

# 基于情感主题建模的负面舆情早期预警研究

崔 骥<sup>1</sup>, 韩益亮<sup>1,2</sup>, 朱率<sup>1</sup>, 李 鱼<sup>1</sup>, 吴旭光<sup>1</sup>

(1. 武警工程大学 密码工程学院, 西安 710086; 2. 武警工程大学 反恐指挥信息工程教育部重点实验室, 西安 710086)

**摘要：**社交网络中的负面舆情事件具有不可低估的影响力，针对基于情感分析的方法不能直接对负面网络舆情进行早期预警的问题，该文提出了一种基于情感分类和主题提取的舆情主题建模方法，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；针对负面舆情预警方法即时性不足的问题，构建网络舆情早期预警模型，从爆发指数、情绪指数、传播指数3个指标综合评估舆情主题的发展态势，设定舆情主题算数指数触发预警值，实现主题词对应的负面舆情事件的早期预警。实验结果表明，在 COVID-19 相关微博情感数据集 TF-IDF 权重排名前10的消极情绪主题词中，最早预警时间比舆情爆发日平均提前161.01h，实现的早期预警平均为2.1次；最早预警时间比舆情峰值日平均提前261.81h，平均早期预警5.8次。所提出的预警模型对社交网络舆情事件具有良好的早期预警效果。

**关键词：**网络舆情；情感分类；主题提取；舆情指数；早期预警

**中图分类号：**TP391 **文献标志码：**A

**DOI：**10.16511/j.cnki.qhdxxb.2024.27.005

网络舆情监测是现代社会中的一项重要工作<sup>[1]</sup>，它为分析和预测互联网上突发事件趋势提供适当的信息，有利于防止社会冲突<sup>[2]</sup>。以2019新型冠状病毒肺炎(corona virus disease 2019, COVID-19)大流行引发的重大公共卫生安全事件为例，社交网络中的负面舆论为科学防疫带来了诸多不良影响，关于负面舆情事件发展情况的信息被聚焦和夸大，伴随而来的便是 COVID-19 舆情热点的突发、频发<sup>[3]</sup>。近年来由于信息传播的规模和速度迅速加快，网络舆情工作的重点已从监测转向预警<sup>[4]</sup>。相关部门如果能尽早地对公众舆论进行建模分析和预警，发布官方新闻反驳谣言等，将有助于公众更好地了解网络舆情真实的发展趋势，避免不必要的恐慌，确保社会秩序的正常和稳定<sup>[5]</sup>。实时预测网络热点事件也有助于避免不良公众舆论引起二次舆情事件的爆发。

网络舆情预警研究的基础是自然语言处理，随着计算机算力提升和机器学习的发展，基于文本分类的方法在舆情预警研究中得到了广泛运用，该方法通过对大量舆情数据进行分类，然后对不同类

别的文本进行分析，实现对负面舆情言论的识别和检测，主要包括假新闻检测<sup>[6]</sup>和情感分析<sup>[7]</sup>等。假新闻检测法中，虚假信息大多为一条独立的文本，可以通过文本分类任务剥离出不实的新闻，准确率越高的分类模型越能更早地分辨出不可靠的信息，达到早期预警的目的<sup>[8]</sup>。然而，假新闻检测法不能直接应用在情感分析的舆情预警任务上，因为舆情事件中的文本信息是无数条信息的集合，不能简单地通过一条消极情绪文本实现预警的效果。情感分析法往往需要先构建文本分类器<sup>[9]</sup>，对大量舆情数据进行情绪分类，然后对不同情绪的文本进行分析，最后通过构建舆情评价指标分析和预警网络舆论热点事件<sup>[10]</sup>。

为实现基于情感分析方法的舆情分析，现有学者采用情感分类和主题挖掘协同分析的方法<sup>[11]</sup>，如 Li 等<sup>[12]</sup>通过对情感主题在舆情生命周期不同阶段的评论数和关注数统计，分析热点话题的舆情倾向，达到舆情防控的目的；张瑜等<sup>[13]</sup>基于离散 Bass 模型对网络舆情演化情况进行分析，设定词频-逆文档频率(TF-IDF)权重阈值，通过对敏感性舆论

收稿日期：2023-09-11

基金项目：国家社会科学基金西部项目(20XTQ007)；国家自然科学基金面上项目(61572521)

作者简介：崔骥(1992—)，男，硕士研究生。

通信作者：韩益亮，教授，E-mail: hanyil@163.com

话题不同主题的情绪趋势进行监测,实现舆情预警。但这种方法通常使用大量事后信息建模拟合舆情发展趋势,分析预测舆情发展动向,所以即时性不足。

为解决情感分析方法不能直接进行早期舆情预警的问题,本文提出了一种基于情感分类和主题提取的舆情主题建模方法,以研究负面舆情主题代替研究负面舆情事件进行早期预警;为解决早期预警即时性不足的问题,构建一种新的早期预警评价指标——舆情主题算数指数(public opinion topic arithmetic index, POI),并设定舆情指标触发阈值,达到早期预警的目的。

## 1 相关工作

研究网络舆情的方法包括文本分类<sup>[14]</sup>、文本挖掘<sup>[15]</sup>和复合模型<sup>[11]</sup>等。在基于文本分类的相关研究中,情感分析法通常需要先构建文本分类器,如机器学习模型、深度学习模型、预训练深度学习模型等。机器学习模型<sup>[16]</sup>如朴素 bayes (naive bayes, NB)<sup>[17]</sup>、K 近邻算法(K-nearest neighbor, KNN)<sup>[18]</sup>、支持向量机(support vector machine, SVM)<sup>[19]</sup>等,可以自动学习和识别数据中的模式和特征,避免了传统手工方法需要手动定义和调整的步骤,提高了效率和准确性,如王晰巍等<sup>[20]</sup>基于 NB 模型从词频、地域和时间 3 个可视化维度,对移动环境下网络舆情用户情感演化进行研究。但机器学习模型性能相对较低,而且很大程度上取决于数据质量和完整性,如果数据存在缺失、噪声或不准确的情况,可能会对模型的准确性和稳定性产生负面影响。大多数深度学习模型如循环神经网络(recurrent neural network, RNN)等,比机器学习模型的性能更优,如 Hu 等<sup>[14]</sup>采用长短期记忆网络(long short-term memory, LSTM)<sup>[21]</sup>构建了基于细粒度情感分析的汽车产品投诉危机智能预警系统;Shen 等<sup>[7]</sup>针对高校网络舆情分析和危机舆情预警的需要,提出了一种基于卷积神经网络(convolutional neural network, CNN)和 LSTM 模型,用于研究短文本舆情信息的情感分析方法;Choudrie 等<sup>[22]</sup>基于预训练大语言模型 RoBERTa<sup>[23]</sup>和迁移学习构建了一种多类情绪分类器系统,用于新冠大流行事件的网络舆情情绪分析研究。

