

小米SU7上市初期视频社交平台用户关注分析

——基于抖音平台的热门视频及评论数据

何 祥

浙江理工大学经济管理学院, 浙江 杭州

收稿日期: 2024年9月13日; 录用日期: 2024年10月14日; 发布日期: 2024年11月27日

摘 要

本文基于抖音平台收集并分析小米SU7上市初期热门视频, 从创作者与观众视角出发探讨用户关注点。基于描述性分析, 发现头部账号创作优势显著, 但尾部账号也能创造高互动视频。基于主题建模, 发现热门视频聚焦于乘坐空间、外观设计、驾驶安全、智能生态、性价比及驾驶感受。通过情感分析和词云图, 发现积极情感评论数量远超负面, 体现消费者对小米SU7的认可; 负面情感评论集中于价格、性能、质量, 并且发现个性化需求会导致期待差异而形成不同情感倾向的评论。研究表明, 视频社交平台助力新品信息传播, 热门视频数据及评论为企业产品优化提供重要反馈。

关键词

社交平台, 文本挖掘, 公众态度, 主题建模, 情感分析

Analysis of User Engagement on Video Social Media Platforms during the Early Launch of Xiaomi SU7

—Based on Popular Videos and Comment Data from Douyin Platform

Xiang He

School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Sep. 13th, 2024; accepted: Oct. 14th, 2024; published: Nov. 27th, 2024

Abstract

This paper collects and analyzes popular videos related to Xiaomi SU7 during its initial launch

文章引用: 何祥. 小米 SU7 上市初期视频社交平台用户关注分析[J]. 电子商务评论, 2024, 13(4): 5212-5220.

DOI: 10.12677/ec.2024.1341755

period on the Douyin platform, exploring user concerns from both creator and audience perspectives. Through descriptive analysis, it is found that while top accounts exhibit significant creative advantages, tail accounts are also capable of generating highly interactive videos. Utilizing topic modeling, the study reveals that popular videos primarily focus on passenger space, exterior design, driving safety, smart ecosystem, cost-effectiveness, and driving experience. Emotional analysis and word clouds indicate that the number of positive audience comments far exceeds negative ones, reflecting a strong endorsement of Xiaomi SU7 by viewers. Negative sentiments are primarily centered on pricing, performance, and quality, and it is observed that individualized demands can lead to varying expectations, resulting in comments with distinct emotional undertones. The research concludes that video social media platforms facilitate the dissemination of information about new products, and the data and comments from popular videos provide crucial feedback for enterprises to optimize their products.

Keywords

Social Media, Text Mining, Public Attitude, Topic Model, Sentiment Analysis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在电子商务与社交媒体深度融合的当下，市场营销策略正经历着从广撒网式的传统模式向精准触达、强调用户互动与参与的新媒体营销范式的深刻转型[1]。这一趋势在汽车行业尤为凸显，特别是新能源汽车领域，其市场格局与消费者行为模式正被新媒体力量深刻重塑。

小米汽车，作为跨界造车的新势力，凭借其在智能手机与智能家居领域的深厚积累与高性价比策略，自诞生之初便吸引了广泛的市场关注与消费者期待，其品牌效应在汽车行业迅速发酵，形成了一股不可忽视的“流量风暴”。2024年3月28日，小米汽车首款车型SU7的正式发布，不仅标志着小米正式踏入汽车制造领域，更在短短四日内，便实现首批交付，迅速成为社交平台上的焦点。据知微事见数据监测平台分析¹，小米SU7上市发布会在抖音平台的影响力指数高达91.2，远超同类事件，这一数据不仅彰显了小米汽车在新媒体环境中的强大传播力，也进一步验证了新能源汽车市场在新媒体营销下的“破圈”潜力。

随着移动互联网技术的飞速发展，视频已成为社交内容生态中的主导力量。据第52次《中国互联网发展报告》权威发布，我国网络视频用户规模已突破10.44亿大关，这一庞大的用户基数为视频社交平台提供了广阔的发展空间。其中，抖音平台作为视频社交领域的领头羊，其日活跃用户数超过6亿，远超以图文为主的微博等平台，凸显了视频内容在吸引用户注意力、促进用户互动方面的独特优势。因此，视频社交平台已成为洞察消费者行为、把握市场脉搏的重要窗口。

