



情报杂志

Journal of Intelligence

ISSN 1002-1965, CN 61-1167/G3

《情报杂志》网络首发论文

题目：基于情感挖掘的网络舆情预警研究
作者：臧振春，李焕，崔春生
网络首发日期：2025-05-15
引用格式：臧振春，李焕，崔春生. 基于情感挖掘的网络舆情预警研究[J/OL]. 情报杂志. <https://link.cnki.net/urlid/61.1167.G3.20250514.1704.010>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于情感挖掘的网络舆情预警研究*

臧振春 李 焕 崔春生

(河南财经政法大学 数据科学与电子商务学院 郑州 450046)

摘要: [研究目的] 为防止网络事件演化发展为网络舆情,构建网络舆情预警模型识别舆情演化关键节点并量化风险,为政府部门及时采取防控措施遏制舆情发酵提供决策支持。[研究方法] 基于微博平台实时舆情数据,整合数据挖掘、可视化分析和情感分析技术,构建多维度预警指标体系:首先通过关键词分析捕捉事件核心争议点,继而运用SnowNLP情感分析进行情感分类并计算情绪强度,综合热度分析从爆发指数(EI)、情绪指数(SI)、传播指数(DI)和搜索引擎指数(SEI)四个维度构建舆情危险指数(HI)量化舆情发展态势,并在危险指数超出预设阈值时自动触发预警机制。[研究结果/结论] 通过对“李佩霞事件”进行分析,模型不仅准确捕捉了网民情绪和事件核心争议点,而且及时发出了预警。预警结果与事件实际走向高度一致,验证了模型在网络舆情事件预警方面的有效性。

关键词: 网络舆情;舆情预警;情感分析;数据挖掘;危险指数

中图分类号:G206

Research on Network Public Opinion Early Warning Based on Emotion Mining

Zang Zhenchun Li Huan Cui Chunsheng

(Henan University of Economics and Law, Zhengzhou 450046)

Abstract: [Research purpose] In order to prevent network events from evolving into network public opinion, this paper builds an early warning model of network public opinion to identify key nodes of public opinion evolution and quantify risks, so as to provide decision support for government departments to take timely prevention and control measures to curb public opinion fermentation. [Research method] Based on the real-time public opinion data of Weibo platform, this paper integrates text mining, visual analysis and emotion analysis technologies to build a multi-dimensional early warning index system: firstly, the model captures the core controversial points of events through keyword analysis, then uses SnowNLP emotion analysis to classify emotions and calculate the emotional intensity, and integrates heat analysis to construct a public opinion hazard index from four dimensions: explosion index (EI), sentiment index (SI), dissemination index (DI), and search engine index (SEI) to quantify the development trend of public opinion and automatically trigger the early warning mechanism when the risk index exceeds the preset threshold. [Research result/conclusion] Through the analysis of the 'Li Peixia incident', it can be seen that the model not only accurately captured the public sentiment and the core issues of the event but also issued timely alerts. The warning results are highly consistent with the actual trend of the event, verifying the effectiveness of the model in early warning of network public opinion events.

Key words: network public opinion; public opinion warning; sentiment analysis; data mining; hazard index

基金简介: 2023 年度教育部人文社会科学研究规划基金项目:不确定视域下网络正能量生成机理及传播机制研究(编号:23YJA860004);2024 年度河南省高等学校哲学社会科学基础研究重大项目:基于不确定理论的网络正能量生成、演化及传播秩序研究(编号:2024-JCZD-27);2024 年教育部人文社会科学研究项目:网络舆情脉络演进、全息图谱及治理策略研究(编号:24YJA860023);2025 年度河南省教育厅人文社会科学研究一般项目:基于脉络结构-演化机理的网络舆情治理措施研究(编号:2025-ZZJH-332)研究成果。

作者简介:臧振春,男,1962 年生,博士,教授,博士生导师,研究方向:舆情预警;李 焕,女,2002 年生,硕士,研究方向:舆情预警;崔春生,男,1974 年生,博士,教授,硕士生导师,研究方向:舆情预警。

