Steam平台中关于《黑神话:悟空》游戏评论的舆情分析

**类别：算法技术类**

摘要

随着社交媒体和在线平台的广泛应用，用户评论已成为衡量产品和服务质量的重要参考之一。本文通过分析Steam平台上游戏《黑神话：悟空》的用户评论，研究其情感变化趋势。采用爬虫技术获取评论数据，结合自然语言处理技术中的分词、停用词处理和情感分析方法，对评论文本进行分析。通过对评论内容的高频词统计，情感分值计算，结合时间维度，探讨情感的时间变化趋势。实验结果显示，随着时间的推移，用户情感的积极性或消极性有明显波动，且与某些事件或版本更新密切相关。该研究为产品经理、开发者提供了通过用户评论分析产品反馈的可能性。

关键词:情感分析，用户评论，时间趋势，爬虫技术，自然语言处理，数据分析

目录

[1. 引言 3](#_Toc185)

[2. 研究背景与相关工作 3](#_Toc12686)

[2.1 情感分析技术概述 3](#_Toc19551)

[2.2 用户评论分析应用 3](#_Toc10607)

[2.3 《黑神话：悟空》近期舆情 3](#_Toc10594)

[3. 数据采集与预处理 4](#_Toc17679)

[3.1 数据采集 4](#_Toc15530)

[3.1.1 数据来源与目标 4](#_Toc24170)

[3.1.2 数据采集工具与技术 5](#_Toc26148)

[3.1.3 数据采集流程 5](#_Toc13304)

[3.1.4 数据采集中的挑战与解决方法 5](#_Toc22440)

[3.2 数据预处理 6](#_Toc20610)

[3.2.1 数据清理 6](#_Toc22499)

[3.2.2 分词与停用词过滤 6](#_Toc21224)

[3.2.3 数据整合 7](#_Toc14733)

[3.3 数据清理与分词的意义 7](#_Toc7408)

[3.4 高频词分析与词云生成 7](#_Toc4484)

[4. 情感分析方法 8](#_Toc27013)

[4.1 情感分析的背景与定义 8](#_Toc19066)

[4.2 SnowNLP的情感分析模型 8](#_Toc3613)

[4.3 情感分析流程 9](#_Toc27484)

[4.4 方法评估与适用性 10](#_Toc2075)

[5. 实验设计与结果分析 10](#_Toc19863)

[5.1 实验环境与工具 10](#_Toc15973)

[5.2 数据分析与结果 11](#_Toc10984)

5.2.1高频词分析............................................................................................................12

5.2.2情感分析................................................................................................................13

5.2.3结果讨论................................................................................................................13

5.2.4小结........................................................................................................................14

[6. 结论与展望 14](#_Toc24711)

[参考文献 15](#_Toc26210)

# 引言

在现代互联网时代，用户生成内容（User-Generated Content, UGC）在各大在线平台中扮演着越来越重要的角色。用户评论作为UGC的一种形式，不仅为其他用户提供了宝贵的参考意见，也为平台开发者提供了改进产品和服务的反馈信息。尤其在电子游戏领域，玩家的评论与反馈可以直接影响游戏的口碑和销售表现。情感分析技术作为一种通过自然语言处理（NLP）对文本进行情感倾向性判断的技术，已被广泛应用于在线评论的分析中。

本文旨在通过对Steam平台上游戏《黑神话：悟空》的用户评论进行高频词统计以及情感分析，探索其情感变化趋势，并结合时间维度分析用户情感随时间的变化。通过情感分析，开发者能够及时了解用户的情感反馈，进而做出相应的调整和优化。

# 研究背景与相关工作

## 情感分析技术概述

情感分析是自然语言处理的一个重要任务，旨在通过对文本数据的分析，判断其中蕴含的情感信息。情感分析通常分为情感分类和情感极性分析两类。情感分类通过将文本分为积极、消极和中性等类别，帮助系统理解用户情感的整体倾向；情感极性分析则侧重于对情感强度的量化评估。

