

ACADEMIC STYLE GRADUATION THESIS DEFENSE

基于小红书用户反馈数据的游戏营销内容选题预测——以《恋与深空》为例

 答辩学生：杨阳

 指导老师：龙思薇、刘杉

GRADUATION

CO

NT

EN

目录

01. 绪论
Introduction

02. 主体
Main Text

03. 结论
Conclusion

CONCLUSION

绪论

GRADUATION

01. 研究背景

- **市场增长：**女性向游戏市场持续增长，情感消费需求显著（如《恋与深空》首月流水5.09亿）。
- **用户特征：**Z世代女性偏好个性化、情感共鸣、沉浸式体验，付费意愿高。
- **政策支持：**政府对女性向游戏采取“扶持+监管”双轨制。
- **行业挑战：**
 - 游戏口碑分化（如《恋与深空》技术表现力强，但剧情节奏、人设单一受争议）。
 - 传统分析方法单一，难以挖掘用户深层需求。

02. 研究问题及意义

- **研究问题：**如何基于小红书平台的用户生成内容，量化玩家对游戏内容的主题偏好与情感需求，并转化为可指导游戏内容营销策略的数据指标。
- **研究意义：**对游戏厂商：优化内容迭代（如剧情、角色、玩法），提升用户留存与付费转化。精准营销，增强社区活跃度与品牌口碑；为《恋与深空》优化提供方向，提升用户体验和口碑，为行业提供新思路。

03. 文献综述

(1) 小红书平台用户行为研究

- 平台特性：强交互性、高聚集性（范淑妍，2023）；用户行为："主动搜索+深度互动"（千瓜数据，2022）
- UGC内容驱动社交传播（陈奕竹，2023）；研究缺口：游戏营销预测分析领域较少涉足

(2) 游戏品牌营销研究

- 核心策略：内容营销+社交传播+私域运营（艾瑞咨询，2024）；情感营销与玩家认同感建立（李一长，2024）
- 口碑传播影响付费决策（李思远，2024）

(3) 《恋与深空》相关研究

- 消费模式："灵活支付者"为主，抽卡系统为核心付费点；社交互动：高活跃UGC社区，玩家自发分享攻略与二创
- 营销现状：跨界联名（如肯德基）、技术营销（裸眼3D广告）；现存问题：剧情节奏、更新速度争议