通信作者:崔春生

0 引言

网络舆情源于网络突发事件,这些事件会在网络中掀起广泛讨论,一些倾向性很强的意见、态度与意愿逐渐聚集,最终形成网络舆情^[1]。从网络事件到网络舆情的演变,是一个复杂、动态的过程,往往起始于某个具体的网络触发点,如一条引人关注的微博、一段引发争议的视频,或是一次热点话题的讨论。在这个过程中,及时发现舆情发展的苗头,控制舆情的发展显得尤为重要。

舆情被视为公众在特定时间和空间内,对公共事务、社会现象或特定事件所持有的群体性情绪、意愿、态度、意见的总和及其表现^[2]。网络舆情的发展趋势呈现出多元化、智能化和理性化的特点^[3]。目前针对网络舆情预警的研究,主要围绕舆情信息收集、舆情演进特征、舆情分析检测系统构建等方面展开。在舆情信息收集方面,唐雨霞^[4]等通过数据挖掘与文本分析技术构建一套网络舆情预警系统,可对网络言论进行分析和预警;王旭娜^[5]等根据重大突发事件的参与主体与内外部因素之间的复杂交互关联关系,构建包括用户、时序、观点、情绪四层子网的超网络模型,基于对超网络的分析,提出公众认知挖掘方法;余铭楷^[6]等利用网络爬虫技术爬取国内舆情排行网站的热点信息,通过关键字检索对网民的评论意见以及发表内容进行全面分析,从而实现网络舆情的动态监测。在舆情演进特征方面;张少芳^[7]等从网络舆情的演化机理出发,通过构建网络舆情预警和溯源机制,能够对网络舆情风险进行早期识别和预测并锁定舆情主体;曾倩倩^[8]等提出一种基于直觉模糊的突发事件网络舆情风险预警方法,通过网络态势分析得出突发事件网络舆情演化过程,根据专家经验确定事件影响程度和传播速度,直觉模糊化处理推理条件以及网络舆情风险等级,计算推理结果和风险等级直觉模糊子集的贴近度,将最大贴近原则作为风险等级判定依据,实现突发事件网络舆情风险预警。在舆情分析检测系统构建方面,杨达吴^[9]在网络舆情负性情绪的视域下构建网络舆情负性情绪预警体系的目标,在此基础上提出建立网络舆情负性情绪预警体系的相关对策建议;崔骥^[10]通过研究消极情绪主题词实现对负面舆情事件统计和量化,实现主题词对应的负面舆情事件的早期预警;刘静^[11]构造 AHP-TOPSIS 模型,通过层次分析法、专家咨询法为指标赋权,从舆情强度、舆情热度以及舆情影响三个维度建立了高校网络舆情监测指标体系。

这些方面的研究都可以深入地了解网络舆情的形成、发展和影响,为政府、企业或组织提供决策支持,以更好地管理和引导网络舆情。但是由于互联网信息的

海量和复杂性,大部分舆情预警系统都无法准确、高效的处理和分析数据;其次现有的研究多集中于特定领域的网络舆情预警,缺乏一套适用于各类网络舆情的预警体系。本文聚焦网络舆情的预警,通过建立信息监测机制,精准捕捉微博社交平台上与突发事件相关的关键数据,并对舆情数据进行可视化分析与情感分析,构建网络事件的情感指数,作为评估舆情态势的重要指标。综合情感指数与搜索引擎指数生成网络舆情的“危险指数”。当危险指数触及预设的警报阈值时,预示着舆情可能向不利方向发展,此时系统将即时触发舆情预警机制。通过这种综合分析方法,不仅提高了网络舆情预警的准确性和时效性,而且增强了对舆情动态的理解和应对能力,有助于预防和减轻网络舆情可能引发的社会影响。