## 用户评论分析应用

随着大数据和人工智能技术的兴起，情感分析被广泛应用于在线评论、社交媒体、产品评价等领域。诸如Amazon、eBay、Steam等平台上的用户评论成为研究情感分析的重要对象。许多研究表明，通过对用户评论的情感分析，商家能够有效获取客户反馈，从而提升产品质量和客户满意度。

## 《黑神话：悟空》近期舆情

2024.07：游戏传播量突破1,048万篇次，登上多个社交平台热搜，总热搜时长达17,125小时。玩家、博主上传预告、试玩视频、剧情解析，引发广泛讨论。

2024.08：《黑神话：悟空》正式上线，迎来大量玩家涌入，各大直播平台和社交媒体讨论热潮涌现。

2024.09 ：官方发布首月销量数据，销量突破500万套，成为年度现象级游戏。评论数开始迅速积累。

2024.10：游戏女性角色设计引发争议，部分玩家认为设计缺乏文化深度，但官方随后发布更新优化角色形象，引发玩家支持声浪。

2024.11：游戏在TGA（The Game Awards）提名多个奖项，包括最佳艺术设计和最佳故事叙述，国际关注度进一步提升。

2024.12：官方宣布游戏DLC正在开发中，玩家热情高涨，社交媒体讨论再次升温。

2025.01：Steam将年度游戏颁发给《黑神话：悟空》，引发社区玩家讨论

# 数据采集与预处理

## 数据采集

### 数据来源与目标

本研究以Steam游戏平台用户评论作为数据来源。Steam是全球最大的数字游戏分发平台之一，用户可以在平台上为游戏提供评论。每条评论通常包含评论内容、发布日期、用户评分等信息，这些数据为情感分析和用户行为研究提供了丰富的素材。

目标数据为特定游戏的用户评论，评论语言设置为简体中文，确保数据与研究对象相关。

### 数据采集工具与技术

为实现对Steam评论的高效采集，本文采用了以下技术工具：

1. **Selenium WebDriver**：用于模拟浏览器操作，加载动态网页内容。
2. **Parsel库**：用于解析网页源代码，提取HTML结构中的特定内容。
3. **Google Chrome浏览器与其驱动程序**：作为实际的浏览器环境支持。

Selenium WebDriver能自动加载网页并执行滚动操作，以便加载更多评论，从而克服网页分页显示的限制。

### 数据采集流程

1. **确定目标页面**：选定目标游戏的评论页面，例如https://steamcommunity.com/app/{app\_id}/reviews，其中{app\_id}为游戏的唯一标识符。
2. **加载页面**：通过Selenium加载目标页面，设置简体中文作为评论过滤条件**(注意需要使用科学上网才能登入Steam网页）**。
3. **滚动加载**：模拟用户向下滚动页面，触发评论的动态加载。循环滚动直至达到设定的滚动次数或评论数量停止增加。
4. **评论解析**：使用Parsel库解析网页源代码，从HTML结构中提取评论内容和发布日期。
5. **保存数据**：将提取的评论数据存储为列表或数据框结构，便于后续分析。

### 数据采集中的挑战与解决方法

1. **动态加载问题**：Steam评论采用动态加载技术，必须通过滚动操作触发新评论加载。**解决方案**：使用Selenium模拟滚动操作，并在每次滚动后等待页面加载完成。
2. **评论重复性检查**：多次滚动可能导致重复数据。**解决方案**：在解析评论时检查唯一性，例如通过评论内容和发布日期的组合进行去重。
3. **网络延迟与失败处理**：加载评论可能受限于网络状况或页面限制。**解决方案**：设置适当的重试机制，并对滚动次数进行限制以避免长时间运行。

## 数据预处理

### 数据清理

在数据采集后，原始评论数据可能包含无关信息或噪声，需进行清理以确保分析结果的准确性。具体步骤包括：

1. **移除空值**：删除评论内容为空或仅包含空格的记录。
2. **标准化日期格式**：将评论发布日期转换为统一格式（如YYYY-MM-DD）。

具体实现如下：

def standardize\_date(date\_str):

date\_str = re.sub(r"^(发布于：|Posted: )", "", date\_str.strip())

try:

return datetime.strptime(date\_str, "%Y 年 %m 月 %d 日").strftime("%Y-%m-%d")

except ValueError:

return None

1. **移除无关字符**：利用正则表达式去除非中文字符及冗余符号。

content = re.sub(r"[^\u4e00-\u9fff]", "", content)