公众舆论的趋势受到互联网上许多因素的影响,各因素之间的关系往往是复杂而紧密耦合的。

文本挖掘方法也可以很好地分析社交网络舆情特征<sup>[24]</sup>,为舆情监督有关部门治理网络生态环境提供参考和借鉴<sup>[25]</sup>。文本挖掘主要包括 2 种方式,一种是文本聚类,如 Iftene 等<sup>[26]</sup>基于语义相似性、文本聚类和事件检测提出了一种早期危机检测模型,通过检测相关事件表达的情绪在地图视图上显示有潜在风险的区域;Chen 等<sup>[1]</sup>提出通过基于熵的网络舆论相似性分析方法,对社交网络的热点和危机舆情事件进行聚类 and 识别。另一种是文本提取,如 Han 等<sup>[27]</sup>基于隐含 Dirichlet 分布模型和随机森林算法提出了一种主题提取和分类模型,通过研究 COVID-19 相关的新浪微博文本在 1 h 内的时间变化和空间分布,分析了 COVID-19 在中国流行早期的公众舆论。Xu 等<sup>[28]</sup>提出一种基于跨媒体分析的突发事件检测与观点挖掘方法,将检测到的突发事件利用观点挖掘分析舆情,然后通过语义扩展技术生成时间序列进行预警。文本挖掘能够有效地对大量的文本数据进行挖掘和分析,更好地理解网络舆情的主题和趋势,揭示数据中的隐藏信息,为预警研究提供有价值的信息。但文本挖掘的效果和准确性受数据预处理和特征提取的影响较大。

在复杂的互联网环境中,有许多因素可以影响公众舆论,单一方法往往不足以对复杂的舆情系统建模和分析,因此许多学者将文本分类、文本挖掘与其他理论模型结合进行舆情预警。Shah 等<sup>[11]</sup>通过对 COVID-19 患者的在线评论进行情感分类,并使用改进的隐含狄利克雷分布模型进行主题建模,分析了 COVID-19 爆发早期新兴和衰退的话题的情感趋势;Sarirete 等<sup>[29]</sup>基于情感分析、主题提取对 COVID-19 期间社交网络对疫苗的情绪态度进行分析;Vicario 等<sup>[30]</sup>充分利用在线社交媒体上传播错误信息中用户的偏见和两极化等特征,基于主题提取和情感分类构建了一个潜在错误信息目标的早期预警框架。

## 2 网络舆情早期预警模型

本文基于情感分类和主题提取的复合方法构建网络舆情早期预警模型,如图 1 所示,首先对 COVID-19 相关新浪微博文章(以下简称博文)进行数据采集,接着进行预处理,之后使用 ERNIE 3.0 情感分类器<sup>[31]</sup>对博文进行情感分类,计算不同情绪的博文数量。然后使用 TF-IDF 算法对含有消极情绪的博文进行主题提取,选择出与负面舆情事件相关联的主题词。接着利用综合爆发指

数 (explosion index, EI)、情绪指数 (sentiment index, SI) 和传播指数 (dissemination index, DI), 对舆情主题建模, 输出消极情绪主题词和其对应的 POI 值, 若模型输出值超出 POI 阈值, 则触发预警; 若未超过, 则继续进行数据采集。

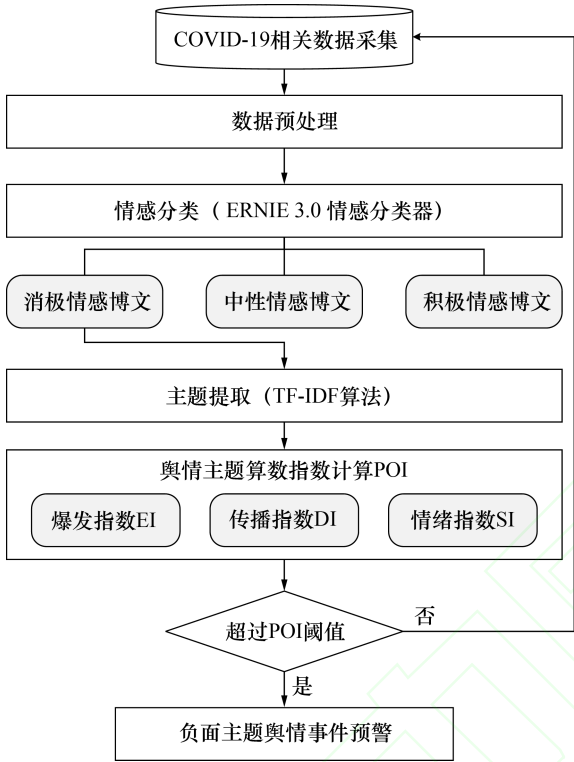


图 1 网络舆情早期预警模型

2.1 ERNIE 3.0 情感分类器

情感分类模块使用 ERNIE 3.0 大语言模型, 如图 2 所示, ERNIE 3.0 是一个持续多范式统一的预训练框架, 融合了自回归和自编码网络, 使得模型可以广泛用于预训练、微调和零/小样本学习。它使用一个共享的 Transformer 网络来完成语义表示任务, 并利用特定的自我注意掩码来控制预测条件的上下文。ERNIE 3.0 使不同的任务范式能够共享在共享网络中学习到的基本抽象特征, 并分别利用在各自的任务特定网络中学习到的任务特定顶层具体特征。此外, ERNIE 3.0 引入 ERNIE 2.0 中的持续多任务学习框架用于学习通用语义表示。

对于情感分类任务的应用, 首先初始化具有预先训练好共享网络参数的 ERNIE 3.0, 然后使用已标注的情感数据集继续进行训练以微调模型参数, 得到 3 种(消极、中性和积极)分类的情感分类器。超参数的设置包括: 学习率为  $4 \times 10^{-5}$ , 批处理大小为 64, 最大文本长度为 185 个字符, epoch 为 10

个轮次。

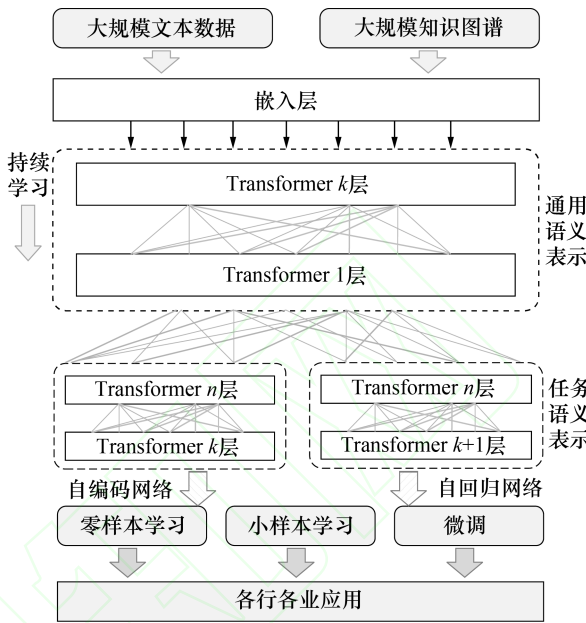


图 2 ERNIE 3.0 情感分类器结构示意图

2.2 TF-IDF 算法

主题提取模块使用 TF-IDF 算法, 词频(TF)表示字词在文本中出现的频率, 逆文档频率(IDF)表示关键词的普遍程度。该算法基于统计的关键词进行主题提取, 反映了字词对一个文件集或语料库中文档的重要性。本文将每个名词视为一个潜在的主题, 每篇博文视为一个文档, 利用 TF-IDF 算法计算频率来捕捉频繁出现的字词, 同时避免在每个文档中通过逆文档频率而出现的不重要的字词, 字词的重要性随着它在文件中出现次数的增加而提升, 在语料库中出现频率的增加而降低。一些通用字词对于表达文章的主题并没有太大作用, 反倒是一些 TF 较低的字词更有效, 所以仅使用 TF 是不合适的。如果包含该字词的文档越少而 IDF 越大, 说明该字词具有很好的类别区分能力。TF-IDF 的权重为 TF 和 IDF 的乘积, 计算方法如式(1)所示。

$$\begin{cases} TF_{t,d} = \log_a(1 + tf_{t,d}), \\ IDF_t = \log_a\left(\frac{M}{1 + df_t}\right), \\ (TF-IDF)_{t,d} = TF_{t,d} \times IDF_t. \end{cases} \quad (1)$$