1 相关理论

1.1 网络舆情生命周期理论

网络舆情生命周期理论是描述公众关注性事件从萌发到消亡的阶段变化及其规律的理论。舆情周期分为潜伏期、发展期、成熟期(或称为高潮期)和消退期四个阶段^[12]。在潜伏期,网络舆情事件的潜在诱因已经存在,但尚未显化为突发事件,此时信息量小且片面模糊,网络舆情力量尚显单薄,走势不稳定^[13]。在发展期,相关信息开始增多,网民对事件的看法、态度和情绪逐渐形成,舆情开始发酵,并可能引发更广泛的关注^[14]。成熟期(高潮期)是网络能量快速集聚的阶段,此时出现意见领袖,热度大幅增加,逐渐形成舆论主导意见,舆情可能经历多次反转、对立和对峙等复杂变化,形成舆论高潮^[15]。这些变化不仅反映了公众对事件的多元解读,也体现了社会价值观的冲突和协商。最终,随着时间推移、相关部门的介入或舆情主体的积极处理^[16],公众对舆情事件的关注逐渐减少,舆情进入消退期并逐渐衰退。这一阶段标志着舆情事件的影响力和关注度的减弱,直至最终被新的社会议题所取代。

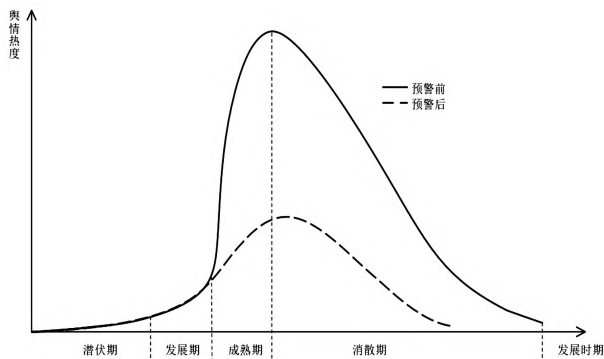


图1 预警前后舆情趋势图

由舆情生命周期理论可知,突发事件发生后,快速

跟踪监测并及早控制极为重要。通过潜伏期、发展期与成熟期关键时刻的挖掘,预测舆情的发展趋势实现有效推迟或消解舆情成熟期,降低可能对社会造成的负面影响。在舆情成熟期之前,舆情的发展呈现出迅猛的上升趋势,公众的情绪和关注度在不断升高。随着预警措施的及时介入,舆情增长势头得到了有效抑制。舆情曲线开始出现平稳甚至下降的趋势,这反映出通过有效的预警机制与防控策略,舆情的热度及其潜在的负面影响得到了逐步的缓解和弱化(如图1所示)。

1.2 SnowNLP 情感分析

SnowNLP 是一种基于 Python 的中文文本处理工具,其情感分析功能主要基于朴素贝叶斯算法。该算法通过训练模型对文本进行情感倾向判断,具体流程包括对文本进行分词处理、去除停用词,然后利用训练好的模型对文本进行情感打分。在舆情分析中, SnowNLP 具有较高的适用性。首先,它能够快速处理大量中文文本数据,适用于社交媒体、新闻网站等平台的舆情监测。其次, SnowNLP 的情感分析模型可以通过自定义训练数据进行优化,以适应特定领域的舆情分析需求。

SnowNLP 情感分析基本原理如下:假设情感分析的分类有正面评价 c_1 和负面评价 c_2 两类,且每一条评论中有 w_1, w_2, \dots, w_n 共 n 个相互独立的文本空间,通过朴素贝叶斯公式可以得出正面评价 $P(c_1 | w_1, w_2, \dots, w_n)$ 和负面评价 $P(c_2 | w_1, w_2, \dots, w_n)$ 的条件概率为:

$$P(c_1 | w_1, w_2, \dots, w_n) = \frac{P(w_1, w_2, \dots, w_n | c_1)P(c_1)}{P(w_1, w_2, \dots, w_n)} \quad (1)$$

$$P(c_2 | w_1, w_2, \dots, w_n) = \frac{P(w_1, w_2, \dots, w_n | c_2)P(c_2)}{P(w_1, w_2, \dots, w_n)} \quad (2)$$

根据全概率公式,可将 $P(c_1 | w_1, w_2, \dots, w_n)$ 表示为:

$$P(w_1, w_2, \dots, w_n) = P(w_1, w_2, \dots, w_n | c_1)P(c_1) + P(w_1, w_2, \dots, w_n | c_2)P(c_2) \quad (3)$$