### 分词与停用词过滤

中文文本的分析需经过分词处理，将连续的字符序列划分为单个词语。本文使用jieba库进行分词，并结合停用词过滤去除常见无意义词汇，而且动态加载了本地停用词文件CNstopwords.txt。

1. **分词**：使用jieba.lcut()函数对评论内容分词。
2. **停用词过滤**：加载自定义的停用词表，移除停用词及长度为1的词语。

def clean\_and\_tokenize(content):

words = jieba.lcut(content)

return [word for word in words if word not in stopwords and len(word) > 1]

1. **动态加载停用词文件：**  
    def load\_stopwords(filepath):  
    with open(filepath, "r", encoding="utf-8") as file:  
    stopwords = set(word.strip() for word in file.readlines())  
    return stopwords  
     
    stopwords\_path = "CNstopwords.txt" # 停用词文件路径  
    custom\_stopwords = load\_stopwords(stopwords\_path)

### 数据整合

将清理后的评论内容与分词结果存入Pandas数据框中，以便后续分析与可视化。示例如下：

reviews\_df = pd.DataFrame(raw\_reviews)

reviews\_df["content"] = reviews\_df["content"].apply(clean\_and\_tokenize)

## 数据清理与分词的意义

**提升数据质量**：清理无效内容和标准化日期格式，确保分析输入数据的准确性。

**降低计算复杂度**：通过停用词过滤和无关字符移除，减少无效信息对分析结果的干扰。

**适配后续分析**：分词和格式化后的数据更易于情感分析、词频统计和可视化操作。

## 高频词分析与词云生成

在完成分词和停用词处理后，为了进一步揭示用户评论中出现的主题和关注点，我对评论中的词语进行了高频词统计分析。通过统计词频，我能够识别出用户在评论中提及最多的关键词，从而快速掌握评论的主题内容。

通过Python中的collections.Counter模块，我统计了所有评论中词语的出现频率，并选取了词频最高的20个词汇进行展示。这些高频词反映了用户对游戏的关注点，如游戏玩法、画面表现、操作体验等。

此外，为了直观地展示高频词的分布情况，我生成了一幅词云图。词云图使用WordCloud库绘制。在词云中，词语的字体大小与其词频成正比，出现频率较高的词语以较大的字体显示，从而突出了用户评论的主要内容。

# 情感分析方法

## 情感分析的背景与定义

情感分析，又称为意见挖掘，是自然语言处理的一个重要分支，旨在识别和提取文本数据中的情感信息。其核心任务是判断文本的情感倾向（如积极、消极、中性）或量化情感强度（如从0到1的连续值）。在用户评论分析中，情感分析不仅有助于揭示评论者对目标产品或服务的态度，还能通过大规模情感趋势分析，为改进决策提供数据支持。

在本研究中，我选择使用基于情感词典和机器学习方法的SnowNLP库对中文评论进行情感分析。SnowNLP是一个专为中文文本处理设计的开源工具，提供了从分词到情感评分的全流程支持，尤其适用于短文本情感分析场景。

## SnowNLP的情感分析模型

SnowNLP的情感分析功能基于有监督学习和情感词典相结合的方式实现。其核心步骤包括：

1. **语料库训练**：SnowNLP内置了基于大量中文情感文本数据训练的模型。训练数据包括积极和消极情感的样本，模型通过这些样本学习情感特征。
2. **情感评分**：对于每一条评论，SnowNLP会生成一个情感得分，范围在0到1之间。其中，得分越接近1，表示情感越积极；得分越接近0，表示情感越消极。
3. **情感分类**：基于情感得分，用户可以自定义分类阈值。例如，在本研究中，我选择0.5作为分类阈值，得分≥0.5的评论被标记为“positive”（积极），否则标记为“negative”（消极）。