其中:  $tf_{t,d}$  为主题词  $t$  在文档  $d$  中出现的频率,  $df_t$  为包含主题词  $t$  的文档数量,  $M$  为文档总数。

2.3 POI

为了深入了解消极情绪主题信息, 本文结合 EI、SI 和 DI 3 个指标综合计算 POI, 以反映消极情



绪主题的影响程度。

### 2.3.1 EI 计算方法

EI 是一种常用于电商购物节销售利润对比的统计方法,用以评估本期的某个指标在前一个时间段内的增速程度。在网络舆情事件中,EI 反映了当前消极情绪主题文本数量比前一段时间内平均文本数量的增加速度。热点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。主题词  $t$  的 EI 计算公式如下:

$$EI_t = \frac{\text{Relu}\left(df_{t,N}^n - \frac{1}{n-1} \sum_{i=1}^{n-1} df_{t,N}^i\right)}{1 + \frac{1}{n-1} \sum_{i=1}^{n-1} df_{t,N}^i}. \quad (2)$$

其中:  $df_{t,N}^n$  和  $df_{t,N}^i$  分别为第  $n$  天和第  $i$  天主题词  $t$  出现的消极(Negative)博文数量。Relu 为激活函数又叫线性整流函数,是人工智能神经网络中常用的方法,该函数在数值  $\leq 0$  时输出 0,  $> 0$  时输出原数值,具体过程如图 3 所示。

$EI_t \geq 1$  代表第  $n$  天主题词  $t$  的消极博文数量已经达到前  $n-1$  天均值的 2 倍及以上,增长程度 100%。当  $df_{t,N}^n$  小于前  $n-1$  天平均数量时,消极情绪舆论呈现负增长,触发 Relu 激活函数,此时  $EI_t$  值为 0,不再为爆发指数提供预警信息。EI 主要反映舆情主题的突然爆发和激增情况。当舆情主题突然爆发并迅速成为公众关注的焦点时,EI 会相应地增加。

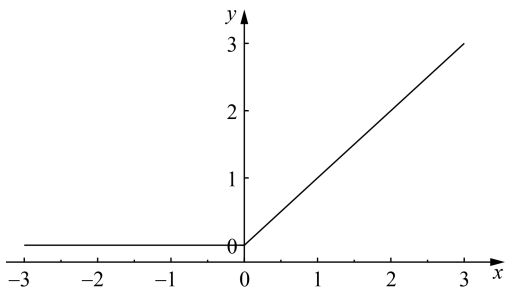


图 3 Relu 函数示意图

### 2.3.2 SI 计算方法

社交网络中善于煽动公众情绪的群体越来越多,大多数社交网络用户在讨论中很难始终维持中立态度,容易受舆论事件和社会情绪的影响而变得情绪化。消极情绪比例在本文是指网络舆情事件中消极博文数量与积极文章数量的比值,SI 可利用消

极情绪比例反映当前社交网络用户对舆情事件的消极情绪比前一时段的增长程度。SI 越高,表明该主题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,舆情事件越严重。主题词  $t$  的 SI 计算公式如下:

$$SI_t = \text{Relu} \left[ \frac{df_{t,N}^n}{1 + df_{t,P}^n} - \frac{\sum_{i=1}^{n-1} df_{t,N}^i}{1 + \sum_{i=1}^{n-1} df_{t,P}^i} \right]. \quad (3)$$

其中  $df_{t,P}^i$  为第  $i$  天主题词  $t$  出现的积极(Positive)博文数量。

$SI_t \geq 1$  代表了第  $n$  天主题词  $t$  的消极情绪比例已达到前  $n-1$  天均值的 2 倍及以上,增长程度大于 100%。当第  $n$  天的消极情绪比例小于前  $n-1$  天均值时,积极情绪占据主导地位,触发 Relu 激活函数,此时  $SI_t$  为 0,不再提供预警信息。

### 2.3.3 DI 计算方法

DI 反映了当前网络空间消极情绪博文传播数量比前一时段的增长程度。消极博文数量较多的舆情事件影响力大,覆盖范围广,危害性高;消极博文数量较少的舆情事件也有可能在某些特定人群传播,不代表影响性较小。因此结合 EI 和 SI 一起计算,能更有效衡量舆情事件的危害程度。主题词  $t$  的 DI 计算公式如下:

$$DI_t = \arctan \left( \log_a \left( 1 + \text{Relu} \left( df_{t,N}^n - \frac{1}{n-1} \sum_{i=1}^{n-1} df_{t,N}^i \right) \right) \right). \quad (4)$$

由于本文使用的数据集为百万级别,所以计算 DI 时的  $\log$  函数的底数  $a$  取 10,这与实际社交网络微博平台上亿级别的信息传输存在一定差距,因此对更大级别数据进行计算时可对  $a$  的值进行适当更改。为防止消极情绪博文基数过大造成持续预警,本文采用  $\arctan$  函数进行归一化处理。当 EI 和 SI 均较低时,DI 不会进行预警;而当舆情主题在短时间内迅速传播并影响到更广泛的人群中时,DI 会相应地增加。

### 2.3.4 综合计算 POI

结合 EI、SI 和 DI 3 个指标综合计算 POI,用以反映消极情绪主题的影响程度。主题词  $t$  的 POI 计算公式如下:

$$POI_t = SI_t + EI_t + DI_t. \quad (5)$$

## 2.4 舆情情感主题建模

舆情情感主题建模主要包括情感分类、主题提

取和 POI 计算。本文以负面舆情事件的主题词为切入点,在情感分类和主题提取后,对舆情主题进行建模。具体来说,情感分类模块将大量的微博舆情信息  $S=\{s_1, s_2, \dots, s_m\}$  使用 ERNIE 3.0 情感分类器进行分类,得到包含  $j$  条消极文本的集合  $N=\{N_1, N_2, \dots, N_j\}$ 、包含  $l$  条中性文本的集合  $E=\{e_1, e_2, \dots, e_l\}$  和包含  $k$  条积极文本的集合  $P=\{p_1, p_2, \dots, p_k\}$ 。然后在主题提取模块使用 TF-IDF 算法对消极情绪文本  $N$  进行主题提取,得到第  $w$  条消极情绪文本  $N_w$  的主题词  $t_w$ ,最后对 POI 值进行计算,假设 POI 的预警值为 2,则输出  $\text{POI} \geq 2$  的主题词和相应的 POI 值。伪代码如图 4 所示。

```
Input: 微博文本数据  $S=\{s_1, s_2, \dots, s_m\}$ 
Output: 消极情绪主题词  $t=\{t_1, t_2, \dots, t_j\}$ 
Begin
for  $S=1$  to  $m$  do  $s_m$  输入 ERNIE 3.0,
 $s_m$  prediction  $=(-1, 0, 1)$ 
 $N_j \leftarrow s_m(-1)$ ;  $e_i \leftarrow s_m(0)$ ;  $p_k \leftarrow s_m(1)$ 
End for
for  $N=1$  to  $j$  do  $n_j$  输入 TF-IDF,
 $t_j \leftarrow N_j$  (TF-IDF),
End for
for  $N=1$  to  $j$  do  $t_j$  计算 EI, SI, DI
 $\text{POI}(t_w) = \text{EI}(t_w) + \text{SI}(t_w) + \text{DI}(t_w)$ 
if  $\text{POI}(t_w) \geq 2$ 
print ( $t_w$ ) and  $\text{POI}(t_w)$ 
End for
End
```