即式(1)可以转化为:

$$P(c_1 | w_1, w_2, \dots, w_n) = \frac{P(w_1, w_2, \dots, w_n | c_1)P(c_1)}{P(w_1, w_2, \dots, w_n | c_1)P(c_1) + P(w_1, w_2, \dots, w_n | c_2)P(c_2)} \quad (4)$$

$P(c_1 | w_1, w_2, \dots, w_n)$ 表示在已知文本中出现了单词 w_1, w_2, \dots, w_n 的情况下,该文本属于积极情感类别 c_1 的概率。这个概率值的范围在 0 到 1 之间,其中:当 $P(c_1 | w_1, w_2, \dots, w_n)$ 越接近于 1,表示给定文本属于积极情感类别的概率越高,即文本更可能表达的是积

极情绪;当 $P(c_1 | w_1, w_2, \dots, w_n)$ 越接近于 0,表示给定文本属于积极情感类别的概率越低,即文本更可能表达的是消极情绪。

2 模型与算法

在突发事件发生后,对网络舆情进行监测与预警。具体流程如图2所示:第一步,对微博平台进行数据采集与数据预处理;第二步,提取数据关键词并进行可视化分析;第三步,结合情感分析和事件热度分析构建舆情危险指数;第四步,设定危险指数阈值,危险指数大于阈值时立即进行舆情预警。

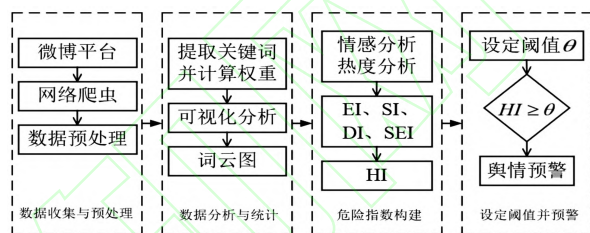


图2 模型流程图

2.1 数据收集与预处理

运用 Python 语言编写网络爬虫程序,在微博平台上迅速捕获热度较高的微博帖子内容及其关联的在线评论数据。然后利用 Jieba 中文分词库,对采集到的评论数据进行分词、移除停用词等处理以降低数据噪声,提炼出更加规范、易于分析的评论数据集。通过以上流程,确保所收集数据的时效性与准确性,并为后续的数据分析工作奠定基础。

2.2 数据分析与统计

为了从网民的视角梳理突发事件的焦点,首先利用 TextRank 算法对经过预处理的数据进行关键词提取。通过对文本数据中各个词语频率的统计,筛选出频率最高的关键词这些关键词能够有效反映事件中最为热议和关注的议题。随后,基于这些高频关键词进行可视化分析,生成词云图,以直观的方式揭示网民讨论的核心议题,从而更好地把握网民的情感走向和讨论趋势。

2.3 舆情危险指数构建

运用 SnowNLP 对爬取的微博评论数据进行情感倾向分析,为了提升情感识别的准确度,首先利用公开的情感分析数据集对 SnowNLP 模型进行训练,使其具备初始的情感分析准确度;其次将微博评论数据分为训练集与测试集,对训练集进行手动情感标注,明确分类为积极与消极两大情感倾向,为了降低主观性对标注结果的影响,本文采用三人独立标注的方式,并通过多数投票法确定最终的情感标签。通过计算 Cohen's Kappa 系数来验证标注者之间的一致性。只有当标注一致性达到 0.8 以上时,才采用该训练集进行模型的

二次训练。最后,使用经过验证的训练集对模型进行再次训练,以提高模型对微博评论情感倾向的识别准确性和客观性。

Cohen's Kappa 系数的公式为:

$$k = \frac{p_o - p_e}{1 - p_e} \quad (5)$$

其中, p_o 是实际一致性比例, p_e 是随机一致性比例。Cohen's Kappa 系数是一种衡量分类问题中评估者之间一致性程度的统计指标。它通过比较实际一致性与随机一致性之间的差异来评估评估者之间的一致性水平。当 k 大于 0.8 时代表了较高的一致性水平。

