旅游地理情感评价模型在粤港澳大湾区应用模拟数据集（2008–2021）

刘 逸[1,2]，陈海龙[1]，肖文杰[1,3*]，保继刚[1]，吴雪涵[1]，徐佳莉[1]

1. 中山大学旅游学院，广州 510275；2. 旅游可持续智能测评技术文化和旅游部重点实验室，珠海 519080；
3. 吉首大学旅游学院，张家界 427000

摘 要：旅游目的地情感评价可为目的地管理提供决策依据。本研究通过自建游客情感词典、构建语义规则和校验情感系数，研发了旅游情感评价（TSE）模型和应用平台，为旅游目的地评价提供了新方法与新工具。以猫途鹰、马蜂窝和携程网为数据源，采集 2008–2021 年粤港澳大湾区 11 个城市的评论，利用平台获得城市情感评价数据集。数据集由 15 个文件组成，包括：（1）11 个城市关注度排名；（2）11 个城市美誉度排名；（3）11 个城市关注度与美誉度排名差异；（4）大湾区整体情感画像；（5）香港情感画像；（6）澳门情感画像；（7）深圳情感画像；（8）广州情感画像；（9）珠海情感画像；（10）佛山情感画像；（11）惠州情感画像；（12）东莞情感画像；（13）中山情感画像；（14）江门情感画像；（15）肇庆情感画像。数据集存储为.xlsx 格式，数据量为 34 KB。数据集表明：（1）以 2016 年为分界点，关注度和美誉度排名由混乱趋于稳定，2016 年前香港位居前列，后期下降；近年广州位居前列，江门和中山位置靠后；（2）游玩体验和观光体验是游客关注要点，基础设施和配套设施的作用也不容忽视。

关键词：情感评价；旅游目的地；TSE 模型；美誉度

DOI: https://doi.org/10.3974/geodp.2023.01.14

CSTR: https://cstr.escience.org.cn/CSTR:20146.14.2023.01.14

数据可用性声明：

本文关联实体数据集已在《全球变化数据仓储电子杂志（中英文）》出版，可获取：https://doi.org/10.3974/geodb.2023.05.06.V1 或 https://cstr.escience.org.cn/CSTR:20146.11.2023. 05.06.V1.

1 前言

旅游目的地评价有利于揭示发展质量与竞争格局[1]，其结果受到各界关注[2]。传统评价模型主要以调查问卷为数据源，效率低下[3]。智能设备生产的海量评论为评价旅游目的地提供新数据。但是如何从海量评论挖掘出游客的偏好等信息是获得游客对旅游目的地的整体评价的关键难题[4]。为此，研究团队基于情感分类理论和词汇匹配技术研发旅游情感

收稿日期：2023-01-06；**修订日期**：2023-03-20；**出版日期**：2023-03-25

***通讯作者**：肖文杰,中山大学旅游学院;吉首大学旅游学院, xiaowj7@mail2.sysu.edu.cn

数据引用方式：[1] 刘逸，陈海龙，肖文杰等. 旅游地理情感评价模型在粤港澳大湾区应用模拟数据集（2008–2021）[J]. 全球变化数据学报, 2023, 7(1): 102–107. https://doi.org/10.3974/geodp.2023.01.14. https://cstr.escience.org.cn/CSTR:20146.14.20223.01.14.

[2] 刘逸，陈海龙，肖文杰等. 旅游地理情感评价模型及其在粤港澳大湾区城市情感评价中的应用模拟数据集（2008–2021）[J/DB/OL]. 全球变化数据仓储电子杂志，2023. https://doi.org/10.3974/geodb.2023.05.06.V1. https://cstr.escience.org.cn/CSTR:20146.11.2023.05.06.V1.

