

# 旅游形象感知研究

## 一、引言

旅游形象感知研究是对旅游景点在大众点评和携程等平台上用户评论进行爬虫获取并进行数据处理、情感分析和可视化展示的研究。本论文主要关注了大众点评和携程两个平台对于布达拉宫和大昭寺两个旅游景点的用户评论，旨在通过分析用户在不同平台对这两个景点的印象和态度，探讨用户画像的特点。

## 二、数据爬取

1. 在大众点评上通过API获取数据：首先通过大众点评网站获取请求参数，获取店铺ID和相关信息，通过start\_search函数判断店铺是否重复，获取评论相关字段，将数据保存为Excel文件。
2. 在携程上通过API获取数据：通过请求Headers和Cookie获取地点ID和名称，调用API获取评论的JSON内容，根据字段定位获取对应文本内容，获取全部评论，保存为Excel文件。

## 三、数据处理

1. 为了处理文本数据，我们可以使用pandas、re、jieba和snownlp库进行数据处理。首先，使用re.sub函数进行数据处理和机械压缩，去除多余的词和无效的标点符号，以及影响情感划分的表情包等数据。然后，利用jieba库中的pseg.cut函数对处理好的文本进行分词。在分词过程中，我们可以根据字的长度、是否为中文数据以及是否存在于停用词文本中进行判断。满足条件的词会被保存下来。接下来，对去重后的分词进行处理，并删除内容为空集的行。

在处理完上述步骤后，我们可以使用snownlp库中的sentiments函数对文本进行情感打分。打分结果大于0.5可以判断为积极态度，小于0.5则为消极态度，其余的为中立态度。

最后，根据处理过的数据进行统计，我们发现，在携程网上有3136条布达拉宫的数据和2994条大昭寺的数据经过处理。而在大众点评上，布达拉宫的数据为9965条，大昭寺的数据为4848条。

通过这些处理步骤，我们可以得到更干净、准确的文本数据，为后续的情感划分和分析提供更可靠的基础。这种分析方法可以帮助我们深入了解用户对不同景点的情感表达，从而指导景点管理和营销策略的优化，提升用户满意度和吸引力。

总结而言，使用pandas、re、jieba和snownlp库对文本数据进行处理和分析，可以帮助我们清洗并提取有用信息。通过这些处理步骤，可以为后续的情感划分和分析提供准确的数据基础，以便更好地了解用户的情感倾向和态度。

## 四、数据分析和可视化

1. 情感分析结果：通过情感分析对用户评论数据进行处理后，我们可以获得景点对应的积极、消极和中立情感数据。为了更好地了解用户对不同景点的情感倾向，我们可以利用pandas中的value\_counts函数对情感数据进行统计，并采用matplotlib库中的Pie函数制作饼图进行可视化展示。在绘制饼图前，需要使用rcParams函数对中文字体进行配置，以保证中文显示。同时，可以使用pie函数的autopct参数将百分比设置为两位小数，并利用legend函数添加图例。

对于布达拉宫，我们在大众点评上得到的情感数据是77.46%积极、22.53%消极和0.01%中立。而在携程网站上，积极态度为78.22%，消极为21.30%，中立为0.48%。在对这些数据进行比较时，我们可以看到在两个网站上，用户的积极情感占据了主导地位，而消极情感占比则较小，中立情感几乎可以忽略不计。

接下来，我们再来看一下大昭寺的情况。在大众点评上，大昭寺的情感数据显示出81.50%的积极态度、18.50%的消极态度和0.00%的中立态度。而在携程网站上，积极态度为83.23%，消极态度为16.60%，中立态度为0.17%。与布达拉宫相比，我们可以看到大昭寺的积极情感比例更高，消极情感比例更低，中立情感比例也更接近于0。

通过这些分析，我们可以更好地了解用户对于不同景点的情感倾向。这些洞察可以被景点管理者和从业者用来改进景点的服务和营销策略，以提高用户满意度和吸引力。同时，这些分析方法也可以为旅游行业从业者和其他研究人员提供参考，以揭示用户对景点的态度和情感偏好，并进一步推动行业的发展和进步。

总而言之，通过情感分析和饼图的可视化展示，我们可以更好地了解用户对不同景点的情感倾向。这些分析结果可以为景点的管理和发展提供有价值的见解，指导相关从业者制定更加精准和有效的业务决策。

2. 发帖趋势分析：为了了解每年发布帖子的频率以及不同地区的热度变化，我们可以利用pandas库中的to\_datetime函数对发帖的时间进行处理，将时间数据类型转换为datetime。接下来，我们可以使用pandas库中的resample函数，并选择"A-DEC"来统计每年的发帖总数。最后，可以使用matplotlib库中的条形图(bar)对数据进行可视化处理，以观察发帖趋势的变化情况。

首先，通过应用to\_datetime函数将评论数据中的时间字符串转换为datetime类型，方便后续的时间序列分析。接着，使用resample函数按年对数据进行采样和聚合，在每个年份计算发帖总数。