舆情危险指数 (Hazard Index, HI) 通过综合考量舆情信息的多个维度,将复杂的舆情信息转化为直观的数值,便于实现实时监测和预警。在进行预警之前,需要构建一个合理的指标体系,以便从多个角度衡量舆情风险^[17]。网络突发事件的产生和传播存在明显的负性偏向^[18],即消极情绪在舆情扩散过程中的影响力更为显著,人们更倾向于关注负面信息、表达负面情绪、持有负面态度等^[19]。因此在评估舆情潜在风险时消极情绪是不可忽视的关键因素。为了深入了解舆情消极情绪主题信息,基于情感分析的结果结合爆发指数 (Explosion Index, EI)、情绪指数 (Sentiment Index, SI)、传播指数 (Dissemination Index, DI) 三个指标以反映消极情绪主题的影响程度,其中 EI 量化舆情事件的爆发速度和严重性,提供对事件紧急性和影响范围的直接度量;SI 反映网民对舆情事件情绪变化趋势,揭示公众情绪的动态演变;DI 衡量消极情绪的传播范围和影响力,指示消极情绪在网络空间的扩散情况。公众对突发事件的关注度也是推动舆情扩散的重要因素,搜索引擎指数 (Search Engine Index, SEI) 是网民追逐和关注热点事件行为的体现,通过监测特定关键词或话题在搜索引擎中的搜索频率和关注度为评估舆情事件的热点和焦点提供了量化视角。危险指数的设定直接关系到舆情态势的界定,通过设定具体的阈值可以明确识别何时需要采取应对措施,从而有效管理和控制舆情风险,确保及时响应并减轻可能的负面影响。综合以上四个关键指标以评估网络舆情的潜在风险,得到:

$$HI = \omega_1 EI + \omega_2 SI + \omega_3 DI + \omega_4 SEI \quad (6)$$

$\omega_1, \omega_2, \omega_3, \omega_4$ 分别为各指标的权重。通过将各指标的加权值相加,得出综合反映网络舆情风险水平的危险指数。这种方法不仅确保了评估过程中的客观性和公正性,而且提高了模型对不同舆情情景的适应性和灵活性。通过这种方式,网络舆情危险指数能够更准确地捕捉和反映舆情的动态变化,为舆情管理和决策提供科学依据。其中, EI、SI、DI、SEI 四个指标的计

算公式如下。

(1) 爆发指数 (Explosion Index, EI)

$$EI = \frac{Text_{now}^{neg}}{Text_{past}^{neg} + Text_{past}^{pos} + 1} \quad (7)$$

其中, $Text_{now}^{neg}$ 表示当前消极情绪的评论数量, $Text_{past}^{neg}$ 表示前一段时间消极情绪的评论数量, $Text_{past}^{pos}$ 表示前一段时间积极情绪的评论数量。

EI 反映当前消极情绪文本数量相比于前一段时间文本数量的增加速度。网络舆情事件中的消极情绪主题词的出现频率通常出现异常陡增。EI 越高表明爆发速度越快,舆情事件越严重;EI 越低表明爆发速度越慢;EI 为 0 或负值时表明负面舆论不增长甚至负增长。

(2) 情绪指数 (Sentiment Index, SI)

$$SI = \frac{Text_{now}^{neg}}{Text_{now}^{pos} + 1} \times \frac{\sum_{i=1}^n S_i}{Text_{total}} \quad (8)$$

其中, $Text_{now}^{neg}$ 表示当前时间积极情绪的评论数量, S_i 表示第 i 条评论的情感得分, $Text_{total}$ 为总评论数。

SI 是指网络舆情中消极文本数量与积极文本数量的比值与情感强度因子的乘积,SI 反映当前社交网络用户对舆情事件情绪变化趋势。SI 越高,表明该事件当前消极情绪人群比例比前一时段更高,引发了大量的消极情绪和情感,舆情越严重。

(3) 传播指数 (Dissemination Index, DI)

$$DI = \frac{Text_{now}^{neg}}{Text_{past}^{neg} + 1} \quad (9)$$

DI 反映当前网络空间消极情绪传播数量比前一时段的的增长程度。消极数量较多的网络舆情影响力大,覆盖范围广,危害性高;消极数量较少的网络舆情也有可能在某些特定人群传播,不代表影响性较小。