图 4 舆情情感主题建模伪代码

3 实验与分析

3.1 数据预处理和分析

本文采用的数据集为 2020 年召开的第二十六届中国信息检索学术会议 (26th China Conference on Information Retrieval, CCIR 2020) 中的评测大赛“疫情期间网民情绪识别”赛题<sup>[32]</sup>, 以下简称为 COVID-19 相关微博情感数据集。该会议的组织者依据与 COVID-19 相关的 230 个主题词进行数据采集, 抓取了 2020 年 1 月 1 日至 2 月 19 日期间共计 100 万条博文数据, 包括 90 万条未标注情绪类别的数据和 10 万条已标注情绪类别的数据, 分别可组成未标注数据集和已标注数据集。

首先对数据进行预处理, 主要包括数据初筛和数据清洗。数据初筛主要针对数据集中少量无法使

用的数据, 从已标注数据集中删除了文本内容空白数据共计 359 条, 删除标签空白和标签错误的数据共计 81 条, 处理后的已标注数据集拥有 99 560 条数据。将标签空白和错误的标注数据加入到未标注数据集中, 处理后的未标注数据集拥有 900 081 条数据。已标注数据集中积极情绪博文的数量  $k$  为 25 376 条, 中性情绪博文的数量  $l$  为 57 283 条和消极情绪博文的数量  $j$  为 16 901 条, 按照 6 : 2 : 2 划分训练集、验证集、测试集, 详细信息如表 1 所示。

表 1 数据初筛后的已标注数据集的划分

划分类别	$j$	$l$	$k$	总计
训练集	10 050	34 389	15 121	59 560
验证集	3 405	11 456	5 139	20 000
测试集	3 446	11 438	5 116	20 000

数据清洗主要针对文本中的信息噪音。本文对已标注数据和未标注数据均进行了清洗, 主要执行以下操作: 1) 删除文本中的“@+用户名”、表情符、邮箱 (abc@demo.com), 2) 删除不必要的空格和换行符, 3) 取消转义 HTML 标记, 4) 用 URL 替换文本中提到的超链接, 5) 将繁体字转换为简体字。

预处理后, 使用 ERNIE 3.0 情感分类器对 900 081 条未标记数据进行情绪分类, 分类完成后将其与已标注数据集融合, 得到全部已标注数据 999 641 条, 需要说明的是, 由于 2020 年 2 月 19 日的博文不完整, 所以不作为后续实验分析的样本。全部已标注数据集中, 中性情绪博文的数量  $l$  为 596 427 条, 占比 59.66%; 积极情绪博文的数量  $k$  为 245 593 条, 占比 24.57%, 消极情绪博文的数量  $j$  为 157 621 条, 占比 15.77%。

COVID-19 相关的博文数量随时间的变化如图 5a 所示, 2020 年 1 月 19 日之前, 相关博文总数和 3 种情绪类别的博文数量占比均保持相对稳定; 1 月 19 日后, 3 种情绪类别的博文数量均呈现逐渐增加的趋势, 并在 1 月 20 日及以后呈现爆发性增长; 同时, 积极和消极情绪博文占比均逐渐增加, 中性情绪博文数量开始减少, 且绝大多数日期中积极情绪博文数量多于消极情绪。

消极与积极情绪博文数量、COVID-19 确诊人数随时间的变化分别如图 5b 与 5c 所示。对比可知, 1 月 20 日后, 消极情绪博文数量的占比最先开始增加, 积极情绪博文数量的占比减少, 而 1 月 21

日后积极情绪博文占比也开始增加。1月23日消极情绪博文数量上升至峰值,1月24日积极情绪博文数量才出现峰值,表明公众的消极情绪大多先于积极情绪爆发。不同情绪类别博文的数量和占比与中国政府相关媒体发布信息的行为密切相关,如2020年1月20日钟南山院士通过中国中央电视台证实了COVID-19存在“人传人”现象,消极情绪博文数

量逐渐增加并于1月23日达到首个峰值,在博文总数中的占比也最大。

因此,当某些主题的消极情绪博文数量出现爆发性增长,或占比远大于积极情绪博文时,该主题所对应的网络事件往往潜伏着严重的舆情危机。通过情绪主题分析舆情事件是一种有效的方法。