评价 TSE 模型应用平台，开发粤港澳大湾区（以下称大湾区）城市情感评价数据集。模型具有较好的信度[4]和较高的准确性[5]，已被用于旅游目的地评价[4]、城市旅游形象捕捉[6]、旅游市场空间结构测度[7]。不仅如此，TSE 模型应用平台还适用于旅游时空行为捕捉和人地互动机制发现等多个地理应用场境。

2　数据集元数据简介

《旅游地理情感评价模型在粤港澳大湾区城市情感评价中的应用模拟数据集（2008–2021）》的名称、作者、地理区域、数据年代、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

表 1　《旅游地理情感评价模型在粤港澳大湾区城市情感评价中的应用模拟数据集（2008–2021）》元数据简表[8]

条　目	描　述
数据集名称	旅游地理情感评价模型在粤港澳大湾区城市情感评价中的应用模拟数据集（2008–2021）
数据集短名	DataSenEvaCitiesGBA_2008-2021
作者信息	刘逸,中山大学旅游学院;旅游可持续智能测评技术文化和旅游部重点实验室, liuyi89@mail.sysu.edu.cn 陈海龙, 中山大学旅游学院, chenhlong5@mail2.sysu.edu.cn 肖文杰, 中山大学旅游学院, 吉首大学旅游学院,xiaowj7@mail2.sysu.edu.cn 保继刚, 中山大学旅游学院, eesbjg@mail.sysu.edu.cn 吴雪涵, 中山大学旅游学院, wuxh68@mail2.sysu.edu.cn 徐佳莉, 中山大学旅游学院, xujli3@mail2.sysu.edu.cn
地理区域	香港、澳门、广州、深圳、珠海、佛山、惠州、东莞、中山、江门、肇庆
数据年代	2008–2021
数据格式	.xlsx
数据量	34 KB
数据集组成	由 4 个部分共 15 个文件组成：(1) 11 个城市关注度排名；(2) 11 个城市美誉度排名；(3) 11 个城市关注度与美誉度排名差异；(4) 情感画像，包括大湾区和 11 个城市的情感画像（城市情感画像指正面、负面和中性评论比例）
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、通过《全球变化数据仓储电子杂志（中英文）》发表的实体数据集和通过《全球变化数据学报（中英文）》发表的数据论文。其共享政策如下：(1)“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；(2) 最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；(3) 增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；(4) 摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源[9]
数据和论文检索系统	DOI，CSTR，Crossref，DCI，CSCD，CNKI，SciEngine，WDS/ISC，GEOSS

3 TSE 模型应用平台

3.1 概述

平台以 TSE 模型为核心，以大数据评论为数据源，适用于旅游目的地评价等多个地理应用场境。

3.2 TSE 模型的构建

TSE 模型构建包括三个步骤[4]。

（1）自建游客情感词库。通过人工深度阅读的方式从旅游攻略和旅游评论中提取表达游客情感的高频情感词汇，并通过与 CNKI（知网）发布的 HowNet 词典合并，最终构建了包括 3,507 个正面词汇和 3,365 个负面词汇的游客情感词库。

（2）构建语义逻辑规则。考虑程度副词、否定词、转折词在情感倾向方面的作用，构建了以三者的组合为句式的 32 条语义规则，详见文献[4]。

（3）校正情感系数。利用长达十年的世界旅游组织开展的问卷调查数据进行校验，发现情感校正系数应该为 4，即只有正面词汇数量超过或等于负面词汇 4 倍时才被判定为正面评论。

3.3 主要功能

平台包括情感计算和共现分析两大功能。情感计算主要采用语法模型计算情感得分，算法如下[5]。

当 $\left|\left(g_{dp}\times\frac{P}{e}-g_{dn}\times N\right)\right|\geq 1$ 且 g_a 为第一类转折词时，采用式(1)计算评论情感得分。

$$\gamma=-1^{g_n+g_a}\times\frac{g_{dp}\times\frac{P}{e}-g_{dn}\times N}{\left|\left(g_{dp}\times\frac{P}{e}-g_{dn}\times N\right)\right|} \tag{1}$$

当 $\left|\left(g_{dp}\times\frac{P}{e}-g_{dn}\times N\right)\right|\geq 1$ 且 g_a 为第二类转折词时，采用式(2)计算评论情感得分。

$$\gamma=-1^{g_n+g_a+1}\times\frac{g_{dp}\times\frac{P}{e}-g_{dn}\times N}{\left|\left(g_{dp}\times\frac{P}{e}-g_{dn}\times N\right)\right|} \tag{2}$$