04. 研究创新点

• 方法创新：

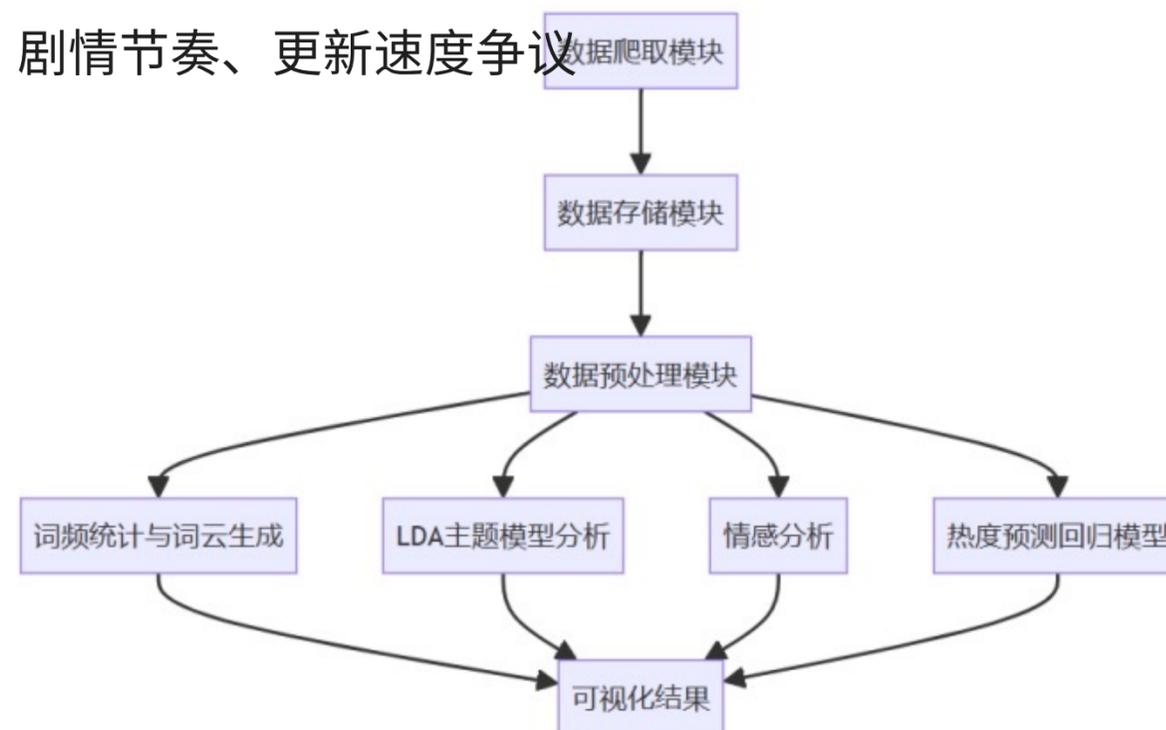
结合LDA主题模型 + SnowNLP情感分析 + 随机森林预测

"主题-情感"双维度量化玩家需求

• 应用创新：

构建"用户反馈-主题分析-营销响应"敏捷决策模型

为中小团队提供低成本运营优化路径



MAIN TEXT

主体

GRADUATION

《恋与深空》产品在小红书平台的内容营销特点分析

• **游戏介绍：**《恋与深空》由上海叠纸互娱研发，2024年1月18日上线，“恋与”系列续作，近未来科幻背景，全3D形式。

• **核心特色：**

1.全3D沉浸式恋爱互动；2.融合卡牌养成+动作战斗；3.AI捏脸系统+多感官交互

• **市场表现：**

1.次月海外收入达首月2倍；2.全球互动叙事手游收入冠军

• **核心竞品 SWOT 分析：**

对比《恋与制作人》等竞品。《恋与深空》优势是技术、玩法、IP；劣势为抽卡、剧情、社区管理问题。机遇是市场增长、用户拓展；威胁是竞品同质竞争、用户审美疲劳。展示各竞品具体优劣势、机会和威胁（表格呈现）

• **《恋与深空》在小红书平台内容营销特点**

1.目标用户与平台属性吻合：小红书月活3亿，游戏玩家1.1亿。用户年轻、女性多、消费强，与游戏核心、次核心用户重合。

2.破圈传播、社区建设与内容互动模式：小红书助力《恋与深空》传播，征集活动效果优，用户形成社区。游戏UGC内容多、浏览量高。评论区促传播，助厂商优化推广。

竞品	优势 (S)	劣势 (W)	机会 (O)	威胁 (T)
恋与制作人	国乙开山鼻祖，IP影响力深厚；剧情成熟，角色人设差异化显著。	2D画面陈旧，玩法单一；用户年龄层固化，难以吸引新玩家。	通过IP衍生内容（如动画、周边）延长生命周期。	3D化趋势下竞争力下降；新竞品分流核心用户。
光与夜之恋	腾讯资源支持，投放力度强劲；18+内容吸引成年女性，差异化定位明确。	剧情争议多（如“强制婚育”桥段）；社交玩法薄弱，用户黏性依赖短期活动。	依托微信生态构建社交裂变；跨界联名（如奢侈品）提升品牌调性。	内容同质化严重；玩家对“饭圈化”运营反感。
未定事件簿	律政题材创新，推理玩法增强策略性；骨骼动画技术提升表现力。	恋爱线占比低，情感满足不足；用户规模较小，社区活跃度有限。	拓展职场女性用户市场；通过案件联动（如公益法律咨询）增强社会价值。	核心玩法与乙女需求错位；难以突破小众圈层。
代号鸢	反套路剧情（如“权谋黑化”）吸引猎奇用户；买断制降低付费压力。	产能不足，更新频率低；技术优化差（如卡顿、闪退）。	差异化叙事填补市场空白；通过二创社区（如Lofter）维持热度。	内容尺度引发监管风险；长线运营能力存疑。
恋与深空	3D技术革新、战斗玩法差异化、IP延续性	抽卡机制争议、剧情同质化、社区管理不足	女性向游戏市场增长、泛用户拓展	竞品同质化竞争、用户审美疲劳

数据采集与预处理

1. 数据采集概况

- 数据来源：小红书平台《恋与深空》相关UGC
- 数据规模：
 - 笔记数据：800条
 - 评论数据：3,751条
- 采集字段：
 - 标题、时间、收藏量、点赞量、分享量、评论内容

2. 数据清洗流程

- 去重处理：drop_duplicates消除重复数据
- 文本清洗：
 - 正则表达式提取中文字符
 - Jieba中文分词
 - 过滤停用词/无效字符
- 结构化存储：
 - 生成"清洗后评论"列
 - Excel标准化存储

笔记id	评论时间	评论正文	评论点赞	头像	评论用户	用户id	ip	评论图片	sec_token_id	meta_id
67b90d4ec	2025-02-1	研德安成	0	https://img	5ea32a1f	江苏		https://r	AB6e83ba667b912fe	67b90d4ec0000000783401b