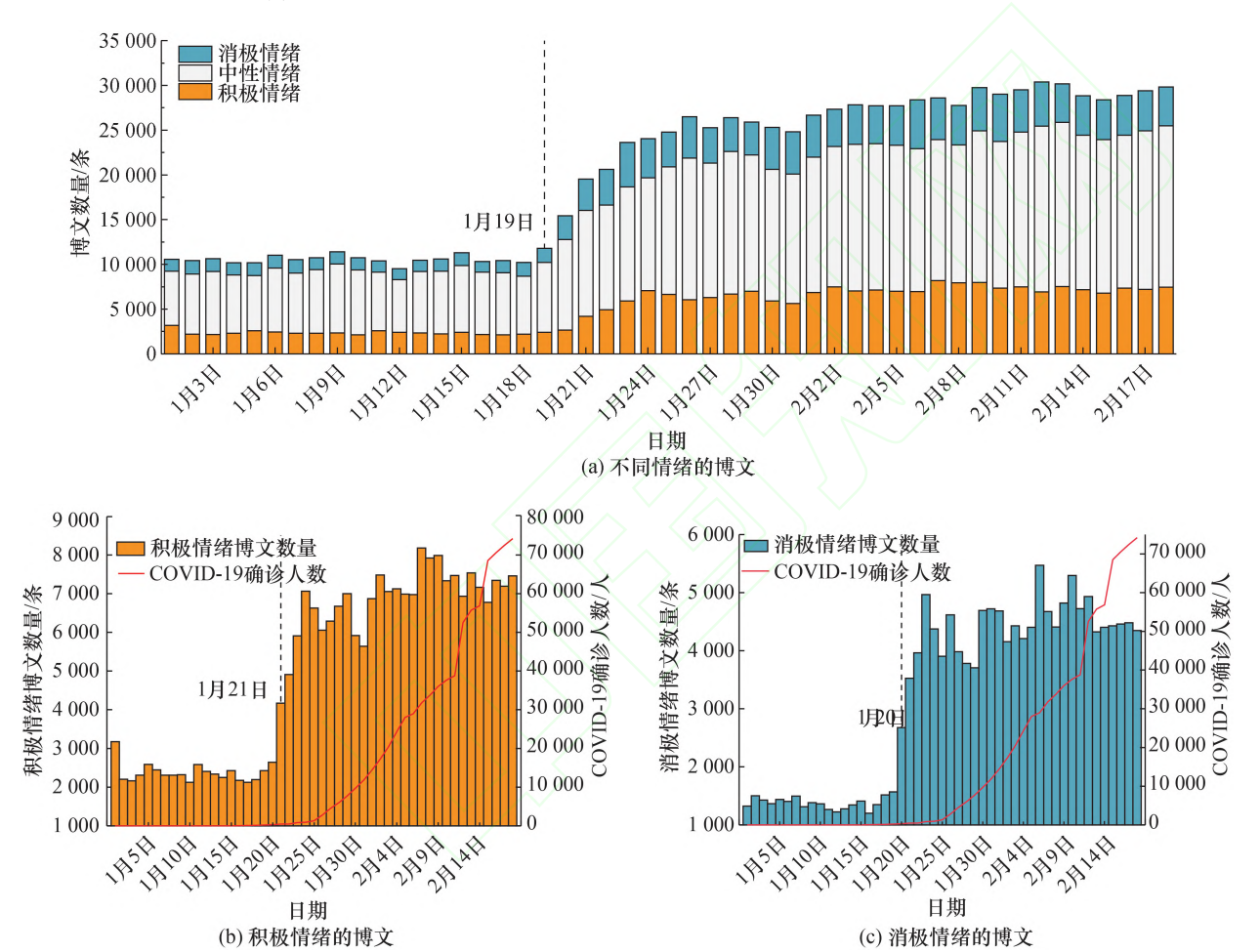


图5 COVID-19 相关博文数据随时间(2020年)的变化

3.2 舆情情感主题分析

使用 TF-IDF 算法针对预处理后的 COVID-19 相关微博情感数据集中 2020 年 1 月 1 日至 2 月 18 日的消极情绪博文进行主题提取,选择出 TF-IDF 权重排名前 10 的主题词,如表 2 所示。这些主题词大致与 4 类舆情事件相关联,分别是:公众自身疫情防护和保障,COVID-19 疫情公共卫生事件,新型冠状病毒起源,医院超负荷运转。不同舆情事件的主题词所关联的舆情事件可能存在交叉。

表 2 消极情绪主题词的 TF-IDF 权重和相关博文数

权重排名	主题词	权重/%	消极情绪博文数量/条
1	口罩	15.32	12 102
2	疫情	14.97	17 108
3	肺炎	10.82	11 528
4	病毒	9.35	15 466
5	野味	8.61	7 945
6	医院	7.42	12 786
7	医生	6.91	9 425
8	蝙蝠	6.14	6 704
9	病例	5.01	6 557
10	新冠	4.27	4 285



主题词“口罩”和“病毒”与公众自身疫情防护和保障的舆情相关。其中“口罩”的 TF-IDF 权重最高(15.32%),相关博文数量为 12 102 条;“病毒”的权重为 9.35%,排名第 4,相关博文数量为 15 466 条。这表明公众对病毒的防护措施非常关心,但由于当时的口罩供应量出现过短期不足,引发了一些不安和消极情绪。

主题词“疫情”、“肺炎”和“新冠”与 COVID-19 疫情公共卫生事件的舆情相关。“疫情”的权重(14.97%)虽然排名第 2,但相关博文数量最多(17 108 条);“肺炎”的权重(10.82%)排名第 3,相关博文数为 11 528 条。这体现了 COVID-19 疫情公共安全事件的发生会导致公众的消极情绪通过社交网络平台快速蔓延。“新冠”的权重(4.27%)和相关博文数量(4 285 条)最少,这可能是因为微博平台的部分用户倾向于使用“疫情”或“肺炎”之类的简略表达。

主题词“野味”和“蝙蝠”与新型冠状病毒起源的舆情相关。“野味”的权重(8.61%)排名第 5,相关博文数量为 7 945 条;“蝙蝠”的权重(6.14%)排名第 8,相关博文数量为 6 704 条。当时围绕病毒的起源产生了大量谣言,干扰了相关部门正常的舆论引导工作,引发了一些消极情绪。

主题词“医院”、“医生”和“病例”与医院超负荷运转的舆情相关。3 个主题词的权重相近,分别为 7.42%、6.91%和 5.04%。公众出于对自身生命健康和卫生安全的考虑,对事件基本情况、进展程度、政府措施、防控手段等信息有较大需求。但随着感染病例的增多出现了医院超负荷运转、医务人员短缺等现象,导致了消极情绪的产生和传染,可能引发更深层次的网络舆情危机。

### 3.3 舆情事件周期分析

根据舆情发展态势,有的学者将舆情周期划分为 3 或 4 个阶段<sup>[33]</sup>。结合社交网络舆情事件特点,本文将舆情周期划分为潜伏期、爆发期、持续期和衰退期。以主题词“医生”为例,结合 2.3 节提出的 EI,分析各阶段的发展情况。

若从 2020 年 1 月 1 日开始分析,则无法统计该日期前的相关博文总数(超出数据集的时间周期),同时为防止样本量过小造成误差,本节选择

1 月 8 日为计算起始日。以 EI 作为划分舆情周期的依据,当 EI 不超过或未连续超过预警值时为潜伏期,该时期的博文总数并没有持续爆发性变化,而是在一定范围波动;当 EI 开始长时间超过预警值并持续增长至峰值时为爆发期,舆情事件呈现爆发性增长,短时间内的讨论热度迅速达到较高水平;当 EI 首次低于预警值并在预警值附近波动时为持续期,该时期虽然博文总数增速减缓,但其基数仍然较大;当 EI 首次小于 0 并在 0 附近波动时,每日的博文总数进入负增长,舆情事件进入衰退期。

如图 6 所示,在主题词“医生”的相关舆情事件中,2020 年 1 月 19 日及以前为潜伏期,1 月 20 日至 1 月 25 日为爆发期,1 月 26 日至 2 月 2 日为持续期,2 月 3 日及以后为衰退期。潜伏期中,大部分日期的 EI 小于 2 并保持小范围波动,说明话题进入社交网络用户视野引发讨论,但讨论量较低;爆发期中,EI 开始大于 2 并且持续增长直至 1 月 20 日达到峰值(7.42),主题词“医生”的博文总数也在同一日达到峰值,说明话题热度很高,社交网络转发和评论的数量迅速增加。持续期中,EI 经爆发期后首次跌至 2 以下,并在 2 左右来回波动,博文总数也基本保持不变,说明随着政府相关部门的逐渐发声,社交网络博文的情绪类别逐渐趋于稳定。衰退期中,EI 首次跌破 1 并逐渐降至 0 甚至负值,说明该话题热度减退,参与转发、讨论信息的博文总数也逐渐减少,舆情事件可能逐渐消散。然而,在持续期和衰退期仍需持续关注舆情,防止二次舆情事件的发生。例如 2020 年 2 月 6 日李文亮医生在抗击疫情时因感染 COVID-19 去世,引发大规模的社交网络用户共情,相关话题一度达到微博热搜第一名,于是当天与主题词“医生”相关的博文总数大量增加,直到 2 月 14 日重新进入衰退期。

本文所提方法主要针对有征兆的网络舆情事件,在其潜伏期和爆发前期尽可能早地检测出负面舆情主题,提出预警防范,为相关部门的舆情监测工作和学术研究提供技术和理论支持。但该方法还无法对急性突发舆情事件做到提前预警,所以仍需完善。

