

基于OLDA模型的社会热点事件演化路径研究

周莎莎

成都信息工程大学统计学院, 四川 成都

收稿日期: 2024年11月23日; 录用日期: 2024年12月19日; 发布日期: 2024年12月25日

摘要

社会热点事件与网络舆情有着紧密联系, 已经成为影响社会稳定的主要因素之一, 热点事件应变能力反映了网络社会波动的相对稳定能力。本文结合社会热点事件和网络舆情的特征和演化规律, 构建社会热点演化指标体系, 并对数据进行赋权和标准化等处理, 通过灰色关联分析方法得出数据之间的关联度, 并进一步根据聚类中心, 将社会热点演化等级分为轻微级、一般级和重大级三类, 最后利用在线主题模型分析总结事件发展各阶段的主题词, 根据主题词推算社会热点事件的演化路径。

关键词

社会热点事件, K-Means聚类, OLDA主题模型

Research on the Evolution Path of Social Hotspot Events Based on OLDA Model

Shasha Zhou

School of Statistics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Nov. 23rd, 2024; accepted: Dec. 19th, 2024; published: Dec. 25th, 2024

Abstract

Social hotspot events are closely related to network public opinion, which has become one of the main factors affecting social stability, and the adaptability of hotspot events reflects the relative stability of network social fluctuations. Combined with the characteristics and evolution laws of social hotspot events and network public opinion, this paper constructs an index system for the evolution of social hotspots, and processes the data with weighting and standardization, obtains the correlation degree between the data through the grey correlation analysis method, and further divides the evolution level of social hotspots into three categories according to the clustering center: slight, general and major, and finally uses the online topic model to analyze and summarize the topic

words at each stage of event development, and deduces the evolution path of social hot spots according to the subject words.

Keywords

Social Hotspot Events, K-Means Clustering, OLDA Topic Mode

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

随着互联网的迅速发展,网络已经成为人们与丰富多彩的现实世界的连接口之一。据第 53 次《中国互联网络发展状况统计报告》,截止 2023 年 12 月,我国网民总量达 10.92 亿人,互联网普及率达 77.5%。互联网具有虚拟性和开放性,这就导致了热点事件,很容易在网民中扩散,在短时间内快速传播。此外合理的推动社会热点事件,有助于维护网络安全,推动社会发展。本文通过 OLDA 主题模型对社会热点事件各时间片的主题词进行追踪,并以此作为社会热点事件路径演化的关键依据,判断社会热点的发展情况,一定程度上指导相关部门做出相应决策。

2. 研究现状

余本功(2017) [1]利用改进 OLDA 模型解决新旧主题混合问题,降低冗余词的概率,更能明确地解释话题。结果表明,改进的 OLDA 模型更为有效地对话题进行检测及演化分析。裴可锋(2017) [2]在 OLDA 模型的基础上提出了一种可变在线 LDA 模型,构建主题相似度矩阵,明确主题变化关系,设计动态权重计算方法及先验参数优化方法。李慧(2019) [3]利用 OLDA 可以实时地追踪热点话题演化的优势以及微博的“话题标签”的特性,提出适合微博的热点话题演化模型 LOLDA,从话题内容和强度两方面分析了话题演化规律,改进的 LOLDA 模型可以准确地发现微博话题演化规律,通过实验验证了该模型较传统模型具有更好地泛化能力。戴长松(2020) [4]基于 LDA 算法对话题动态建模,在变分推断主题模型基础上建立衡量话题内容和热度变化的流行话题模型,对新闻媒体平台的重大事件进行话题演化建模分析。

3. 社会热点演化路径分析

3.1. 样本选择与数据来源

由于社会热点与网络舆情具有传播速度快、爆发性强、反应时间短的特点,本文对社会热点的演化指标的确定,一方面指标数据获取的难易程度,另一方面考虑指标科学性,最终将热点热度、热点强度、热点变化度、热点主体设为最终一级指标,下属 19 个二级指标。“哈尔滨冰雪旅游”这一事件作为最近具有代表性和广泛影响力的社会热点事件被选作为本文案例样本,通过 Python 爬虫技术抓取了新浪微博和百度两个平台 2023 年 12 月 20 日~2024 年 2 月 17 日共 5578 条数据,由于各指标信息量不同,本文运用熵权法对各指标赋予不同权重,得到赋权后数据并进行下文分析。

