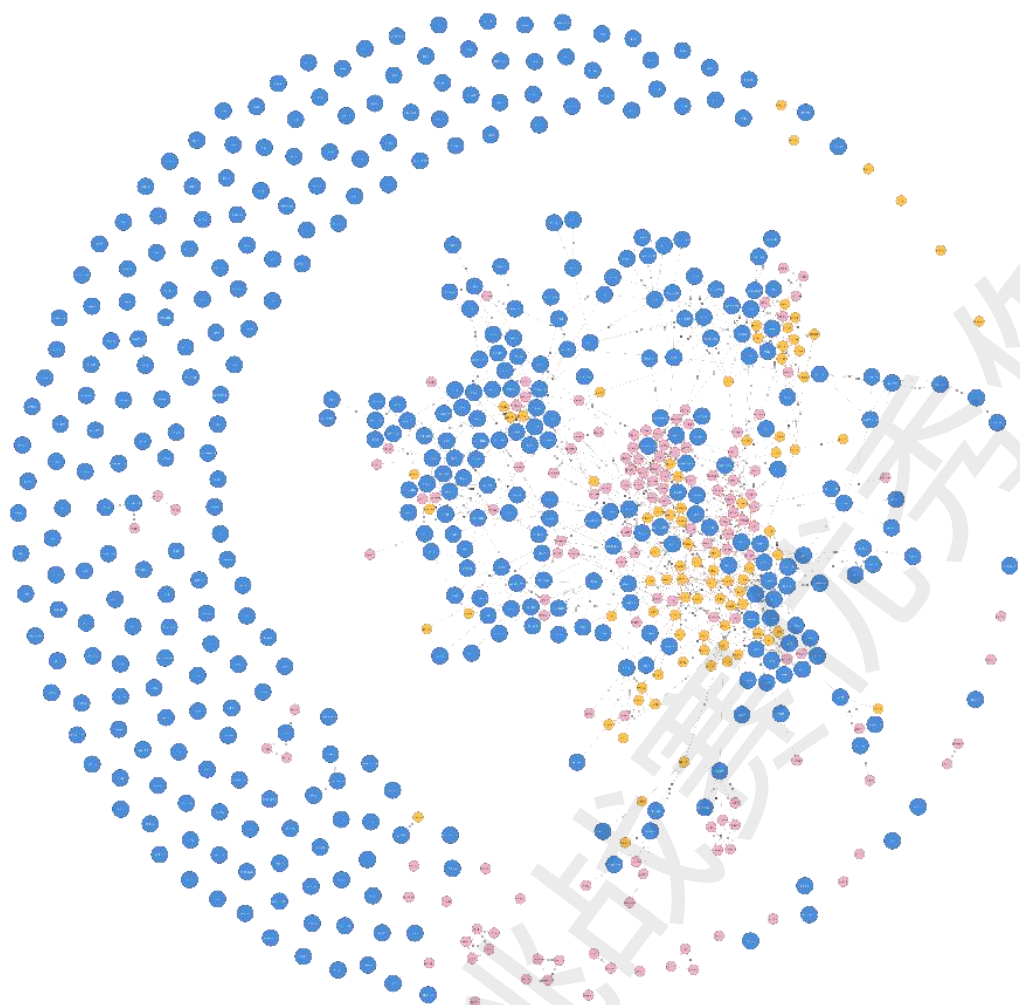
综上，茂名是个山清水秀的风水宝地，在此基础上，若管理部门能把握疫情前后的变化趋势，重点完善服务设施、打造平台影响力、提升用户粘性和活跃度，当地旅游业势必迎来更加美好的未来。以上是本团队的一些拙见！感谢你们！

参考文献

- [1] 丛昕,殷敏,丁绍刚,王小文,张清海.游客视角的中国古典园林景点热度感知评价与传播途径研究——以苏州古典园林留园为例[J].中国园林,2021,37(08):56-61.DOI:10.19775/j.cla.2021.08.0056.
- [2] 张春晖.旅游流与目的地耦合:模型与要素贡献度研究[D].陕西师范大学,2015.
- [3] 刘宇鹏,栗冬冬.基于 BLSTM-CNN-CRF 的中文命名实体识别方法[J].哈尔滨理工大学学报,2020, v.25(01):119-124.
- [4] 刘江舟,段立.基于 Albert-TextCNN 的网络安全文本多标签分类方法[J].舰船电子工程,2022,42(03):114-118+170.
- [5] 叶施仁,杨英,杨长春,朱明峰.孤立点预处理和 Single-Pass 聚类结合的微博话题检测方法[J].计算机应用研究,2016,33(08):2294-2297.]
- [6] 张源淇.影响网络舆情热度评价的识别因素探讨[J].新闻研究导刊,2022,13(01):127-129.
- [7] 张菁芳,李佳承,陈俊国,薛晓,任家顺.基于层次分析法的医院财务信息化绩效评价指标体系研究[J].西南大学学报(自然科学版),2017,39(02):73-83.
- [8] 李磊.文本型 Web 内容中隐含实体关联的挖掘、推理与应用[D].云南大学,2019.
- [9] 黄秀娟,刘伟平.森林公园旅游产品项目开发的影响因素——外部关联性及产品项目竞争力要素分析[J].绿色中国,2004(06):59-60.
- [10] 姚占雷,许鑫,李丽梅,杨勇.sss 网络游记中的景区共现现象分析——以华东地区首批国家 5A 级旅游景区为例[J].旅游科学,2011,v.25;No.123(02):39-46+72.DOI:10.16323/j.cnki.lykx.2011.02.008.1
- [11] 方远平,谢蔓,毕斗斗,肖佑兴.sss 中国入境旅游的空间关联特征及其影响因素探析——基于地理加权回归的视角[J].旅游科学,2014,v.28;No.142(03):22-35.DOI:10.16323/j.cnki.lykx.2014.03.005.
- [12] 暴向平,薛东前,李庆雷,郭瑞斌.陕西省城市旅游经济辐射场强格局演变研究