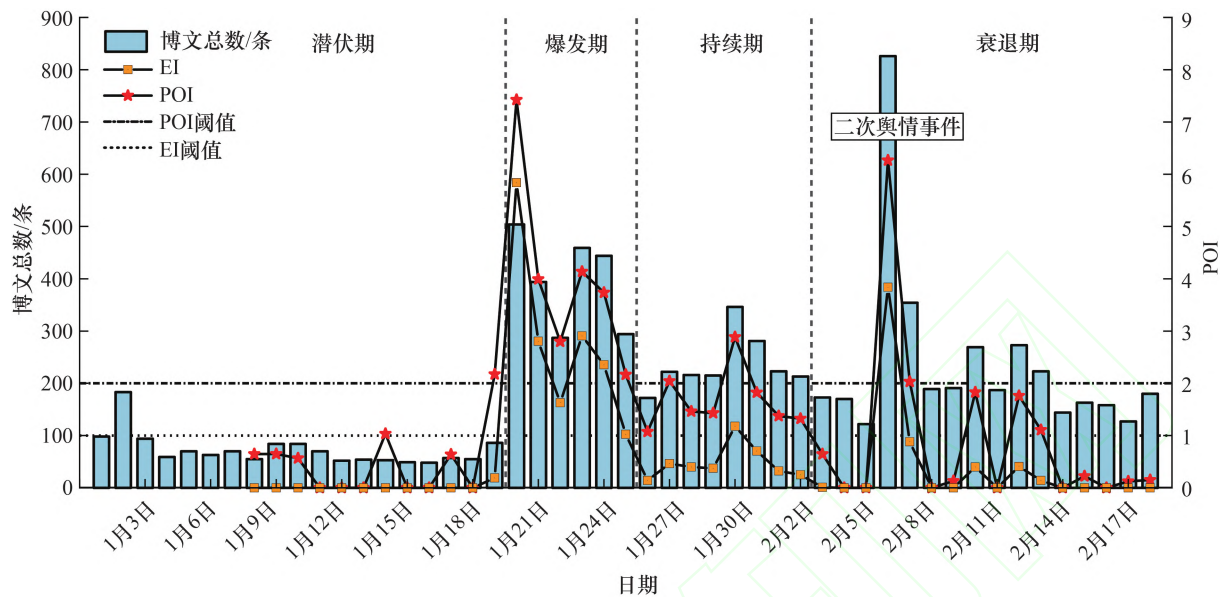


图 6 主题词“医生”在舆情周期内的相关博文总数与 EI、POI 的关系(2020 年)

3.4 预警评价指标分析

针对 POI 及其 3 个子指标(EI、SI、DI)的预警值进行实验分析和选择。为了方便分析每个指标数值的正负波动氛围,本节将式(2)—(5)中的 Relu 函数删去后再进行计算。EI、SI、DI 也可以单独用于分析网络舆情特定方向,预警值设定均不宜过高,要尽量囊括大多数主题词在大部分潜伏期内的数值;也不宜过低,让无隐患主题的小幅波动影响预警效果。TF-IDF 权重排名前 10 的主题词在潜伏期内的预警评价指标统计结果如表 3 所示。

EI 代表着舆论主题短时间内增长程度。主题词“野味”在潜伏期的 EI 最低,为-0.87;“疫情”在潜伏期的 EI 最高,为 1.95;大部分主题词的 EI 方差少于 0.26,而“疫情”的方差最高,达到了 0.70,说明该主题词在潜伏期内的 EI 波动较大,不适合作为预警主题词。在去掉过大的 EI 异常值(1.95),并考虑其他所有主题词的 EI 最大值为 0.94 的情况下,将舆情主题词的 EI 预警值设定为 1,当 EI>1 时表明舆情主题有突然爆发的趋势,属于异常状态,需要引起关注。

SI 代表着社交网络用户对舆情事件的消极情绪比前一时段的增长程度。大部分主题词在潜伏期的 SI 最大值超过 1,“野味”和“蝙蝠”的 SI 最高,均为 6.0;“野味”的 SI 最低,为-11.75;大部分主题

词的 SI 方差超过 1,“野味”甚至达到了 24.55,说明主题词在潜伏期的 SI 值波动均较大,因此本文对 SI 预警值的选择并不依赖潜伏期,而是爆发期及以后。以主题词“口罩”为例,由图 7a 可知,SI 在潜伏期和爆发期前期的波动较大,爆发期中后期及以后时期的波动较小;但 1 月 18 日以后,SI 大多处于 1 以下。因此将 SI 预警值设定为 1,当 SI>1 时表明社交网络用户的消极情绪有突然爆发的趋势,属于异常状态,需要引起关注。

DI 代表着当前负面舆情主题词数量比之前平均数量的增长程度,“病毒”的方差和标准差均最高,分别为 0.55 和 0.74;大部分主题词的方差位于 0.20~0.44,说明 DI 整体波动不大;“医院”的 DI 最高,为 0.99。为了囊括所有主题词,将 DI 预警值设定为 1。

综合上述 3 个子指标对 POI 进行分析。以主题词“口罩”为例,由图 7b 可知,POI 在爆发期和潜伏期后期波动较大,而且 POI 达到峰值的日期整体早于博文数量达到峰值的日期,体现了 POI 进行网络舆情事件早期预警的有效性。舆情前期 DI 较低, EI 与 SI 较高,舆情中期开始 DI 较高, EI 和 SI 数值呈现跳跃性,但总体数值较低。因此将 POI 预警值设置为 2,能综合各子指标在不同时期的特点,体现整个舆情周期内的状态。



表 3 TF-IDF 权重前 10 的消极情绪主题词在潜伏期内的预警评价指标统计

预警评价指标		主题词									
		口罩	疫情	肺炎	病毒	野味	医院	医生	蝙蝠	病例	新冠
EI	最大值	0.45	1.95	0.21	0.94	0.80	0.24	0.19	0.37	0.26	0.80
	最小值	−0.38	−0.74	−0.59	−0.44	−0.87	−0.50	−0.39	−0.35	−0.52	−0.18
	平均值	0.06	−0.10	−0.28	0.05	−0.34	−0.23	−0.22	0.03	−0.12	0.09
	方差	0.09	0.70	0.07	0.17	0.26	0.04	0.03	0.06	0.09	0.17
	标准差	0.30	0.84	0.26	0.41	0.51	0.20	0.18	0.24	0.29	0.41
SI	最大值	4.50	5.17	1.38	2.37	6.00	2.00	0.65	6.00	3.93	1.00
	最小值	−4.67	−6.00	−0.75	−1.51	−11.75	−1.22	−1.05	−4.47	−1.63	−1.00
	平均值	0.41	0.10	0.34	0.29	−0.25	0.11	0.11	−0.27	0.43	0.00
	方差	6.65	8.37	0.55	1.14	24.55	1.05	0.55	9.63	2.52	0.36
	标准差	2.58	2.89	0.74	1.07	4.95	1.03	0.74	3.10	1.59	0.60
DI	最大值	0.57	0.91	0.65	0.94	0.83	0.99	0.86	0.75	0.67	0.27
	最小值	−0.54	−0.77	−0.85	−0.82	−0.86	−1.07	−1.00	−0.75	−0.82	−0.09
	平均值	0.08	−0.27	−0.48	−0.08	−0.39	−0.80	−0.68	0.07	−0.13	0.02
	方差	0.20	0.36	0.34	0.55	0.38	0.33	0.32	0.36	0.44	0.02
	标准差	0.45	0.60	0.58	0.74	0.62	0.57	0.57	0.60	0.67	0.15
POI	最大值	4.72	8.03	1.96	4.26	6.41	2.72	2.17	7.12	4.56	2.07
	最小值	−3.71	−7.51	−1.93	−2.49	−13.39	−2.79	−2.42	−4.18	−2.91	−1.27
	平均值	0.55	−0.27	−0.42	0.26	−0.98	−0.91	−0.79	−0.17	0.18	0.10
	方差	6.77	16.64	1.96	3.85	29.87	2.26	1.71	10.63	4.96	1.15
	标准差	2.60	4.08	1.40	1.96	5.47	1.50	1.31	3.26	2.23	1.07