67b90d4ec	2025-02-28	AMurIax	0	https://	647872dc	湖北		AB6e83ba667b90w0c	67b90d4ec0000000783401b	
67b90ce7c	2025-02-28	01:16:	1	https://	60a2a79e	广东		https://r	AB6e83ba667b911ed	67b90ce7c0000000281e406
67b90c84c	2025-02-1	刀我[哭惹]	1	https://	6417dbdf	美国		https://r	AB6e83ba667b90d7c	67b90c84c0000000990d5d4
67b90c84c	2025-02-1	老师，大	1	https://	67b9c9c	辽宁		AB6e83ba667b9c9c	67b90c84c0000000990d5d4	
67b90c84c	2025-02-1	新的好好	1	https://	6417dbdf	美国		https://r	AB6e83ba667b90da9c	67b90c84c0000000990d5d4
67b90c84c	2025-02-1	不要这样	1	https://	67b90dab	重庆		https://r	AB6e83ba667b90dab	67b90c84c0000000990d5d4
67b90c84c	2025-02-1	补药啊	1	https://	67b90dab	江西		https://r	AB6e83ba667b90dab	67b90c84c0000000990d5d4
67b90c84c	2025-02-1	啊啊啊啊	1	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90c84c0000000990d5d4
67b90c84c	2025-02-1	[哭惹]	1	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90c84c0000000990d5d4
67b90c84c	2025-02-1	小孩子半	0	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90c84c0000000990d5d4
67b90c1d	2025-02-1	没有防	0	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90c1d0000000990d5d4
67b90b6d	2025-02-1	没有防	0	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90b6d0000000990d5d4
67b90b6d	2025-02-1	这是哪	0	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90b6d0000000990d5d4
67b90b6d	2025-02-1	不知道	0	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90b6d0000000990d5d4
67b90b6d	2025-02-1	啊特	0	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90b6d0000000990d5d4
67b90b6d	2025-02-1	最后响	0	https://	67b90d11	广东		https://r	AB6e83ba667b90d11	67b90b6d0000000990d5d4