通过在matplotlib中使用条形图(bar)，我们可以将每年的发帖数量进行可视化展示。横轴表示年份，纵轴表示发帖数量。可以根据条形的高度来观察发帖数量的变化趋势。对于不同地区，可以分别生成相应的图表，以比较不同地区的热度变化。

根据我们的分析结果，可以得出以下结论：在大众点评上，布达拉宫的发帖趋势是逐年上涨的。特别是在2018年，布达拉宫的发帖数量达到了历史最高峰。之后，趋势开始下滑，但在2022年再次出现了第二次高峰。而在携程网上，布达拉宫的发帖数量则在2023年达到了历史最高峰。

这些分析结果可以帮助景点管理者和从业者更好地了解用户对不同景点的关注程度和热度变化。通过观察发帖趋势，他们可以及时了解用户需求的变化，根据数据指导业务决策并优化服务和推广策略。

总结而言，通过使用pandas中的to\_datetime函数对时间数据进行处理和转换，并结合resample函数对每年发帖频率进行统计，我们可以使用matplotlib绘制条形图对数据进行可视化展示。通过观察发帖趋势的变化，我们可以更好地了解不同地区的热度变化，并根据结果指导旅游行业的发展和经营决策。

3. 用户画像分析：为了更好地了解不同景点在不同平台上的用户画像，可以利用词云图进行可视化展示。通过使用stylecloud技术，我们可以提取用户评论中的高频词，并将其以词云的形式呈现出来，以便更直观地观察和比较用户关注的重点。

词云图是一种以词语为单位，用不同大小、颜色、布局的方式展示特定文本数据中词语重要程度的图形化工具。通常，词云中的词语出现频率越高，它们在图中的显示位置和大小就越突出。因此，通过词云图我们可以快速了解用户在不同平台、不同地区对不同景点的关注点和兴趣。

具体而言，在进行词云图的制作时，我们可以选取用户评论数据中的前100个高频词。这些高频词可能涵盖了用户对景点的印象、体验、特色等方面的关键词。通过分别制作不同平台和不同地区的词云图，可以对比不同群体的关注点差异，进一步了解用户的喜好和期望。

通过这种可视化展示方式，我们可以直观地观察到不同平台和地区用户的关注点，并从中提取有价值的洞察。这些洞察将有助于景点管理者和从业者了解用户需求，制定更精准的营销和服务策略，提升用户体验，优化景点的形象。并且，这种可视化分析方法也可以为未来的研究提供参考，洞察用户对不同景点的态度和兴趣，推动旅游行业的发展和创新。

总而言之，利用词云图进行用户画像的可视化展示是一种有效的方式，可以帮助我们了解不同景点在不同平台上的用户关注点和偏好。这种分析方法能够提供有关用户需求和期望的重要见解，对景点的管理和服务提供有价值的指导。

## 五、结论

通过对大众点评和携程上的用户评论进行爬取、数据处理和分析，对布达拉宫和大昭寺的旅游形象感知进行了研究。根据情感分析结果和发帖趋势分析，得出不同平台上用户对这两个景点的态度和热度变化。通过词云图可视化分析不同平台和地区的用户画像，发现用户关注的点有所不同。

最后，本研究的研究结果对于景点管理者和旅游行业从业人员提供了重要的参考和指导，帮助他们更好地了解用户对旅游景点的感知和态度，进一步改进景点的形象与服务，提升用户的旅游体验。

## 六、讨论

在本研究中，我们通过对大众点评和携程两个平台上用户对布达拉宫和大昭寺的评论进行爬取和情感分析，得出了一些有趣的结果。然而，我们也要意识到研究存在一些局限性和可以改进的地方。

首先，本研究仅针对了大众点评和携程这两个平台上的评论数据，而其他社交媒体平台和旅行指南上的评论也可能对形象感知产生影响。因此，进一步研究可以考虑包括更多的数据来源，以获取更全面、多样化的用户评论数据。

其次，情感分析是基于文本内容进行的，但用户对景点的主观感受可能无法完全通过文字表达。由于缺乏直接面对面的沟通，文本中可能存在着隐含的语义和情感信息。因此，结合其他技术手段，如图像识别、声音分析等，可以更全面地捕捉用户的情感和感受。

另外，本研究的用户画像分析仅仅基于高频词的展示，而没有分析不同用户群体之间的差异。未来的研究可以利用用户标签数据、用户行为分析等方法，对用户画像进行更深入的探究，挖掘出更多的用户行为特征。

此外，由于本研究仅针对布达拉宫和大昭寺两个景点进行了形象感知研究，因此研究结果的推广范围有一定限制。如果希望更全面地了解旅游景点的形象感知，可以扩大研究范围，包括更多、不同类型的景点进行比较分析。