3.5 早期舆情预警分析

本文将舆情主题词进入爆发期后，相关消极情绪博文数量出现首个峰值的日期设定为峰值日，将EI首次大于1并持续增长至峰值的日期设定为爆发日。表4总结了TF-IDF权重前10的消极情绪主题词在爆发日和峰值日的各指标信息。

由表可知，主题词“病毒”在峰值日的消极情绪博文数量最多，为761条，“新冠”在峰值日的消极情绪博文数量最少，为137条；“医生”在爆发日的消极情绪博文数量最多，为504条，“新冠”在爆发日的消极情绪博文数量最少，为16条；“新冠”在爆发日前的大部分日期中的消极情绪博文数量为0，这是因为世界卫生组织在2020年1月11日将造成不明原因肺炎的病毒暂时命名为COVID-19，简称为新冠肺炎。“疫情”在爆发日的POI最高(88.27)，“蝙蝠”在峰值日的POI最高(59.57)；“蝙蝠”在爆发日的EI最高(8.91)，“新冠”在峰值日的EI最高(29.82)，且大部分主题词的EI在爆发日比峰值日低，这是因为爆发日为爆发期第一天，之后的上升速度可能加快。与EI相反，大部分主题词的SI在爆发日比峰值日高，这是因为SI在潜伏期和爆发期前期的波动较大，在爆发后期较为稳定，这也

与图5a中积极情绪博文数量在绝大部分日期多于消极情绪的结论相呼应。预警提前时间是指触发预警阈值的第一条博文发帖时间，比舆情峰值日当天零时提前了多少个h。

本文提出的爆发日整体早于博文数量峰值日，舆情主题词的爆发日比社交网络舆情爆发日期整体提前，舆情事件早期预警时间比爆发日期整体提前，说明本文提出的网络舆情早期预警模型能够及早地感知舆情事件，负面舆情事件也能够得到提前预警。在这10个消极情绪主题词中，“口罩”“疫情”“肺炎”“病毒”主题的爆发日最早，均为2020年1月17日；该日期之后，与新型冠状病毒起源舆情事件相关的“野味”“蝙蝠”主题词，和与医院超负荷运转舆情事件相关的“医院”“医生”“病例”主题词也相继爆发。网络舆情早期预警模型最早于1月8日08时19分开始针对相关主题词进行预警，并且大部分的预警时间远早于实际爆发日。主题词“新冠”的消极情绪博文在爆发日只有16条，但最早在1月10日13时28分就开启了预警，说明该模型的早期预警不一定需要依赖大量数据样本即可实现。主题词“疫情”的最早预警时间比峰值日提前的时间最多，为423.52h，早期预警次数也最多，为9次；主

题词“新冠”的最早预警时间比爆发日提前的时间最多,为 250.53 h,早期预警次数为 2 次。本文所提模型的最早预警时间比舆情爆发日平均提前了

161.01 h,平均早期预警 2.1 次;最早预警时间比舆情峰值日平均提前 261.81 h,平均早期预警 5.8 次,表明该模型具有良好的早期预警能力。

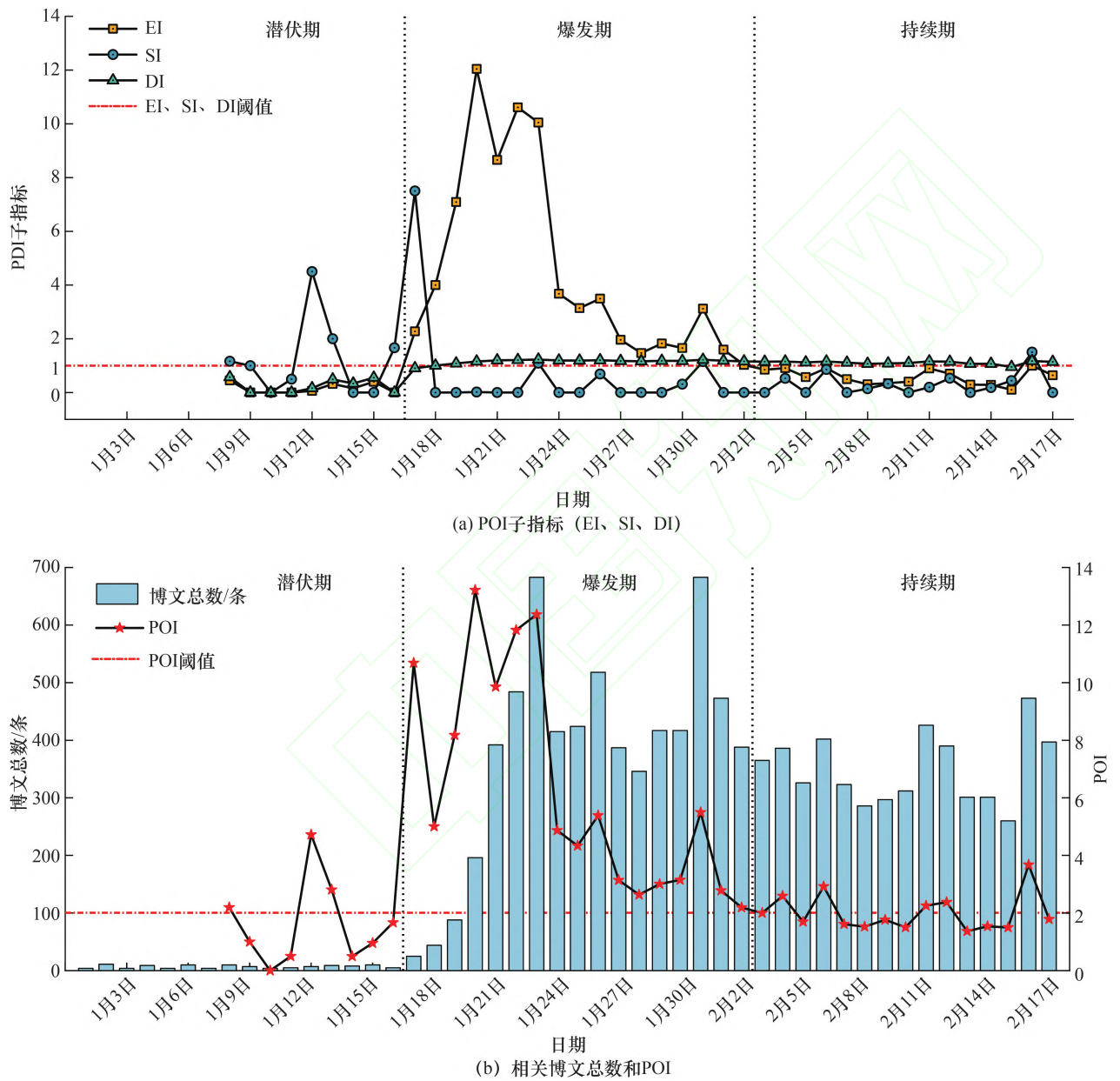


图 7 主题词“口罩”的舆情分析结果(2020 年)