(4) 搜索引擎指数^[20] (Search Engine Index, SEI)

$$SEI = V_{now}^{keyword} / EV_{time period}^{zone_j} \quad (10)$$

其中, $V_{now}^{keyword}$ 表示针对当前突发事件关键词的搜索数量, $EV_{time period}^{zone_j}$ 表示在突发事件发生期间的社交媒体中所有关键词的平均搜索数量。

SEI 是当前网络事件关键词搜索数量与过去一段时间搜索引擎里所有关键词平均搜索数量的比值,SEI 反映当前网络舆情关注程度,SEI 越高,表明当前网络事件受关注度越高,网络舆情越严重。

2.4 设定阈值并预警

网络舆情危险指数的阈值是网络舆情监测和分析中用于判定舆情风险级别并触发预警机制的关键临界点。为了准确设定这一阈值,以便及时干预和引导网络舆情的发展态势,本文通过模拟仿真方法,综合考虑了社会网络结构、个体属性、信息属性、心理属性和观

案件的呼声。网民对对受害者表示了同情和关注,对涉及的个人和机构的行为表示了不满和质疑,并对事件的公正处理持有强烈的期待。如下表 1 所示。这些

信息对于理解事件的公众影响和网民的态度至关重要。

表 1 “李佩霞事件”关键词分析表

关键词	分析
党委书记、县委书记、举报人、 省纪委、调查、强奸犯	事件涉及党委书记,存在性侵犯指控或举报,是事件的核心争议点。 公众讨论焦点集中在事件涉及的官员和机构上。
被害人、受害者	公众对受害者遭遇表示关注和同情。
狗咬狗、鱼死网破、见怪不怪、管不住	网民对涉及个人或机构的行为感到愤怒和不满。
办公室、公务员	事件涉及公职人员的工作场所行为,公众对公务员的行为标准有较高的期待。
互联网、调查组、处理结果、调查结果	事件在互联网上引起广泛关注,公众期待事件处理结果,可能需要成立调查组进行调查。
公信力、政府部门、官官相护	网民对事件处理的公正性和透明度有所质疑,对相关部门的公信力表示怀疑。

3.3 情感分析

运用 SnowNLP 工具对采集自微博平台的评论数据进行情感倾向分析。利用公开情感倾向数据集训练完成预训练后,依据情感倾向将微博提取评论数据的 1/3 手动分类为积极与消极两类,形成训练集。标注结果的一致性系数 $k = 0.85$,表明标注结果可靠,有效

降低了主观性影响。这一步骤旨在使模型更好地适应微博评论文本的特定语境,从而提高情感分析的准确度。训练完成后,将模型应用于收集数据的情感倾向分析,结果(部分)如表 2 所示。基于情感分析结果,绘制了情感分布直方图(图 5),展示了不同情感得分区间的评论数量分布。

表 2 情感分析结果

评论人 id	评论内容	情感倾向	情感得分
茉妮咖儿	当你看见一只蟑螂的时候……	消极	0.012217963
雍虾	希望有后续吧不要等热度没了又没后续了。	消极	0.38887196
最爱上饶	上饶市委反应迅速[good][good][good]	积极	0.784288505475263
曼联 1992	男盗女娼	消极	0.267817853
一一麻麻是个少女	为啥都在说这个女生,这么明显的是无奈,推脱。如果真的想勾引领导,就不是这样无奈祈求了。受害者不一定完美,但加害者一定有罪。	消极	0.000037554525756045