鉴于上述背景，本文旨在通过深入分析小米SU7上市初期在抖音平台上的热门视频及其评论数据，探索创作者和消费者对于该车型的关注的焦点。本文将围绕以下几个核心问题展开研究：一是热门视频数据集中创作者的粉丝分布情况，探讨普通用户是否具备创造热门视频内容的潜力；二是从视频创作者视角出发，分析视频创作者对于“小米SU7上市”事件的关注的焦点与创作主题；三是从消费者视角出发，解析各主题视频下评论的情感倾向与内容，以揭示消费者对于小米SU7的真实看法与期待。通过这一系

¹<https://ef.zhiweidata.com/>。

列分析, 本文期望为企业在新媒体环境下的市场营销提供有价值的参考与启示。

2. 相关研究

2.1 视频社交平台与市场营销的深度融合

视频社交平台的崛起, 无疑为社交媒体领域注入了新的活力, 标志着这一平台正步入一个全新的发展阶段。美国学者 Gibbs 与 Colin 的先见之明, 早已预见到移动视频将成为未来媒体市场竞争的焦点[2], 这一趋势在随后的市场发展得到了充分验证。在商业应用层面, 视频以其独特的互动性和沉浸感, 成为提升品牌忠诚度、深化市场渗透力的强大工具[3]。视频内容的消费模式, 正逐步向社交化、碎片化、分众化及定制化方向发展, 这些特征不仅丰富了用户的视听体验, 也为市场营销提供了更为精准、高效的传播路径[4]。视频营销以其独特的魅力, 在促进品牌认知、激发购买欲望等方面发挥着不可替代的作用[5]。

在强调用户体验的方面, 视频社交平台的优势尤为突出。例如在旅游营销方面, 张跃先和王雪莹(2024)基于 SIPS 模型, 通过详尽的问卷调查, 深入剖析了旅游短视频如何激发顾客灵感, 进而促进顾客融入的复杂机制。他们发现, 关系强度在这一过程中起到了关键的调节作用, 为旅游行业如何利用视频社交平台进行精准营销提供了宝贵的洞见[6]。徐同谦和王志轩(2023)则从 S-O-R (刺激-机体-反应)理论框架出发, 深入探讨了种草类短视频如何通过情感唤起机制影响消费者的购买决策。他们的研究表明, 短视频的有用性、原生性和社交性, 是激发消费者积极情感、促进购买意愿的关键因素, 为短视频营销策略的制定提供了科学依据[7]。此外, 鲁楠(2024)的研究聚焦于抖音等短视频平台在网络营销中的实际应用, 强调了平台高曝光率与用户高粘性对于品牌宣传和产品推广的重要性。进一步指出, 通过加强用户互动、深化与网红的合作, 可以有效增强品牌与用户之间的信任关系, 为品牌的长远发展奠定坚实的基础[8]。

综上所述, 视频社交平台与市场营销的深度融合, 不仅为品牌传播提供了更为广阔的空间, 也让消费者更沉浸式的体验到产品进而增加消费的机率。

2.2. 短文本主题建模

LDA (Latent Dirichlet Allocation)模型被学者广泛用于网络热点内容的主题建模, MA 等在 LDA 模型的基础上加入时间因素, 形成连贯的文本内容, 大大提高话题识别的聚类质量[9], GAVHANE 等对社交媒体信息进行定期重复性筛查和相对标准化周期性检查, 从而区分内容文本所隐含的话题[10], LU 等通过优化语义模板, 基于时间和空间的相似性, 解决了话题识别研究过程中的模糊性[11]。虽然已有不少学者运用基于 LDA 主题模型以及基于该模型改进的模型, 但由于 LDA 主题模型的准确性过于依赖文本长度, 所以在长文本数据集上效果较好, 而随着当下移动社交的发展, 碎片化内容逐渐占据主流, 如微博、推特等文本社交平台上, 短文本是内容的主力, 在新兴的视频平台, 短视频流行, 其包含的文本内容必然较短, 因此对于短文本的主题建模研究也是网络内容主题建模研究的一个重点。Cheng Xueqi 等人提出了一种适用于短文本的主题模型 BTM, 在短文本中主要的问题便是文本稀疏的问题, 如 LDA 中每个文档对应一个主题分布, 每个词对应一个主题 z , 但由于词汇量少, LDA 模型计算的主题分布并不具备统计的意义。而 BTM 是通过滑动窗口内的词共现行为来扩展短文本, 使得短文本的词汇量成倍增长。因此, BTM 主题模型十分适用于各种短文本研究中[12]。