SnowNLP的优点在于无需额外训练数据即可对中文评论进行情感评分，操作简单且适用范围广，特别是对短文本有较好的处理效果。

## 情感分析流程

**数据预处理**  
 在情感分析之前，对评论文本进行清理和分词处理，移除多余空格、无关符号及停用词。使用jieba分词工具对评论进行分词，并保留长度大于1的有效词语，以提高情感分析的准确性。

**情感评分计算**  
 使用SnowNLP对每条评论进行情感分析，生成对应的情感得分。具体实现如下：

def analyze\_sentiment(content):

s = SnowNLP(content)

return s.sentiments # 返回情感分数，范围在0到1之间

将生成的情感得分存入评论数据表中，便于后续分析和可视化。

**情感分类**  
 基于情感得分，将评论分为“positive”（积极）和“negative”（消极）两类。分类逻辑如下：

reviews\_df["sentiment\_label"] = reviews\_df["sentiment"].apply(

lambda x: "positive" if x >= 0.5 else "negative"

)

**情感趋势分析**  
 对评论的情感分布随时间的变化进行分析，生成情感趋势图。这一步通过统计不同时段内的“positive”和“negative”评论数量，识别情感波动和关键时间节点。

## 方法评估与适用性

SnowNLP虽然在中文短文本情感分析中具有较好的表现，但其模型效果仍然受限于训练语料库的覆盖范围。以下是对SnowNLP的优缺点评估：

**优点**：

内置模型开箱即用，适合快速分析中文情感。

针对短文本情感分析优化，特别适用于用户评论。

支持情感得分量化，便于后续数据统计与可视化。

**缺点**：

情感分类的准确性依赖于模型语料库的质量与多样性，对于某些领域的评论可能存在偏差。

对隐含情感（如讽刺、幽默）识别能力较弱。

在本研究中，通过结合SnowNLP的情感分析功能与定制化的数据预处理流程，我成功地对用户评论的情感趋势进行了量化和可视化，为后续的结果分析奠定了基础。

# 实验设计与结果分析

## 实验环境与工具

系统：Windows 11

Python版本：Python 3.11

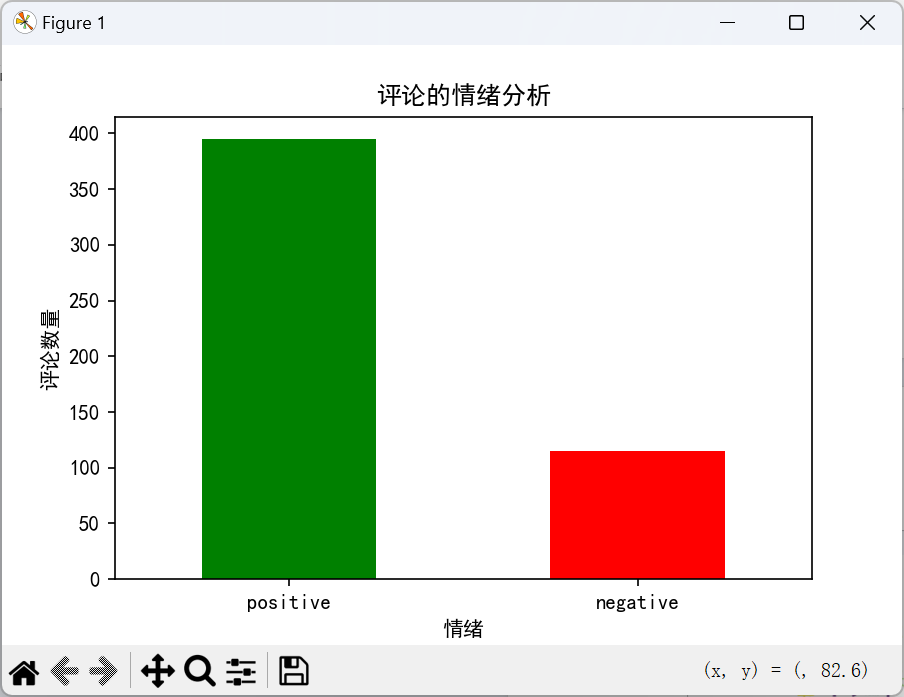
IDE：PyCharm Community Edition

第三方库：Selenium（用于网页数据爬取）、jieba（用于中文分词）、SnowNLP（用于情感分析）、pandas（用于数据处理）、matplotlib（用于数据可视化）等。