当 $\left|\left(g_{dp}\times\frac{P}{e}-g_{dn}\times N\right)\right|<1$，采用公式（3）计算评论情感得分。

$$\gamma=0 \tag{3}$$

式中，γ 为评论的情感得分，包括 1（正面）、–1（负面）、0（中性）三种结果，g_n 为否定副词的数量，g_a 为转折连词的数量，g_{dp} 为正面词前的程度副词数量，g_{dn} 为负面词前的程度副词数量，P 为正面词数量，N 为负面词数量，e 为情感乘数。

共现功能则根据用户提供的关键词生成关键词共现矩阵和相邻矩阵。

4　数据研发

4.1　数据抓取

以知名度、评论丰富性、用户活跃度、评论长度为筛选标准，选择猫途鹰、马蜂窝和携程网三个旅游网站作为数据来源。通过 Python 抓取 11 个城市 2008–2021 年的旅游评论，字段包括评论时间、评论内容、评论得分等。

4.2　关注度和美誉度计算

运用情感计算功能，获得 2008–2021 年 11 个城市每条评论的情感分类结果（负面、中性或正面），以评论数量为关注度、以正面评论比例为美誉度，按年份统计旅游评论情感分类结果，得到 11 个城市关注度、美誉度以及排名差异。

4.3　情感画像

根据情感计算结果，分别以 11 个城市的正面评论和负面评论为数据源，利用高频词分析功能，获得正面评价和负面评价的高频词，选择排名前 200 的高频词作为关键词文件，应用共现分析功能生成相邻矩阵（高频词之间的关系矩阵），利用 Gephi 软件绘制各年份 11 个城市情感画像。

5　数据结果与验证

5.1　数据集组成

《旅游地理情感评价模型在粤港澳大湾区城市情感评价中的应用模拟数据集（2008—2021）》存储为.xlsx 格式，包括 15 个统计表，分别是：（1）11 个城市关注度排名；（2）11 个城市美誉度排名；（3）11 个城市关注度与美誉度排名差异；（4）大湾区整体情感意向；（5）香港情感画像；（6）澳门情感画像；（7）深圳情感画像；（8）广州情感画像；（9）珠海情感画像；（10）佛山情感画像；（11）惠州情感画像；（12）东莞情感画像；（13）中山情感画像；（14）江门情感画像；（15）肇庆情感画像。

5.2　数据结果

图 1 和图 2 分别报告了 2008–2021 年 11 个城市关注度和美誉度变化。

关注度方面，11 个城市在 2016 年之前排名波动较大，但 2016 年及之后整体趋于稳定，广州、深圳、佛山、惠州、肇庆、中山、江门 7 个城市的排名自 2014 年之后较为平稳，广州稳居高位，肇庆、中山、江门始终保持较低水平，香港和澳门下降明显。

美誉度方面，11 个城市 14 年间波动上升。前 6 年剧烈波动，后 8 年稳中有升。除江门、惠州外的 9 个城市在 2018 年及以后均高于 0.80，澳门保持相对稳定的高位，香港、中山相对稳定。

情感画像方面，对广州 2008–2021 年间正负面评价的情感画像进行分析。

由图 3 可知，广州游客正面评价集中于旅游吸引物和旅游基础设施。整体而言，城市美景、良好的生态环境与岭南首府的千年文化底蕴共同突出现代与历史相融的花城形象；负面评价则集中反映了对旅游吸引物开发以及旅游配套设施的不满，存在夜游游船较为拥挤、娱乐景点项目单调、环境卫生状况堪忧、景区饭菜贵且难吃、停车难收费贵等问题。