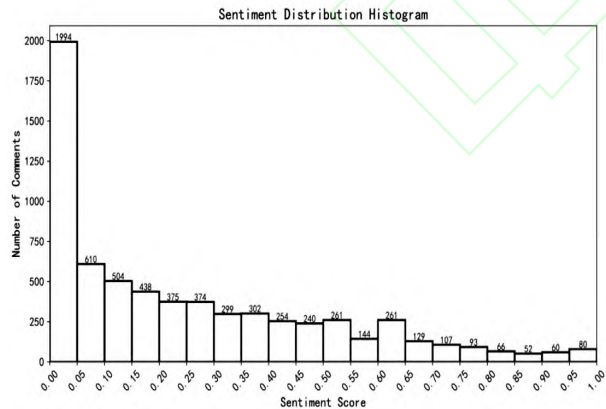


图 5 情感分布直方图

情感分析结果大多低于 0.6,网民的情绪普遍呈现消极倾向,显示负面情绪在舆论中占主导地位。消极情绪在网民评论中的显著倾向性表明大多数网民对于事件持质疑、不满和担忧的态度。结合网民评论内容,许多网民对事件的透明度和公正性表示怀疑,对涉事官员和相关部门的公信力表示不满;也有部分网民呼吁严查事件,追求真相和正义。舆情动态显示,随着事件的持续发酵,公众对事件的关注和讨论热度持续上升,对官方的回应和处理结果抱有期待,但也夹杂着

不信任和质疑。总体来看,网民情绪激烈,情感倾向以消极为主,对事件进展及处理结果保持高度关注,舆情态势紧张且走向复杂,需要相关部门及时、透明、公正地处理和回应,以缓解公众的不满情绪,恢复公信力。

3.4 危险指数及预警

为了计算网络舆情的危险指数,本文通过爬虫技术收集了 2024 年 7 月 25 日至 28 日期间微博平台上关于“李佩霞事件”的 6644 条评论。为了对比事件前后的情感变化,选取了 7 月 25 日晚至 7 月 26 日的评论数据作为基线数据,代表事件发生初期的舆情态势,而整个数据收集期内的评论则代表了当前的舆情状况。通过运用数据筛选和排序工具,从评论中提取了积极与消极情绪评论的数量,进而计算得到 EI、SI、DI,为舆情分析提供了量化的视角。利用 Python 对数据进行分类分析,结果显示,在基线期间的 2 223 条评论中,积极评论 ($Text_{past}^{pos}$) 303 条,消极评论 ($Text_{past}^{neg}$) 1 920 条;在整个数据收集期内的 6643 条评论中,积极评论 ($Text_{now}^{pos}$) 848 条,消极评论 ($Text_{now}^{neg}$) 57 95 条。基于这些数据,计算出 EI 为 2.607、SI 为 1.722、DI 为 3.017。结合微博热搜榜与云合数据,得到

“李佩霞事件”在7月28日的搜索量($V_{now}^{keyword}$)为1018087,而7月25日至28日期间微博平台主要关键词的平均搜索量($EV_{time\ period}^{zone}$)为345005。据此计算SEI为2.950。假定四个指标对网络舆情危险指数时具有均衡的影响,将每个指标赋予相等的权重,即每个参数的权重为1/4。这一权重分配方案经过向量相似度法的验证,证明其有效性。综合以上指数,得出“李佩霞事件”网络舆情危险指数(HI)为2.574。

通过多维参数融合与动态优化机制确定舆情危险指数的预警阈值。具体而言,综合考量微博平台的社会网络结构、个体属性、信息属性、心理属性及观点属性等多个参数,并基于历史舆情事件数据(2020–2023年共10起同类事件)的危险指数分布分析,通过模拟仿真确定舆情危险指数的阈值 θ 为1.75。此阈值既能综合考量各参数,又能覆盖历史事件的风险波动范围,确保模型在复杂舆情环境下的稳健性。鉴于“李佩霞事件”的危险指数显著超出该阈值,触发预警机制,需要立即采取网络舆情防控措施以应对潜在的风险。随后的监测显示,该事件在微博平台上持续引发热议,搜索量与讨论量持续上升,直至8月2日才呈现下降趋势。若采用本文所提出的模型对微博平台及类似事件进行监测,将有助于及早识别舆情发酵迹象,及时采取防控措施,从而减轻民众的负面情绪,降低社会影响,并维护政府的公信力。