关键词	计数
可爱	60
好看	44
衣服	38
剧情	35
生气	28
游戏	28
好看	27
失望	22
啊特	19
主控	18
祁煜	18
制服	15
动作	15
制服	12
幸福	12
不好	11
恋爱	11
男主	10

3. 词频统计结果

- 高频词示例：
 - 角色相关：好帅、祁煜、制服
 - 体验相关：剧情、互动、失望
 - 付费相关：销冠、氪金

一、模型构建

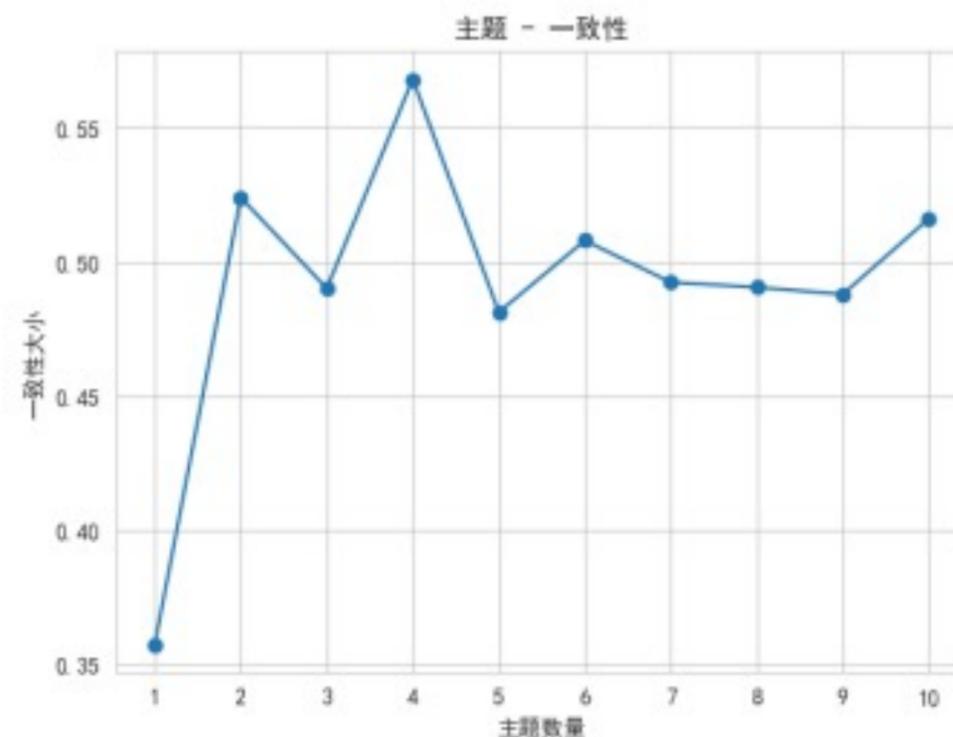
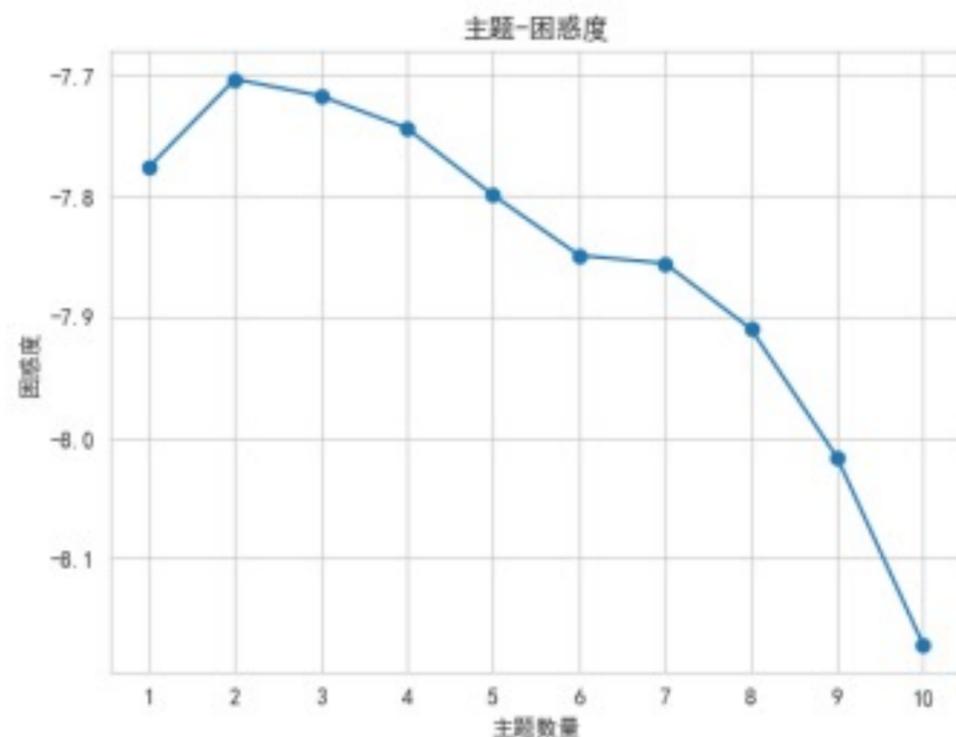
(一) 构建字典和语料库