3.2. 社会热点演化等级监测

原始数据准备得当后,需要得出各数据之间的关联程度。在灰色关联分析方法中,大部分情况选择各数组中的最大值作为参考序列,也可以根据需要自主选择一个序列。因此本文将参考序列定义为各指

标字段中的最大值，并且将采集到的 61 个时间片的详细数据设定为比较序列，最后利用 Python 进行灰色关联分析得到比较序列与参考序列之间的关联度。接着利用 K-Means 聚类算法，在计算得到的 61 的关联度的基础上，进行聚类分析，从而确定社会热点事件演化等级分级阈值，并将社会热点事件演化等级分为轻微级别、一般级别和重大级别三个等级，结果如下表 1。

Table 1. Correlation degree clustering intervals
表 1. 关联度聚类区间

		聚类类别		
预警等级	轻微级	一般级	重大级	
聚类区间	[0.63507, 0.72231]	[0.74582, 0.79797]	[0.80221, 0.86440]	

本文对社会热点事件各个时间段演化等级的区分采用不同的颜色。轻微级用“绿色”来描述，此时社会热点事件演化等级较低，舆论传播速度可控，相关部门要有识别舆论发展的能力，尽早关注，提前适当引导。一般级用“黄色”来描述，此时舆论传播速度较快，相关部门需要高度关注社会热点事件演化，并进行严密监测，采取多元多方位的引导策略，既能避免引导失效带来的不利影响，又能从多方位促进社会热点事件发展，从而促进社会多元化发展。重大级用“红色”来描述，此时网络舆论传播速度极快，往往是舆论风向转化的关键时期，需要相关部门应该对舆情快速做出反映，控制社会热点事件发展方向。“哈尔滨旅游”事件的社会热点演化等级如下图 1。

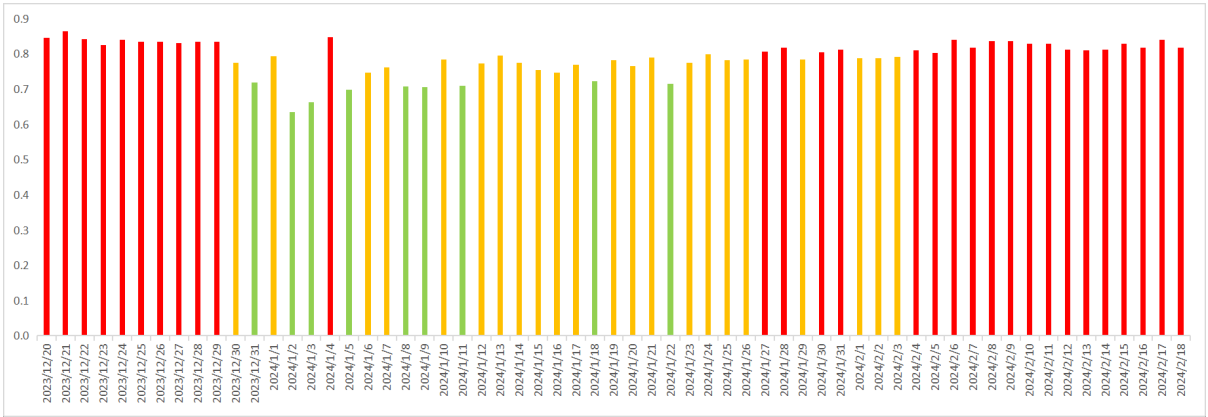


Figure 1. Bar chart of correlation degree and evolution level
图 1. 关联度与演化等级条形图

由上图 1 可以看出，“哈尔滨旅游”事件只有较少一部分时间片处于轻微级，且经过自然生长或者合适引导策略与手段能够将社会热点事件演化等级转化为一般级和重大级；同时社会热点事件演化等级变化并不遵守渐变规律，存在突变点，比如 2024 年 1 月 3 日~2024 年 1 月 4 日的时间片区中演化等级就由轻微级直接变为重大级，因此相关部门在处理社会热点事件过程中，除了时刻严密关注外，识别潜在舆论转变契机也非常重要。

3.3. 基于 OLDA 模型的社会热点路径分析

针对“哈尔滨旅游”事件在各个时间片下的不同预警等级变化的作用机制，本文通过 OLDA 主题模型提取各时间片下主题内容，从中研究该舆论事件的演变规律，结合实际情况作出判断，给社会热点事件演化规律提供一定参考。“哈尔滨旅游”事件各个时间片内微博的条数信息如下图 2。