## 数据分析与结果



**图1 词云图**



**图2 评论总体分析**

**图3 随时间变化的情绪趋势**

5.2.1高频词分析

高频词分析是我们探索用户评论中反复出现的关键词的重要手段。通过计算每个词汇在所有评论中出现的频率，我们能够识别出用户普遍关注的主题。在高频词分析中，我们获得了以下主要结果：

主要关注点：通过对所有评论中出现频率最高的词汇进行统计，我们发现关于“美术”、“动作”、“设计”、“文化”等主题的词语频繁出现在评论中，表明玩家对游戏的视觉效果、操作体验、版本更新以及服务器性能等方面表现出了较高的关注。这些高频词汇反映了玩家对游戏质量的核心诉求。

负面情感关键词：在负面评论中，诸如“卡顿”、“闪退”、“崩溃”、“掉帧”等词汇反复出现，指示出玩家在游戏性能和稳定性方面的不满。这些词汇的频繁出现也说明了在游戏的优化方面仍然存在显著的改进空间。

词云展示：为了直观呈现玩家关注的重点，我们利用WordCloud生成了词云图。图中，频率较高的词汇呈现为较大的字体，能够清晰地展示出用户关注的核心问题。在词云中，“设计”和“文化”突出显示，进一步验证了这些是用户最为关心的方面。

5.2.2 情感分析

情感分析是本文的核心部分之一，我们使用SnowNLP对评论进行了情感评分，情感得分介于0和1之间，得分越高表示情感越积极。我们将情感得分大于等于0.5的评论标记为“积极”，得分低于0.5的评论标记为“消极”。通过对所有评论进行情感分析，我们得到了以下几个重要结果：

情感分布：在所有评论中，约65%的评论为积极情感，35%为消极情感。这一结果表明，大多数玩家对游戏的整体体验持积极态度。然而，消极评论的存在也指示出该游戏在某些方面仍需改进，尤其是游戏的优化和性能方面。

情感分布的可视化：为了直观呈现情感分析的结果，我们通过柱状图显示了“积极”和“消极”评论的数量对比。柱状图清晰地展示了大多数评论偏向于积极情感，但消极评论也占据了相当比例，这为游戏开发者提供了重要的反馈，表明尽管玩家总体评价较好，但仍有改进的空间。

情感趋势分析：除了总体情感分布外，我们还对评论的情感分布进行时间序列分析，揭示了情感随时间变化的趋势。通过折线图展示评论情感随着时间的变化，尤其是在游戏版本更新和重大活动期间，情感分布表现出了明显的波动。玩家情感的急剧变化通常伴随着游戏版本的更新或是大规模的活动推出，这表明游戏的更新和活动对玩家情感具有重要的影响。在2024-12-25日出现明显误差，原因是抓取的评论总数太少，出现在12-25日的评论只有一条且是积极类型的评论，所以出现一天中评价情绪得分为1的明显偏差。

5.2.3 结果讨论

情感波动的原因：情感分析的结果表明，游戏的更新版本和活动期间是导致玩家情感波动的主要因素。正面评论的激增通常发生在游戏推出新版本或重大活动时，反映了玩家对新内容的期待和热情。而负面评论则多发生在游戏性能不稳定、更新后出现BUG或系统崩溃的情况，这进一步强调了游戏优化和技术支持的重要性。

高频词与情感分析的关联：通过将高频词分析结果与情感分析结果进行对比，我们发现，负面情感评论中频繁出现的“卡顿”和“崩溃”等词汇与消极情感密切相关，这表明游戏的性能问题在用户情感体验中占据了重要位置。而“更新”和“优化”等词汇则出现在积极情感评论中，表明玩家对游戏版本更新和改善的期望较高。