从本地CSV文件读取已分词的评论数据并进行预处理，预处理包括分词和词频统计，根据词频统计结果重构文本数据，将高频词按出现次数进行重复扩展，形成符合LDA模型输入的文本集合，使用gensim库中的corpora.Dictionary来构建词典，并将文本转换为词袋（Bag-of-Words）格式。

(二) 确定最优主题数

LDA主题模型中，确定最优主题数是一个关键问题，通常通过困惑度（Perplexity）和一致性（Coherence）两个指标来评估。困惑度是一种用于评估语言模型预测能力的指标，反映了模型对给定数据的不确定性程度。

由图4.1可知，主题词的最优主题数4效果最好。



二、LDA主题分析结果

(一) 主题词云可视化

主题0：服装设计（最大字号：衣服）

主题1：剧情体验（突出词：恋爱/生气）

主题2：角色魅力（关键词：好帅+失望）

主题3：玩法优化（整改显著）

(二) 核心发现

用户关注维度：

角色外观设计（41%）

剧情叙事体验（28%）

玩法系统评价（19%）

付费机制讨论（12%）

主题	占比	核心关键词	情感倾向
0	41%	衣服、销冠、心动	正面
1	28%	剧情、生气、恋爱	复杂
2	19%	好帅、祁煜、失望	两极
3	12%	可爱、狼尾、整改	负面