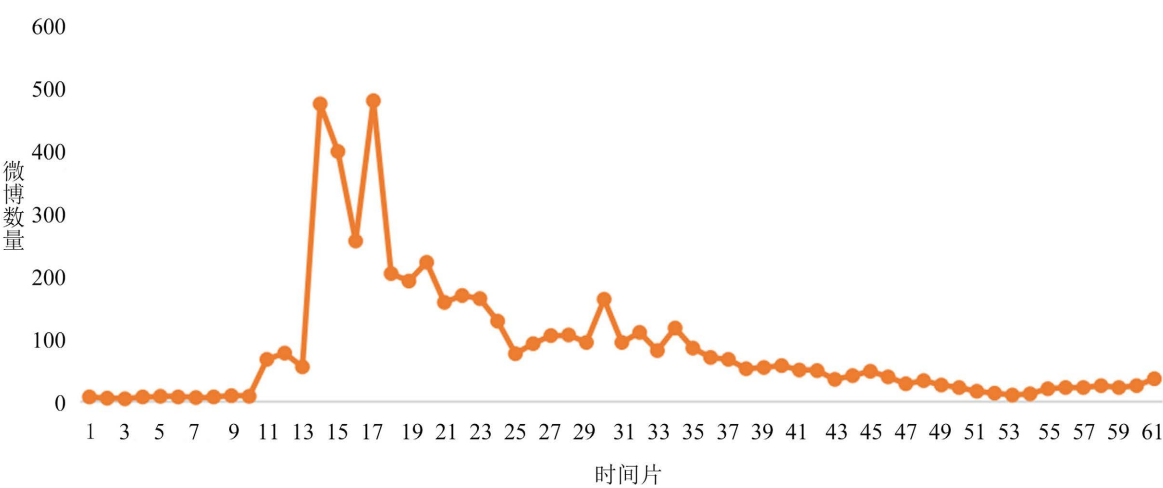


Figure 2. Line chart of the number of microblogs in each time of the “Harbin Tourism” incident

图 2. “哈尔滨旅游”事件各时间片内微博条数折线图

Table 2. Examples of sample subject terms for “Harbin Tourism”

表 2. “哈尔滨旅游”事件样本主题词示例

时间片	预警等级	主题词
1	重大级	冰雪大世界、事件、音乐节、现场、搭子、雕刻、进场、轿厢、路灯、世界
2	重大级	冰雪大世界、门票、铁路、游玩、影响、免费、专生、朋友、有空、专四
3	重大级	文化、旅行、雾凇、旅游局、长廊、银装素裹、酒店、退票、艺术界、美食
4	重大级	活动、剧本、餐厅、艺术节、氛围、美食、建筑、冰宫、体验、爱好者
5	重大级	冰雪大世界、企鹅、现场、极地馆、排队、大家、免费、木屋、星空、日落
6	重大级	冰棍、开学、面包、运动量、特色、星光、大雪、商品、西餐厅、舞蹈
7	重大级	地铁、免费、马铃薯公主、旅行、冰雪大世界、老铁、老客、美食、网约、结果
8	重大级	旅行、体验、童话世界、私家车、游客、步伐、极具、本地人、星星、下馆子
9	重大级	雪乡、旅行、搭子、特种兵、江鱼、服装、朋友、景点、教堂、一家子
10	重大级	雪乡、音乐、梦幻、音乐家、酒店、蘑菇、美食、童话、积雪、电话
11	一般级	南方小土豆、雪乡、本地人、冰雪大世界、美食、免费、中华、老板、热情、建筑
...	...	...
50	重大级	携程、视频、旅行、游客、来源、南方小土豆、冰雪大世界、鄂伦春族、民族、横道河子
51	重大级	旅行、风光、摄影、地点、锅包肉、钻石、鄂伦春族、童话世界、火烧云、世界
52	重大级	口音、旅行、印象、刻板、诱因、热度、气温、小品、姐妹、大家
53	重大级	纪实、摄影、小品、家庭、文旅、花式、流量、旅行、新车、复兴号
54	重大级	南方小土豆、教堂、调查、花式、柚子、文旅、住宿、中心、时代、数据
55	重大级	摄影、冰雪大世界、水彩、水彩画、美食、南方小土豆、单身、上线、绘画、秘密
56	重大级	冰雪大世界、历史、会场、世界、遗留、融化、大家、人民、部队、舞狮
57	重大级	冰雪大世界、情人节、动物、流浪狗、流浪、教堂、救助、滑梯、情感、世界
58	重大级	冰雪大世界、铁路、教堂、工农子弟、革新、机票、旅行、大家、博物馆、体验
59	重大级	携程、冰雪大世界、乐园、世界、样子、旅行、南方小土豆、酒店、美食、天气
60	重大级	摄影、冰雪大世界、早市、旅行、美食、状态、脉动、排队、啤酒、南方小土豆
61	重大级	手机、冰雪大世界、美食、教堂、行程、花花、排队、啤酒、司机、口罩