## 七、结论

本研究通过对大众点评和携程平台上用户评论数据的爬取和分析，对布达拉宫和大昭寺的旅游形象感知进行了研究。通过情感分析，我们了解了用户对这两个景点的态度和情感倾向，通过发帖趋势分析，我们观察到用户对这两个景点的热度变化。通过用户画像分析，我们发现不同平台和地区用户关注的点有所不同。

这些研究结果对旅游目的地管理者和从业者具有重要的指导意义，可以帮助他们更好地了解用户对景点的感知和态度，并根据用户需求进行服务和策划的优化。同时，本研究也为未来类似研究提供了一些方法和启示，如扩大数据来源，结合多模态分析等。

在未来的研究中，可以考虑改进本研究的局限性，如扩大数据来源、引入多模态分析等，以获取更全面、准确的信息。此外，还可以开展更多旅游形象感知研究，深入探讨不同景点和用户群体之间的差异，从而提供更有针对性的服务和管理策略。

## 八、建议和未来方向

基于本研究的结果和讨论，我们提出以下几个建议和未来研究方向：

1. 多数据源的综合分析：除了大众点评和携程这两个平台外，还可以考虑其他社交媒体平台、旅行博客、微博等的用户评论数据。通过综合分析多个数据源，可以更全面地了解用户对旅游景点的形象感知。
2. 考虑用户群体的差异：不同的用户群体可能对景点的形象有不同的看法和期望。因此，未来的研究可以针对不同的用户群体进行分析，比如不同地区、不同年龄段、不同旅游偏好等用户特征。
3. 深入挖掘用户行为和需求：除了情感分析外，还可以结合用户的浏览行为、点击率、收藏等数据，分析用户对景点形象感知的影响因素，判断用户的偏好和需求，为景点管理者提供更具针对性的策划和服务。
4. 应用机器学习和自然语言处理技术：通过应用机器学习和自然语言处理技术，可以提高情感分析的准确性和效率，使分析结果更具可靠性。此外，还可以利用这些技术来自动识别用户评论中的关键词和主题，进一步挖掘用户对景点的态度和需求。

5. 结合在线调查和访谈：为了更全面地了解用户对旅游景点的形象感知，可以结合在线调查和访谈等定性研究方法。通过与用户进行深入交流，可以获取到更丰富、具体的意见和见解，从而更好地理解用户的需求和期望。

综上所述，旅游形象感知研究是一个复杂而有意义的领域。通过在大众点评和携程平台上的用户评论数据的爬取、处理和分析，我们可以揭示用户对旅游景点的态度和需求，并为景点的管理和运营提供参考和指导。未来的研究可以继续拓展数据源、分析更多的因素，并结合定量和定性研究方法，以提升研究的全面性和准确性。

## 九、数据隐私与伦理问题

在进行用户评论数据的爬取和分析时，需要注意数据隐私和伦理问题。以下是一些需要考虑的方面：

1. 数据隐私保护：用户评论数据涉及个人的意见和评价，可能包含敏感的个人信息。在处理这些数据时，需要确保用户的隐私得到保护，采取安全措施，如数据匿名化、加密和存储保护等。
2. 合法获取数据：在进行用户评论数据的爬取时，需要确保获取数据的合法性。需要遵守平台的使用条款和政策，并尽量避免侵犯用户的权益。
3. 知情同意和匿名处理：如果研究中使用了用户评论数据，应该在适当的情况下征得用户的知情同意。并且，对于用户的身份和个人信息，需要进行匿名处理和保护，以保障用户的隐私。
4. 伦理审查：在进行涉及人类参与的研究时，尤其是涉及敏感数据和个人隐私的情况下，可能需要进行伦理审查，并遵守相关的伦理规范和准则。

在进行数据收集和分析时，研究者需要充分考虑数据隐私和伦理问题，确保研究的合法性、合规性和道德性。遵守数据隐私和伦理规范不仅是研究者的责任，也是保护用户权益和建立信任的重要环节。

## 十、总结

本研究通过对大众点评和携程两个平台上用户对布达拉宫和大昭寺的评论进行爬取和情感分析，探讨了用户对这两个景点的形象感知。研究结果显示，用户对布达拉宫和大昭寺的评价和情感倾向相对较正面，同时用户的关注度也在不同时间段有所变化。通过用户画像分析，我们还发现了不同平台和地区用户的关注点差异。

在未来的研究中，可以进一步拓展数据来源、分析用户群体差异、挖掘用户行为和需求，以获取更全面、准确的信息。此外，要注意数据隐私和伦理问题，在数据收集和分析中确保用户隐私的保护和研究的合法性、道德性。

这些研究成果对旅游景点的管理者和从业者具有重要指导意义，能够帮助他们更好地理解用户对景点的感知和态度，并根据用户需求进行服务和策划的优化。同时，本研究也为类似领域的未来研究提供了一些方法和启示，如多数据源综合分析、用户需求深入挖掘等。

通过进一步研究和探索，能够更好地理解和满足用户的期望，提升旅游体验，推动旅游业的发展和创新。