图 4.3 主题 0 词云图



图 4.4 主题 1 词云图



图 4.5 主题 2 词云图



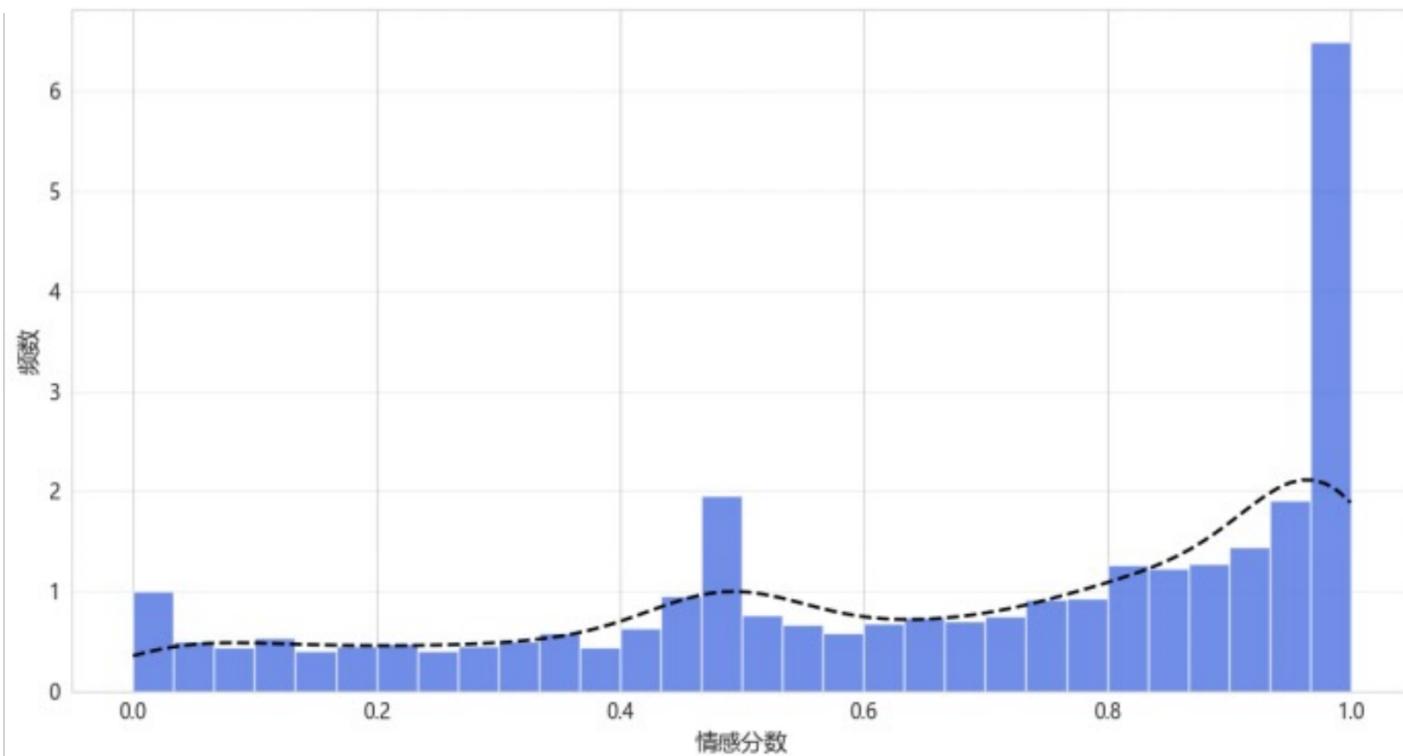
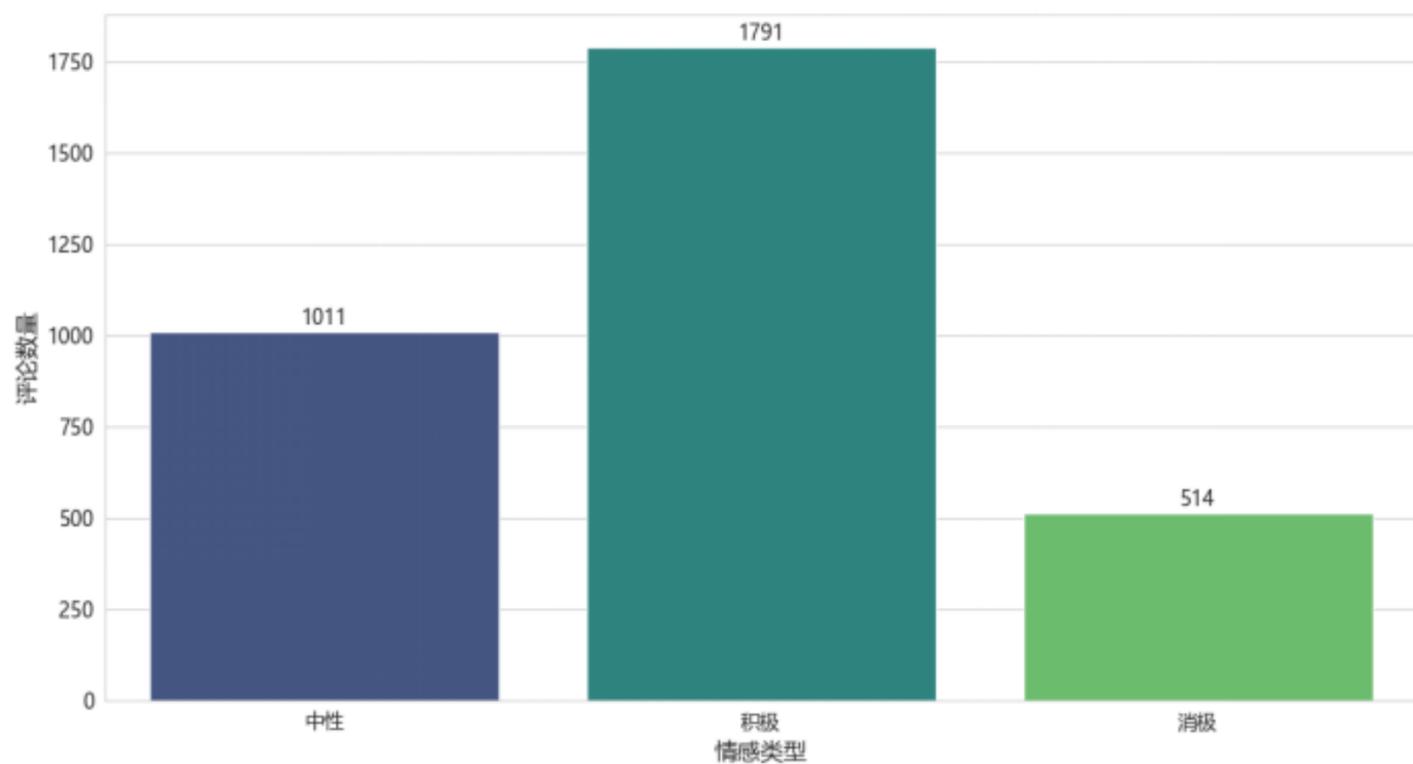
图 4.6 主题 3 词云图

一、情感分析与特征工程

针对小红书《恋与深空》用户评论数据进行了系统性的数据预处理工作，包括编码检测与处理、特殊数值转换、缺失值处理，基于研究需求，构建**互动等级分类**、**文本特征提取**、**时间特征工程**、**情感标签编码**4类关键特征变量