3. 研究方法

3.1. BTM (Biterm Topic Model)主题建模

BTM 模型通过引入“biterm”(词对)这一创新概念, 即文本中任意两个词汇构成的无序词对, 有效地

捕捉了词汇间的潜在关联与相互作用,从而显著提升了主题建模的精确度和效率[12]。BTM 模型的基本假设是,每个文档由多个潜在主题构成,而每个主题则进一步由多个 **biterm** 组成。这一假设契合人类对于文本内容的自然理解方式,即文本内容往往围绕多个主题展开,每个主题又通过一系列紧密相关的词汇得以体现。因此,BTM 模型在解决短文本词汇稀疏性问题上展现出了独特的优势,尤其适用于社交媒体平台上广泛存在的短文本数据分析任务。

3.2. SnowNLP 中文文本情感分析

SnowNLP 作为一款专为中文设计的自然语言处理库,其情感分析功能基于朴素贝叶斯分类器构建,通过精细的文本预处理、特征提取及情感分类流程,能够准确判断文本的情感倾向,并在中文情感分析领域展现出卓越的性能[13]。该算法对每篇文本处理后,会输出一个介于 0 至 1 之间的情感得分,其中得分越接近 0 表示情感越消极,反之则越积极。为便于分析,我们设定情感得分小于 0.3 为负面情感,大于 0.7 为积极情感,其余为中性情感。

4. 数据获取及实验分析

4.1. 数据选取及描述性统计

本文将收集数据的时间范围设定为 2024 年 3 月 20 日至 2024 年 5 月 4 日(即小米 SU7 上市发布会前一周至首次参加北京国际车展这一段时间)。本文利用 Python 编程语言配合 DrissionPage 包实现对抖音平台的自动化爬虫,针对关键词“小米汽车”和“SU7”在抖音平台上进行视频数据的爬取,获取以下字段信息,如:作者名称、作者链接、标题、描述、粉丝数、发布时间、点赞数、评论数、分享、视频 id。由于平台的限制和数据获取的复杂性,本文尽可能地扩大了搜索范围。在经过去重和剔除数据缺失的视频后,本文选取了点赞量、评论量和分享数各前 1000 条视频作为小米 SU7 热点事件的热门视频样本,再运用爬虫技术对该样本集的视频的评论数据进行爬虫。最终形成了一个包含 1077 条视频案例的热门视频数据集,以及 1,912,649 条评论数据集。由于平台政策的限制,本文未能获取到所有的评论数据。

在分析中,为了探究不同粉丝基数的创作者是否能创作出热门视频,本文根据创作者的粉丝数将他们分为三个层级:头部账号(粉丝数超过 1,000,000)、中部账号(粉丝数在 100,000 到 1,000,000 之间)和尾部账号(粉丝数少于 100,000)。相关的描述性分析结果如表 1 所示。

从表 1 中可以观察到,头部账号在数据集中的占比相对较大,达到了 41.7%,并且在互动量上展现出了显著的优势。具体来看,头部账号的视频平均点赞数为 41281.852,平均评论数为 5018.01,平均转发数为 17599.315,这些数据均高于中部和尾部账号的相应均值。此外,头部账号的视频时长均值为 157.26 秒,高于中部账号的 131.7 秒和尾部账号的 92.62 秒,说明头部账号创作的长视频更有竞争力。尽管头部账号在互动量上占据优势,但中部和尾部账号同样展现出了创作热门视频的潜力。特别是尾部账号,尽管其粉丝数较少,但依然有视频能够获得超过百万的点赞数,这一成绩甚至超过了热门视频中头部账号的均值。这一现象表明,即使是粉丝数较少的尾部账号,也有可能通过创作高质量的内容,在抖音平台上获得高互动量,成为热门视频的创作者。

4.2. 创作者关注焦点

本文采用主题建模的方法分析创作者的关注焦点。为了对热门视频进行主题建模,首先需要将视频数据转换为视频内容文本数据。本文运用基于深度学习框架的语音识别工具包 Vosk,该工具以其高度的简洁性与执行效率在语音转文本领域广受赞誉。尽管通过数据集进行深入训练可能能够进一步优化模型的识别精度,但鉴于社交平台中热门视频创作者的声线特征纷繁复杂,实施针对性的模型训练不仅耗时