4 总结与展望

本文提出了基于情感挖掘的网络舆情预警模型,通过数据挖掘和情感分析技术对突发事件发生后的微博平台数据进行实时监测和分析;构建了包含EI、SI、DI和SEI四个维度的网络舆情危险指数(HI)对舆情的发展态势进行量化;设定阈值并在危险指数超过阈值时触发预警,以便及时采取防控措施应对舆情态势蔓延。以“李佩霞事件”为案例,本文展示了模型在数据收集、关键词提取、情感倾向分析和危险指数计算等方面的应用。模型不仅有效地揭示了事件的关键要素和网民的情感倾向,而且其预警结果与事件的实际发展态势相吻合,验证了模型的有效性和实用性。未来的研究可在以下几个方面进行拓展:首先,优化情感分析算法,增强模型对复杂情感表达的识别精度,包括对隐喻、讽刺等微妙语言现象的解析能力;其次,动态调整各指标权重,以适应不同舆情事件的特点;最后,将模型应用于更广泛的社交平台,增强模型的普适性。通过这些改进,网络舆情预警模型将更加精准、高效,为政府和企业提供更有力的决策支持。

参考文献

- [1] 王芳,董绮梦.基于fsQCA的事件信息跨平台传播效果影响因素研究[J].情报理论与实践,2024,47(10):94–106.
- [2] 刁龙,赵阳,王亚飞,等.舆情对课外补习治理政策制定的影响分析——基于多源流理论的大数据分析[J].当代教育科学,2021(8):77–86,95.
- [3] 袁子桐,杨馨溟.舆情智能化的原理解析及治理优化——基于传播政治经济学的视角[J].哈尔滨工业大学学报(社会科学版),2024,26(2):133–141.
- [4] 唐雨霞,左尚扬,易业曦,等.一套基于数据挖掘技术的网络舆情预警系统开发研究[J].电脑知识与技术,2024,20(21):67–70.
- [5] 王旭娜,张立凡.基于超网络的重大突发事件公众认知演化与预警策略研究[J].情报理论与实践,2024,47(12):183–194.
- [6] 余铭楷,王涛.网络舆情监控及预警系统构建与应用研究[J].山西电子技术,2024(3):110–113.
- [7] 张少芳,李献军,刘延锋.信息生态视域下网络舆情预警、溯源及治理策略研究[J].公关世界,2024(10):151–153.
- [8] 曾倩倩,张月琴.突发事件网络舆情风险等级判定与预警仿真[J].计算机仿真,2024,41(5):281–285.
- [9] 杨达昊.自媒体时代网络舆情负性情绪的预警研究[J].新闻前哨,2024(10):51–53.
- [10] 崔骅,韩益亮,朱率率,等.基于情感主题建模的负面舆情早期预警研究[J].清华大学学报(自然科学版),2024,64(10):1771–1784.
- [11] 刘静,孙玉河,张正东,等.基于AHP-TOPSIS法的高校舆情预警方法研究[J].大数据时代,2024(3):27–30.
- [12] 张澜.网络舆论与高考政策互动关系研究——基于江苏新高考选科舆情处置案例分析[J].中国考试,2024(5):28–35.
- [13] 王哲,何飞.高校网络舆情危机演化及管理机制研究[J].西南民族大学学报(人文社会科学版),2022,43(8):157–162.
- [14] 龚艳.面向公共卫生安全网络舆情预警的弱关联挖掘方法研究[J].情报科学,2022,40(6):19–24.
- [15] 李磊,支梅,李睿涵.新冠肺炎突发公共卫生事件网络舆情传播动力学[J].西安科技大学学报,2022,42(1):47–54.
- [16] 朱贺.社会突发事件网络舆情的传播管理研究[J].新闻世界,2024(4):62–64.
- [17] 管雨翔,王娟,兰月新,等.基于LDA和TextCNN的跨平台网络舆情风险预警研究[J].情报探索,2024(10):109–115.
- [18] 张梅,丁书恒,刘国芳,等.网络突发事件中的负性偏向:产生与表现[J].心理学报,2021,53(12):1361–1375.
- [19] 谢乐,杨波,杨美芳,等.企业危机事件网络舆情负性偏向驱动机制研究——基于40个案例的两阶段fsQCA分析[J].管理评论,2023,35(7):339–352.
- [20] 陈涛,林杰.基于搜索引擎关注度的网络舆情时空演化比较分析——以谷歌趋势和百度指数比较为例[J].情报杂志,2013,32(3):7–10,16.