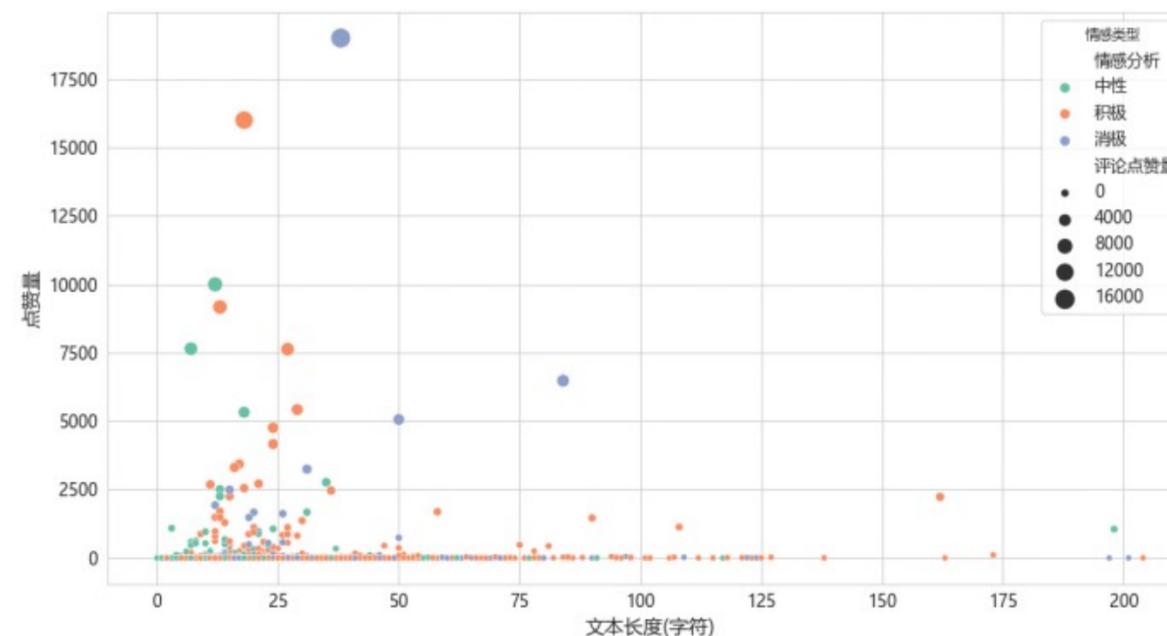
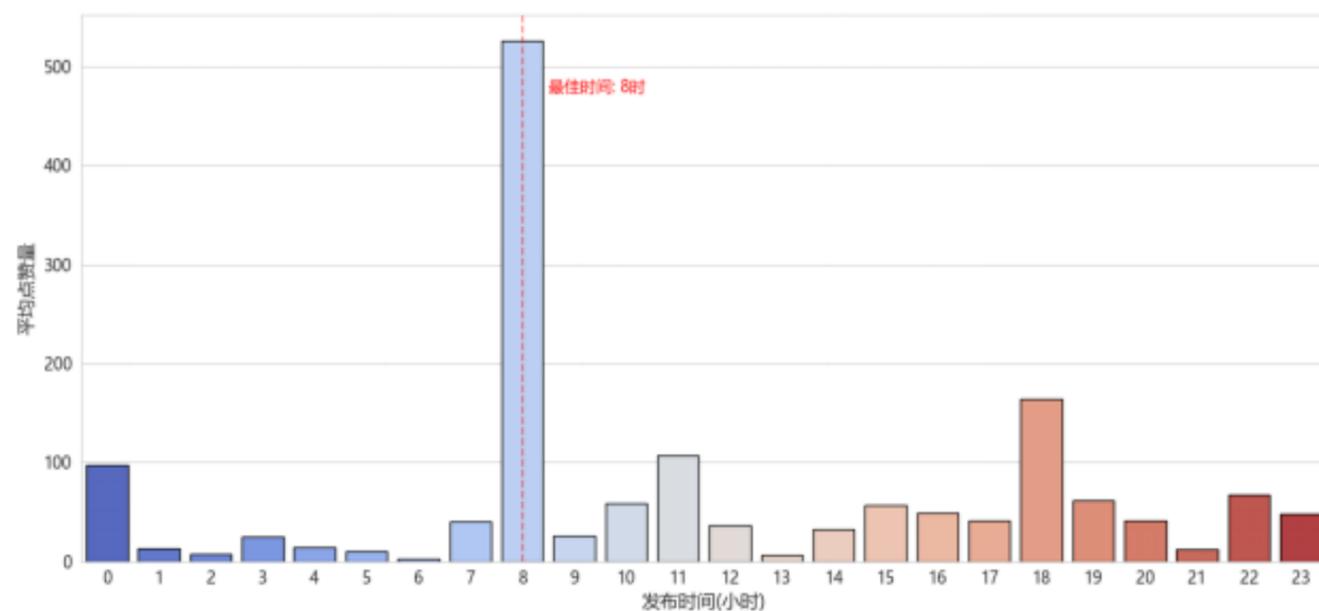
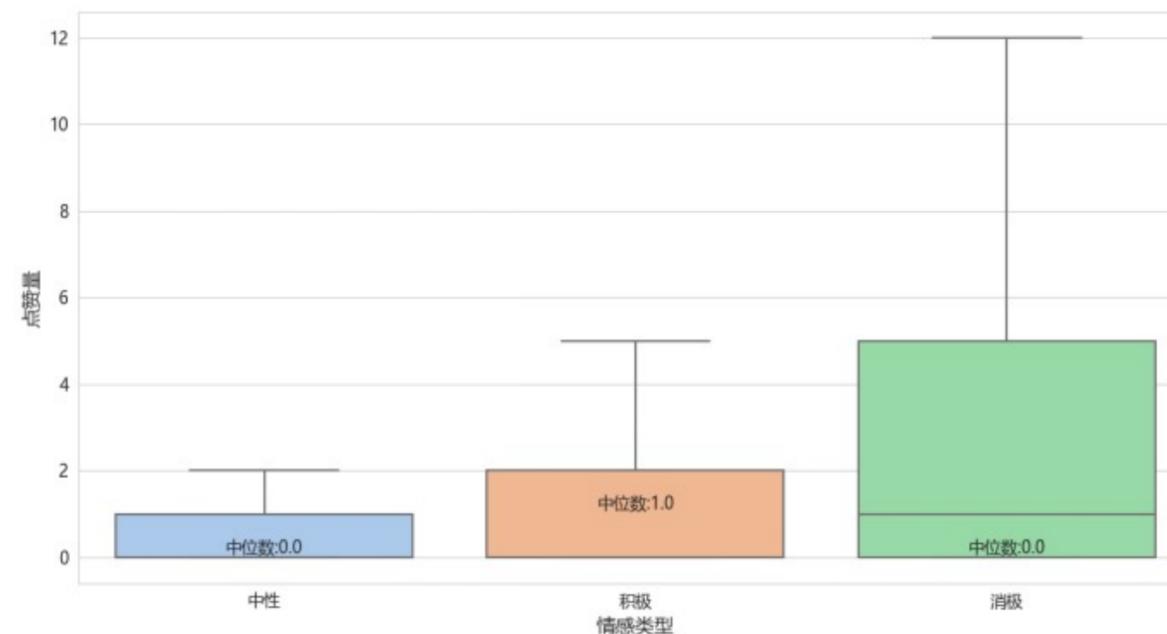
通过图5.1可见，积极评论占比最高（1791条），显著高于中性（约750条）和消极评论（约500条），表明《恋与深空》整体用户反馈**正向**。