耗力，且效果难以普适化。鉴于此，本文决定直接采用预训练的 Vosk 模型(vosk-model-cn-0.22)，该模型在广泛的语言环境下展现出良好的识别性能。为确保数据质量，本文随后对自动生成的文本进行了细致的人工复核与校正，力求每一段文本都能准确无误地反映原始音频的内容。在获取视频内容文本后，本文运用在中文自然语言处理中广泛运用的 jieba 分词工具配合自定义的停用词，同义词库，完成对视频内容文本的预处理，以满足之后进行主题建模的需求，部分分词结果示例如表 2 所示。

Table 1. Descriptive analysis of raw data from popular videos
表 1. 热门视频原始数据的描述性分析

	参数	最小值	最大值	均值	标准差
头部账号 (N = 450, 41.7%)	粉丝数	1,002,000	160,000,000	5779673.3	10,369,334
	点赞数	1000	3985,000	93645.113	271096.9
	评论数	37	277,000	9517.8111	25605.87
	转发数	12	1,651,000	35661.262	130771.2
	视频时长(s)	5	2394	131.7	190.1087
中部账号 (N = 311, 28.8%)	粉丝数	104,000	995,000	403077.17	267850.4
	点赞数	484	671,000	41281.852	84671.78
	评论数	74	67,000	5018.01	9769.545
	转发数	22	623,000	17599.315	61938.4
	视频时长(s)	6	2721	157.26	326.265
尾部账号 (N = 316, 29.3%)	粉丝数	41	99,000	20268.345	26997.12
	点赞数	189	1,072,000	42136.63	106864.4
	评论数	43	118,000	4717.709	10676.98
	转发数	0	954,000	17981.582	65109.71
	视频时长(s)	5	1789	92.62	168.733

Table 2. Example of word segmentation results for video content
表 2. 视频内容分词结果示例

原文本	分词结果
今天我接到邀请来到北京亦庄小米汽车工厂来带大家看一看小米如何造车的这可能是我见过的最大的汽车工厂了整个园区相当于故宫的面积每隔七十六秒就有一台新车从这里开出小米 SU7 就是诞生在这里	接到 邀请 北京 亦庄 工厂 看一看 造车 工厂 园区 相当于 故宫 面积 七十六 秒 一台 新车 诞生

在 BTM 主题建模中运用瑞丽熵(Renyi entropy)来判断最合适的主题数，通常选择瑞丽熵为最小时的 主题数。热门视频内容主题建模的瑞丽熵随着主题数增加的趋势图如图 1 所示。可以看到当主题数达到 6 时，瑞丽熵最小，因此确定主题数为 6。

在采用 BTM 主题建模方法深入分析后，本文成功提取了各主题下最显著的前十个关键词。基于这些关键词及对应的热门视频案例，本文系统地对各主题进行了命名与总结，如表 3 所示。值得注意的是，主题 2 “外观设计” 凭借其独特的吸引力，赢得了最广泛的创作者关注，共关联了 311 条热门视频。外观设计作为产品留给消费者的第一印象，是视频中频繁分享与讨论的焦点。关键词中“保时捷”的频繁出

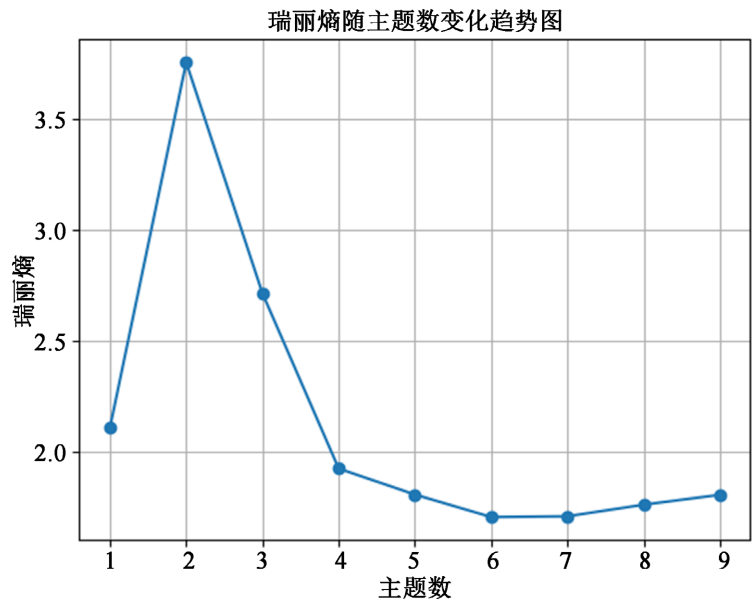


Figure 1. Trend chart of Rényi entropy variation with number of topics
图 1. 瑞丽熵随主题数变化趋势图