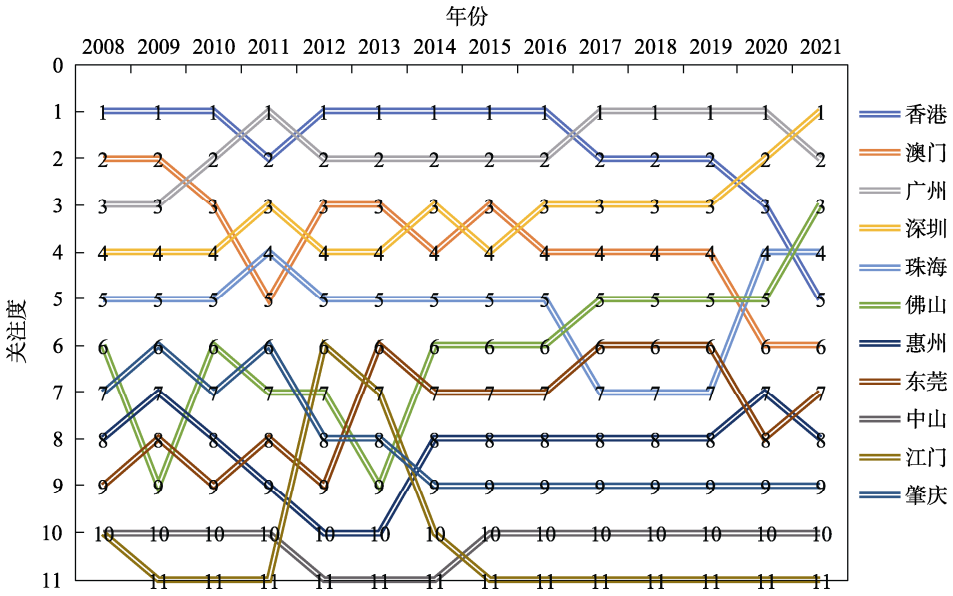

图 1　2008–2021 年大湾区城市关注度排名变化图[10]

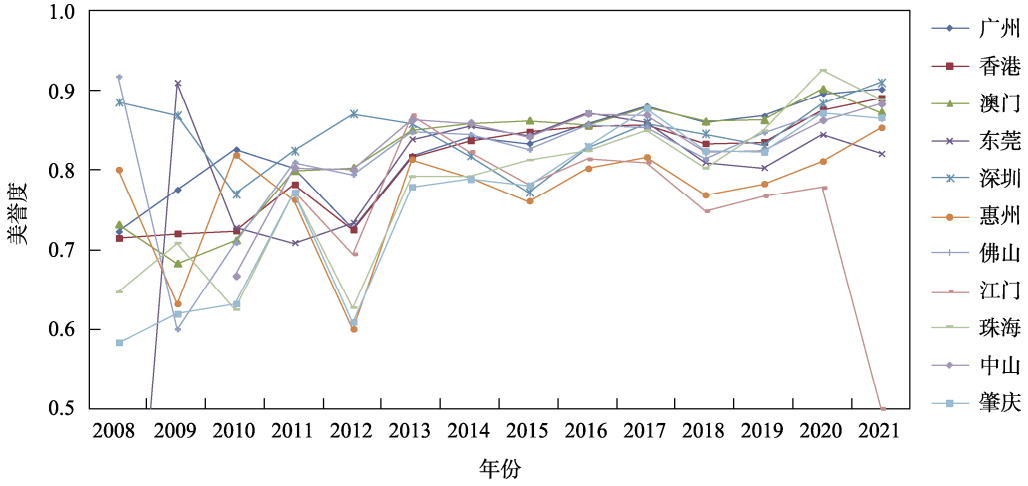

图 2　2008–2021 年大湾区城市美誉度变化图[6, 10]

5.3　数据验证

利用长达十年的问卷调查数据进行校验，证实模型具有较好的信度，详细结果见文献[4]。与 6 个机器学习模型对比，TSE 模型具有稳定的精确度，详细结果见文献[5]。

6　讨论和总结

《旅游地理情感评价模型在粤港澳大湾区城市情感评价中的应用模拟数据集（2008–2021）》是对大湾区城市影响力、美誉度的宏观描述，有效揭示了 11 个城市的竞争格局以及游客关注要点，可为目的地管理提供决策依据。