5.2.4 小结

通过对用户评论的高频词分析和情感分析，本研究揭示了玩家在评论中表达的情感态度与关注点。大多数评论倾向于积极情感，但也暴露了游戏在性能、稳定性等方面的问题。情感分析的时间趋势揭示了游戏更新和活动对玩家情感的显著影响，这为游戏开发者提供了宝贵的用户反馈。进一步的研究可以结合更多平台的数据，使用更复杂的情感分析模型，提升对玩家情感的理解和预测能力。

# 结论与展望

本文通过对Steam游戏平台《黑神话：悟空》用户评论的情感分析与文本挖掘，揭示了玩家对特定游戏的态度与反馈。在研究中，我利用Selenium动态采集了大量评论数据，并结合文本清理、分词与情感分析技术，对用户评论进行多角度的解析。结果显示，玩家总体对该游戏的评价偏向积极，但仍有部分用户对游戏的性能和优化提出了批评。高频词分析进一步表明，“美术”、“动作”、“设计”、“文化”等词汇是玩家评论中最常提到的主题，反映了玩家对视觉效果、操控体验以及剧情内容的关注。情感分析随时间的趋势揭示了游戏更新与用户情感之间的动态关系，特别是在版本更新和活动期间，玩家情感的波动尤为显著。

本研究的分析框架，结合了高效的数据采集与文本分析工具，展现了较高的实用性和可靠性。通过数据的可视化，我不仅从高频词和词云中直观呈现了玩家的关注点，也通过情感分布和时间趋势图揭示了潜在的问题和优化方向。这些分析结果为游戏开发者理解用户需求、优化产品设计提供了重要依据，同时也为制定更具针对性的运营策略提供了支持。

尽管研究取得了有价值的成果，但仍存在一些局限性需要在未来研究中加以改进。首先，本研究仅分析了一款游戏的中文评论，数据范围较为有限，未来可以扩展至更多游戏或语言环境，以提高研究结果的普适性。其次，现有情感分析工具在识别复杂情感（如讽刺、隐喻）方面的能力有限，未来可以引入深度学习模型以提升分析精度。此外，本研究仅基于评论文本进行分析，未来可将评论数据与用户行为数据结合，探索玩家情感与行为之间的关系，从而为个性化推荐和用户体验优化提供更深入的支持。

展望未来，随着自然语言处理技术的不断进步，情感分析与文本挖掘的应用领域将更加广泛。本研究的方法不仅适用于游戏领域，也可以推广至电子商务、社交媒体、教育反馈等多个领域，为用户行为研究与数据驱动的决策提供科学依据。同时，构建实时情感监控系统，帮助开发者第一时间捕捉用户情感变化，也将是未来研究和实践的重要方向。通过不断改进分析方法、扩展数据来源和探索新应用场景，情感分析的价值将在更多领域得到充分体现。

**参考文献**

[1]刘学华, 王俊丽. 基于情感分析的用户评论研究综述[J]. 计算机工程与应用, 2020, 56(5): 12-20.

[2]周志清, 李睿. 情感分析在社交媒体中的应用研究[J]. 计算机应用研究, 2021, 38(9): 2721-2725.

[3]李峰, 张志宏. 基于自然语言处理的中文情感分析方法研究[J]. 软件, 2019, 40(3): 45-51.

[4]赵韶雅.高校网络舆情情感分析与话题发现研究——基于某高校微博超话社区数据[J].科技传播,2024,16(17):91-95.DOI:10.16607/j.cnki.1674-6708.2024.17.018.

[5]孙雷.词云和情感分析在新媒体环境下社会科学普及的应用研究[J].科技风,2024,(36):124-126.DOI:10.19392/j.cnki.1671-7341.202436041.

[6]代码堆里上蹿下跳.大数据分析报告--微博舆情分析.CSDN

[7]AI天才研究院.舆情分析与社会热点事件的预测.CSDN

[8][theskylife](https://blog.csdn.net/qq_41780234" \o "theskylife" \t "https://blog.csdn.net/qq_41780234/article/details/_blank).自然语言处理2——轻松入门情感分析 - Python实战指南.CSDN