Table 3. Keywords of popular video content topics
表 3. 热门视频内容主题关键词

主题名	主题 1 乘坐空间	主题 2 外观设计	主题 3 驾驶安全	主题 4 智能生态	主题 5 性价比	主题 6 驾驶感受
视频数	117	311	143	163	262	89
主题词	空间	保时捷	刹车	功能	价格	运动
	后排	喜欢	测试	屏幕	续航	刹车
	喜欢	颜色	碰撞	小爱同学	特斯拉	驾驶
	特别	好看	赛道	模式	版本	底盘
	座椅	价格	技术	驾驶	公里	动力
	冰箱	发布会	造车	生态	配置	轮胎
	防晒	品牌	系统	显示	时间	加速
	智能	特别	速度	调节	充电	性能
	腿部	外观	轮胎	智能	电池	悬架
	后备箱	设计	性能	冰箱	加速	操控

现，反映了小米 SU7 在设计语言上与保时捷产品的某种相似性，这一特征在视频中被广泛提及。此外，“颜色”“外观”“设计”等词汇的反复出现，进一步凸显了消费者对汽车外观美学及个性化需求的重视。紧随其后的是主题 5“性价比”，该主题关联的视频数量达 262 条，位居第二。小米产品长期以来以其高性价比著称，而 SU7 作为小米进军新能源汽车市场的首款产品，其性价比自然成为消费者考量的重要因素。主题 4 聚焦于小米 SU7 的智能配置，体现了消费者对智能科技的高度关注。主题 3 则针对 SU7 作为一款电动汽车的安全性进行探讨，视频内容涵盖了刹车、电池及碰撞等多个方面的安全性测试，彰显了消费者对小米汽车这一新兴品牌在车辆安全性能方面的深切关注。此外，主题 1 关注于汽车的乘

坐空间，而主题 6 则侧重于驾驶感受的分享。这些主题的设置与分析，不仅揭示了小米 SU7 上市初期热门视频内容的多元化特征，也反映了视频创作者对汽车产品不同维度的创作，更是反映出消费者的期待与需求。

4.3. 消费者视角

在深入分析小米 SU7 上市初期评论数据的过程中，本文沿用了前节所述的预处理方法，并借助 SnowNLP 工具进行了情感倾向的细致剖析，部分评论数据分词及情感值计算结果示例如表 4 所示。结果显示，积极情感评论高达 406,396 条，远超负面情感评论的 43,361 条，彰显出公众对小米 SU7 上市的普遍正面情感态度。为进一步探究这一情感倾向背后的动因，本文分别制作词云图，直观展现了评论中的关键词汇，从而揭示了消费者的多维度关注点。

Table 4. Example of word segmentation and sentiment value results for comment texts

表 4. 评论文本分词及情感值结果示例

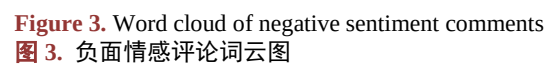
原评论文本	分词结果	情感值
不要太相信智驾，方向盘在自己手上才是真的，机器始终是机器，时间久了，或者一些情况下它也会犯错	不要 相信 智驾 方向盘 手上 机器 始终 机器 时间 久 情况 犯错	0.140455
小米汽车真正吸引我的地方是：在百万豪车面前也不会自卑	吸引 百万 豪车 面前 自卑	0.995616
小米颜值高，配置高，性价比高，物超所值	颜值 配置 性价比 物超所值	0.999615
我的小米 su7，提车第二天车机卡死，第三天自动泊车撞了	提车 车机 卡死 自动 泊车 撞	0.280745