据上图2及表2相关主题词可得如下结论：(1)“哈尔滨旅游”事件在第一到第十时间片处于爆发初期阶段，演化等级达到重大级，虽然微博数量较少，但网民讨论的主题词大致为冰雪大世界、门票、退票和免费等涉及地方特色和财产等话题，迅速在网民之中传播；(2)“哈尔滨旅游”事件在第十一到第二十一时间片处于爆发盛期阶段，演化等级大部分处于一般级和轻微级，加之处于年度假期，社交平台上的微博数量也最多，而网民热议的话题则包括冰雪大世界、景点、旅行、南方小土豆和公主等具有地方特色、幽默和互联网热梗等话题，在网民中引起迅速传播，且更多的网民在社交网络直接发表相关社会热点意见，与爆发初期的观望状态大不相同，传播速度更快；(3)“哈尔滨旅游”事件在第二十二到第三十五时间片处于稳定期阶段，演化等级处于一般级，网民所参与的微博数量较多，其中网民热议话题主要为冰雪大世界、文旅、热情、雪乡和冻梨等结合地方特色与旅行体验的内容，传播速度虽快，但事件热点波动较小；(4)“哈尔滨旅游”事件在第三十六到第六十一时间片处于衰退期阶段，虽然网民对该事件的关注度较高，但随着时间增加事件逐渐常态化，网民在社交平台上的分享内容也有所减少，网民的热议话题也由上述主题转变为冰雪大世界、行程、博物馆和排队等。

4. 结论与建议

4.1. 结论

社会热点事件对社会发展有着重要影响，因此研究社会热点事件的演化路径非常重要。本文通过对社会热点事件演化路径进行研究得出以下结论：(1)社会热点事件发展有着不同的等级，大致表现为潜伏期、发展期、爆发期、稳定期和衰退期，其中发展期是社会热点事件的关键时期，相关单位部门应当高度注意此期间的社会热点发展；(2)社会热点事件爆发期一般处于重大级的演化等级，此时需要积极正确的引导，在发展期阶段，预警等级一般为轻微级和一般级，在社会热点事件的成熟阶段，预警等级一般为重大级；(3)在社会热点事件中，社会热点事件的演化等级在一定程度上由社会热点的主题内容决定，本文通过主题模型分析“哈尔滨旅游”事件的主题词可以发现，整个舆情事件整体围绕“冰雪大世界”等词语展开，讨论话题在发展期之后网民的关注点越集中，讨论的话题也较为集中。

4.2. 建议

本文透过“哈尔滨旅游”事件，建立 OLDA 主题模型，探讨社会热点事件的演化路径，结合相关单位基本情况，提出以下建议：由于社会热点事件有着三个演化等级，且存在发展过程，通过对社会热点事件演化路径进行分析，采集相应的网友的讨论话题，集中关注网友重点讨论话题，从中判断社会热点事件的性质、情感方向和敏感度，并加以预防和引导，采取相应的对策；为了保障网络社会的安全性和稳定性，各大网站应当构建并实施相关规章制度，对网友发布的信息进行规范管理，同时组建专业监控团队，对平台上的信息进行监督和管理，保证网友能自由发表言论的同时规范语言的规范性，防止社会热点事件对社会安定产生不利影响，同时进一步促进社会发展。

参考文献

- [1] 余本功, 张卫春, 王龙飞. 基于改进的 OLDA 模型话题检测及演化分析[J]. 情报杂志, 2017, 36(2): 102-107.
- [2] 裴可锋, 陈永洲, 马静. 基于 OLDA 的可变在线主题演化模型[J]. 情报科学, 2017, 35(5): 63-68.
- [3] 李慧, 王丽婷. 基于话题标签的微博热点话题演化研究[J]. 情报科学, 2019, 37(1): 30-36.
- [4] 戴长松, 王永滨, 王琦. 基于在线主题模型的新闻热点演化模型分析[J]. 软件导刊, 2020, 19(1): 84-88.