- [J].干旱区资源与环境,2014,28(08):183-188.DOI:10.13448/j.cnki.jalre.2014.08.033.
- [13]刘中艳,罗琼.省域城市旅游竞争力测度与评价——以湖南省为例[J].经济地理,2015,35(04):186-192.DOI:10.15957/j.cnki.jjdl.2015.04.026.
- [14]胡丁婷.桂林抗战文化遗产旅游开发价值评价研究[D].广西大学,2020.
- [15]张吉祥,张祥森,武长旭,赵增顺.知识图谱构建技术综述[J].计算机工程,2022,48(03):23-37.
- [16]侯颖,张海洲,陆林.sss 现象学如何提升旅游体验的经验研究? ——融合手段-目的链方法的尝试[J].旅游科学,2022,v.36;No.188(01):73-91.
- [17]蔡卫民,彭晶,覃娟娟.sss 韶山的全国网络关注热度矩阵及推广策略研究[J].旅游科学,2016,v.30;No.155(04):61-72.

附件:



“泰迪杯”挑战数学竞赛作品

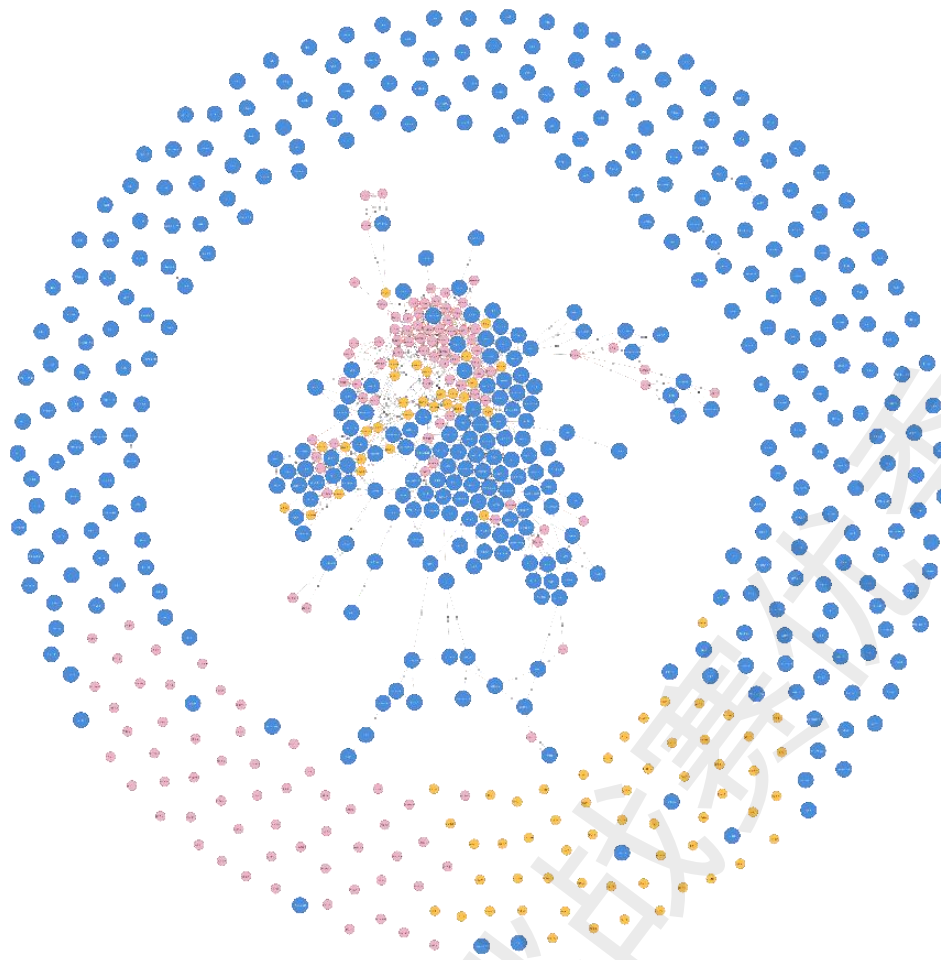


图 1 疫情前后全关联模式