Figure 2. Word cloud of positive sentiment comments

图 2. 积极情感评论词云图

如积极情感评论词云图，图 2 所示，“喜欢”“支持”“赞”及“加油”等词汇的显著出现，映射出消费者对小米 SU7 寄予厚望，并愿意以实际行动支持其汽车新品的推出，这种情感基础或可归因于小米品牌在智能手机和家电的良好口碑与成功积累。同时，“体验”“试驾”等词汇频繁现身，凸显了消费者对小米 SU7 实际使用感受的高度关注，预示着其在用户体验方面可能具备的竞争优势。技术层面，“智能”“技术”“性能”及“续航”等词汇的聚集，表明消费者对小米 SU7 在智能驾驶、电池技术等方面的创新与性能抱有浓厚兴趣。此外，“奔驰”“保时捷”等高端品牌的提及，既体现了消费者对小米 SU7 代表中国国产新能源电动车与国外传统汽车品牌竞争的期待，也折射出对国产汽车崛起的自豪与支持。价格与性价比方面，“性价比”“便宜”等词汇反映出消费者对价格敏感，期待高性价比产品的心理。而“外观”“颜色”及“颜值”等词汇的强调，则显示出消费者对小米 SU7 外观设计美学的认可。情感与



导,体现了对小米品牌的信任与期待,以及对SU7技术创新、性价比及用户体验的认可;负面评论虽少,但针对价格、性能及质量的反馈对小米后续改进具有参考价值。本文研究揭示了视频社交平台在新品上市信息传播中的优势,提供了消费者真实需求的第一手资料,有助于企业精准把握市场动态,优化产品与市场策略,促进企业与消费者的良性互动与共同发展。

参考文献

- [1] 邓倩. 新媒体营销研究综述与展望[J]. 科学决策, 2020(8): 67-88.
- [2] Gibbs, C. (2007) Short-Form May Be Long-Tail for Mobile Video. RCR Wireless News.
- [3] Li, Z. and Zhang, J. (2023) How to Improve Destination Brand Identification and Loyalty Using Short-Form Videos? The Role of Emotional Experience and Self-Congruity. *Journal of Destination Marketing & Management*, **30**, Article 100825. <https://doi.org/10.1016/j.jdmm.2023.100825>
- [4] 朱杰, 崔永鹏. 短视频: 移动视觉场景下的新媒介形态——技术、社交、内容与反思[J]. 新闻界, 2018(7): 69-75.
- [5] Qian, L. (2021) Analysis of Short Video Marketing Strategy under the Background of Social E-Commerce. 2021 *2nd International Conference on E-Commerce and Internet Technology (ECIT)*, Hangzhou, 5-7 March 2021, 20-25. <https://doi.org/10.1109/ecit52743.2021.00012>
- [6] 张跃先, 王雪莹. 旅游短视频营销对顾客融入的影响机制研究[J]. 管理学报, 2024, 21(2): 261-268.
- [7] 徐同谦, 王志轩. 种草类短视频对消费者购买意愿的影响机制[J]. 中南民族大学学报(人文社会科学版), 2023, 43(8): 151-158, 187.
- [8] 鲁楠. 抖音短视频中的网络营销策略[J]. 山西财经大学学报, 2024, 46(S1): 142-144.
- [9] Ma, T., Li, J., Liang, X., Tian, Y., Al-Dhelaan, A. and Al-Dhelaan, M. (2019) A Time-Series Based Aggregation Scheme for Topic Detection in Weibo Short Texts. *Physica A: Statistical Mechanics and Its Applications*, **536**, Article 120972. <https://doi.org/10.1016/j.physa.2019.04.208>
- [10] Gavhane, S.A., Bhadave, S.B. and Vengatesan K., (2019) Review on Latest Trending Topic Detection in Twitter with Stream Processing (Using Fission Pattern). *International Journal of Applied Evolutionary Computation*, **10**, 43-47. <https://doi.org/10.4018/ijaec.2019040106>
- [11] Lu, Z., Lin, Y., Huang, X., Xiong, N. and Fang, Z. (2016) Visual Topic Discovering, Tracking and Summarization from Social Media Streams. *Multimedia Tools and Applications*, **76**, 10855-10879. <https://doi.org/10.1007/s11042-016-3877-1>
- [12] Cheng, X., Yan, X., Lan, Y. and Guo, J. (2014) BTM: Topic Modeling over Short Texts. *IEEE Transactions on Knowledge and Data Engineering*, **26**, 2928-2941. <https://doi.org/10.1109/tkde.2014.2313872>
- [13] 蔡增玉, 韩洋, 张建伟, 等. 基于 SnowNLP 的微博网络舆情分析系统[J]. 科学技术与工程, 2024, 24(13): 5457-5464.