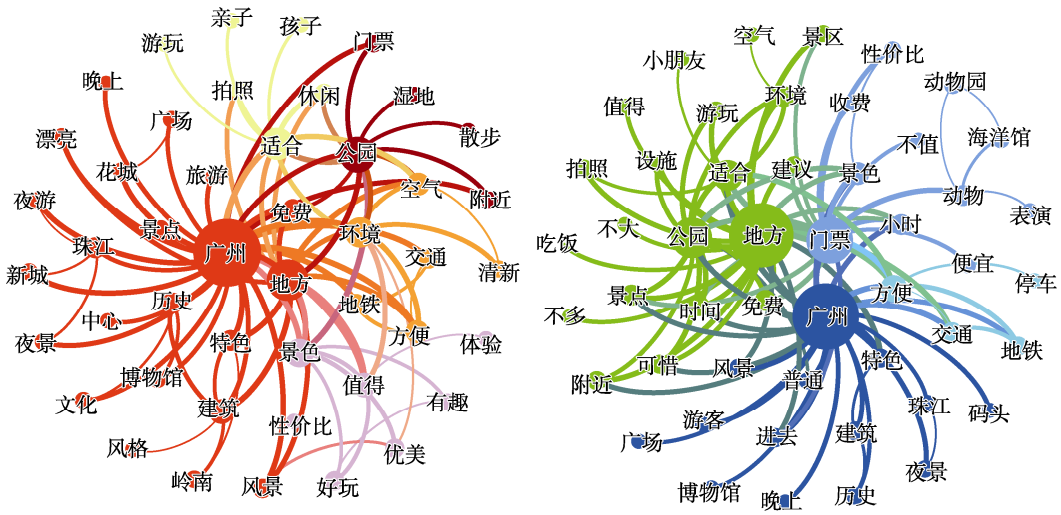

图 3　广州游客正面和负面评价情感意象图[10]

作者分工：刘逸、陈海龙、肖文杰、保继刚对数据集的开发做了总体设计；吴雪涵、徐佳莉采集和处理了原始评论和美誉度等数据；刘逸、肖文杰设计了模型和算法；肖文杰做了数据验证；陈海龙、肖文杰、吴雪涵撰写了数据论文；刘逸、保继刚参与论文修改。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

[1] 龚箭, 杨舒悦. 基于网络评论的旅游目的地评价研究——以我国 31 个省市自治区为例[J]. 华中师范大学学报(自然科学版), 2018, 52(2): 279–286.

[2] 史春云, 张捷, 沈正平等. 区域旅游竞合研究进展[J]. 地理与地理信息科学, 2005(5): 85–89.

[3] 甄峰, 王波. “大数据”热潮下人文地理学研究的再思考[J]. 地理研究, 2015, 34(5): 803–811.

[4] 刘逸, 保继刚, 朱毅玲. 基于大数据的旅游目的地情感评价方法探究[J]. 地理研究, 2017, 36(6): 1091–1105.

[5] 刘逸, 孟令坤, 保继刚等. 人工计算模型与机器学习模型的情感捕捉效度比较研究——以旅游评论数据为例[J]. 南开管理评论, 2021, 24(5): 63–74.

[6] 刘逸, 徐晓静, 赵莹. 基于 TSE 和 IPA 模型的城市旅游形象感知和差异研究——以广东省为例[J]. 旅游论坛, 2019, 12(6): 41–49.

[7] 刘逸, 陈海龙, 曹轶涵. 基于旅游评论数据的旅游地客源市场空间结构探析与理论修正[J]. 世界地理研究, 2023, 32(5): 113–124.

[8] 刘逸, 陈海龙, 肖文杰等. 旅游地理情感评价模型及其在粤港澳大湾区城市情感评价中的应用模拟数据集（2008–2021）[J/DB/OL]. 全球变化数据仓储电子杂志, 2023. https://doi.org/10.3974/geodb.2023.05.06.V1. https://cstr.escience.org.cn/CSTR:20146.11.2023.05.06.V1.

[9] 全球变化科学出版系统. 全球变化科学研究数据共享政策[OL]. https://doi.org/10.3974/dp.policy.2014.05 (2017 年更新).

[10] 刘逸, 林雪颖, 张童等. 2021 年基于大数据的粤港澳大湾区游客情感意象研究报告[M]//徐红罡, 保继刚. 粤港澳大湾区蓝皮书: 粤港澳大湾区旅游业发展报告（2021）. 北京: 社会科学文献出版社, 2022: 168–210.