通过图5.2可见，情感分数分布呈现右偏态特征，核密度曲线峰值位于0.6-0.8区间，进一步验证**用户情感倾向以正面为主**。



二、互动特征可视化分析

- 通过图5.3可见，积极评论的点赞量中位数最高（1.0），显著优于中性（0.0）和消极评论（0.0）。箱线图显示积极评论的点赞量分布范围更广，存在多个高互动异常值。
- 通过图5.4可见，散点图显示文本长度与点赞量呈弱正相关（ $R^2=0.32$ ），中等长度内容更易获得高互动，积极情感的评论在相同长度下普遍获得更高点赞量，体现情感倾向的调节作用。
- 通过图5.5可见，日间8时发布的评论平均点赞量最高（峰值较次高峰高出23%），建议作为核心内容发布时间，凌晨3-5时为互动低谷期，平均点赞量不足高峰时段的30%。

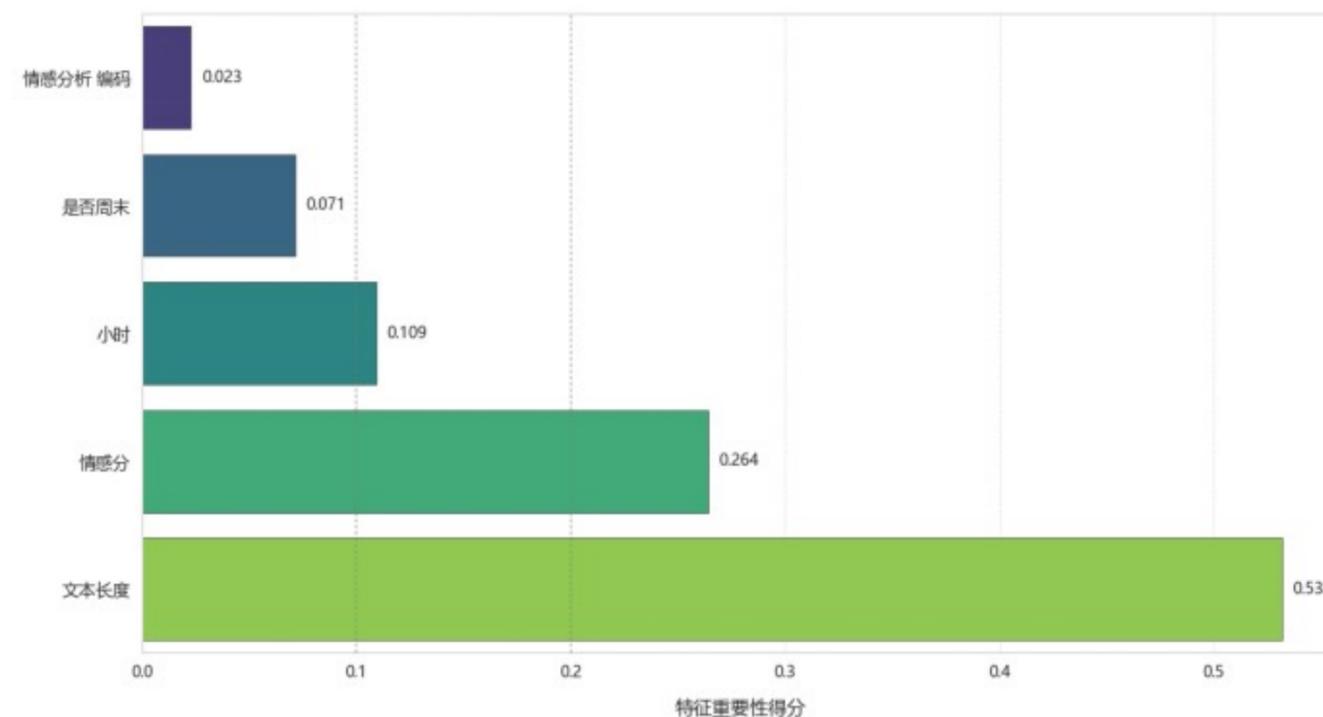


三、预测模型构建

本研究采用**随机森林**算法构建用户互动预测模型，特征选择选取**"情感分"**、**"文本长度"**、**"发布时间（小时）"**、**"是否周末"**和**"情感分析编码"**作为输入特征；将评论点赞量高于中位数的样本标记为**"高互动"**（1），其余为低互动（0）；按7:3比例划分训练集与测试集，采用分层抽样（stratify=y）保证类别分布一致；设置模型参数，例如决策树数量（n_estimators）为150，最大树深度（max_depth）为5，构建模型并训练。

该模型准确率为61.6%，F1-score中正类（高互动）为0.599，负类为0.632，召回率与精确度基本平衡（约0.59-0.64），表明模型无显著偏倚。显示各特征对预测高互动评论的贡献度，其中关键特征**发布时间（小时）**重要性得分最高（0.471），验证了前文发现的8时发布效果最佳的时间规律；第二重要特征**情感分**（0.221），与情感分析结论一致，正面情感更易引发互动；**文本长度**贡献度0.109，中等长度内容更具传播性；次要特征**是否周末**重要性0.071，周末发布效果略优于工作日；**情感类型编码**：影响最小（0.023），说明具体情感分数比粗粒度分类更具预测力。

	precision	recall	f1-score	support
0	0.625	0.639	0.632	513
1	0.606	0.591	0.599	482
accuracy			0.616	995
macro avg	0.616	0.615	0.615	995
weighted avg	0.616	0.616	0.616	995



四、推荐系统架构设计

- 输入示例：(积极,100字,12时,非周末)
- 输出结果：高互动概率：69.06%
- 数据驱动的精准确营销
- 提升用户互动率

预测内容方向	关联关键词	优先级得分	建议迭代形式
战斗难度调整	流浪体、时间不够	8.6	新增简单模式+战斗教学优化
CP剧情扩展	追妻、哄人、火葬场	9.1	推出角色专属剧情DLC
抽卡系统改良	五星、概率、推	7.8	保底机制+免费抽卡券
UGC创作支持	剪辑、素材、文	8.3	举办同人创作大赛，设立分赛道

情感

文本长度 84

小时 12

周末 0 1

高互动概率：75.69%

推荐示例：

- 圈外人：这款游戏是为恋爱而烦恼
圈内人：这款游戏是打不过流浪体 战斗中时间不够 思念等级不够 抽不到自己推的五星 见不到自己推而烦恼！
没错 我就是刚刚打不过流...
- 天哪，他笑的那一瞬间，谁懂，那种男友明明知道你在和他玩，他不至超级无敌配合，不是逼迫自己配合爱意的眼睛，里面的爱意瞬间倾...
- 哥这个痛到呲牙咧嘴的表情好涩啊，特别涩有没有人能懂一下。痛苦和隐忍就是哥哥的感情啊...这份畸形的妹妹掌控...

示例文本片段	提取关键词	关键词类型
"打不过流浪体"	战斗难度、失败体验	游戏机制痛点
"彻彻追妻火葬场"	追妻、火葬场、CP剧情	角色关系热点
"太太们剪辑出来好"	UGC创作、二创、素材	玩家创作行为
"抽不到自己推的五星"	抽卡机制、角色获取、概率	付费系统讨论

CONCLUSION

结论

GRADUATION

一、模型验证结果

随机森林模型：

- 准确率61.6%，F1-score 0.62
- 特征重要性：
 - 发布时间 (0.471)
 - 情感分 (0.221)
 - 文本长度 (0.109)

(二) 核心研究发现

- 用户情感特征：
 - 积极评论占比50.3% (显著高于中性/消极)
 - 情感分数峰值0.6-0.8 (右偏态分布)
- 互动规律：
 - 最佳内容：积极情感+80-120字 ($R^2=0.32$)
 - 黄金时段：8时发布 (互动量↑23%)

方向	优先级	实施案例
CP剧情扩展	9.1	角色专属DLC
战斗优化	8.6	公会协作玩法
UGC激励	8.3	同人创作大赛
抽卡改良	7.8	动态概率调整

不足之处

- **数据局限：**
 - 仅小红书单平台数据
 - 未覆盖微博/B站等多元场景
- **模型局限：**
 - 女性向游戏特异性强
 - 未验证开放世界/竞技类游戏

展望

- **多平台整合：**
 - 构建微博+B站+TapTap跨平台数据库
 - ARIMA+LSTM舆情监测系统
- **AI内容生成：**
 - GPT剧情分支生成器
 - Stable Diffusion自动素材生产
 - 热点响应式资源库

THANK YOU FOR WATCHING

恳请各位老师批评指正

 答辩学生：杨阳

 指导老师：龙思薇、刘杉

GRADUATION