3.6 模型局限性

为了说明网络舆情早期预警模型的局限性,进一步对 TF-IDF 权重排名前 100 的消极情绪主题词进行了早期预警。部分主题词没有达到早期预警效果或预警错误,分析其可能原因如下:

1) 模型对无预兆的、突发性的舆情事件的早期预警效果不佳。以主题词“医生”为例,1 月 19 日前“医生”相关消极主题博文数量保持相对稳定,1 月

20 日“朝阳医院眼科大夫陶勇医生被砍”事件登上微博热搜,由于媒体对于事件进展的深度报道和信息核对需要一定时间,该信息空白期内出现了网络负面舆情的爆发,而模型对此无法做到提前预警。

2) 模型中的 TF-IDF 算法会错将部分常用语的修饰性名词提取为主题词,如“时间”“视频”等。这些词语会出现在消极情绪博文中且占据了较大比例,但实际上并不适合作为主题词。

表 4 TF-IDF 权重前 10 的消极情绪主题词的早期预警结果

权重排名	主题	最早预警时间	日期	消极博文数量/条	POI	EI	SI	DI	预警提前时间/h	早期预警数/次
1	口罩	1 月 8 日	爆发日 1 月 17 日	25	10.68	2.28	7.50	0.91	193.20	3
		22:48	峰值日 1 月 23 日	683	12.37	10.05	1.09	1.23	337.20	9
2	疫情	1 月 8 日	爆发日 1 月 17 日	81	88.27	8.44	78.75	1.08	207.52	1
		08:19	峰值日 1 月 26 日	734	7.58	6.36	0.00	1.22	423.52	9
3	肺炎	1 月 11 日	爆发日 1 月 17 日	41	2.57	1.07	0.56	0.94	120.72	1
		23:17	峰值日 1 月 23 日	707	7.68	6.45	0.00	1.23	264.72	6
4	病毒	1 月 11 日	爆发日 1 月 17 日	75	4.26	0.94	2.37	0.94	129.36	2
		14:22	峰值日 1 月 21 日	761	9.17	7.93	0.00	1.24	225.63	5
5	野味	1 月 9 日	爆发日 1 月 19 日	23	9.37	1.11	7.42	0.85	230.45	4
		9:33	峰值日 1 月 23 日	502	10.60	9.39	0.00	1.21	326.45	8
6	医院	1 月 9 日	爆发日 1 月 20 日	392	4.96	2.41	1.36	1.18	246.22	2
		17:47	峰值日 1 月 24 日	717	4.49	3.27	0.19	1.22	342.22	6
7	医生	1 月 19 日	爆发日 1 月 20 日	504	7.42	5.84	0.37	1.21	1.15	1
		22:51	峰值日 1 月 20 日	504	7.42	5.84	0.38	1.21	1.15	1
8	蝙蝠	1 月 12 日	爆发日 1 月 21 日	218	29.67	8.91	19.60	1.16	214.47	4
		17:32	峰值日 1 月 22 日	287	59.57	8.44	49.95	1.18	238.47	5
9	病例	1 月 17 日	爆发日 1 月 18 日	66	3.28	2.25	0.00	1.03	16.20	1
		7:48	峰值日 1 月 23 日	524	9.24	8.03	0.00	1.21	136.20	6
10	新冠	1 月 10 日	爆发日 1 月 21 日	16	16.51	3.13	6.00	0.88	250.53	2
		13:28	峰值日 1 月 24 日	137	30.95	29.82	0.00	1.13	322.53	3

4 结 论

本文基于情感分类和主题提取的方法构建了网络舆情早期预警模型，采用研究消极情绪主题词的方法对负面舆情主题进行量化分析，提出了舆情主题算数指数(POI)，并结合 POI 的 3 个子指标，即爆发指数、情绪指数和传播指数，分别剖析了当前舆情主题消极情绪博文数量比前一时段的增加程度，当前社交网络用户对舆情事件消极情绪比前一时段的增长程度，当前网络空间消极情绪博文传播数量比前一时段的增长程度。通过设定各指标预警值，进行早期舆情预警。利用 COVID-19 疫情爆发初期的微博情感识别数据集验证了该模型具备较优的网络负面舆情事件早期预警能力。

参考文献 (References)

[1] 李静怡. 中国高校网络舆情的传播与应对策略研究 [D]. 济南: 山东大学, 2021.  
LI J Y. Research on internet public opinion communication and countermeasures of Chinese universities [D]. Ji'nan: Shandong University, 2021. (in Chinese)

[2] LIAN Y, DONG X F, LIU Y J. Topological evolution of the internet public opinion [J]. Physica A: Statistical Mechanics and its Applications, 2017, 486: 567 – 578.  
[3] 李雪. 双黄连抢购事件下的舆情传播网络特征及演化博弈研究 [D]. 西安: 西安电子科技大学, 2022.  
LI X. Research on network characteristics and evolutionary game of public opinion dissemination under Shuanghuanglian panic buying event [D]. Xi'an: Xidian University, 2022. (in Chinese)  
[4] KUMAR P, HAMA S, OMI DVARBORNA H, et al. Temporary reduction in fine particulate matter due to ‘anthropogenic emissions switch-off’ during COVID-19 lockdown in Indian cities [J]. Sustainable Cities and Society, 2020, 62: 102382.  
[5] YAN L, ZHANG H T, GONCALVES J, et al. An interpretable mortality prediction model for COVID-19 patients [J]. Nature Machine Intelligence, 2020, 2(5): 283 – 288.  
[6] GALLOTTI R, VALLE F, CASTALDO N, et al. Assessing the risks of ‘infodemics’ in response to COVID-19 epidemics [J]. Nature Human Behaviour, 2020, 4(12): 1285 – 1293.



- [7] SHEN L, XU M H. Student public opinion management in campus commentary based on deep learning [J]. *Wireless Communications and Mobile Computing*, 2022, 2022: 2130391.
- [8] 武兴隆. 基于主体关系情感的社交网络事件热度的预测 [D]. 哈尔滨: 哈尔滨工业大学, 2018.
- WU X L. Prediction for popularity of events in social network based on subject relationship emotion [D]. Harbin: Harbin Institute of Technology, 2018. (in Chinese)
- [9] CUI S, HAN Y, DUAN Y, et al. A Two-Stage Voting-Boosting Technique for Ensemble Learning in Social Network Sentiment Classification[J]. *Entropy*, 2023, 25(4): 555.
- [10] RAO D N, MIAO X, JIANG Z H, et al. STANKER: Stacking network based on level-grained attention-masked BERT for rumor detection on social media [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. EMNLP, 2021: 3347–3363.
- [11] SHAH A M, YAN X B, QAYYUM A, et al. Mining topic and sentiment dynamics in physician rating websites during the early wave of the COVID-19 pandemic: Machine learning approach [J]. *International Journal of Medical Informatics*, 2021, 149: 104434.
- [12] LI J H, PAN F Y, ZHOU B, et al. Research on the early warning of college online public opinion under the big data environment [C]//Proceedings of the IEEE 3rd International Conference on Computer and Communication Engineering Technology. Beijing, China: IEEE, 2020: 14–18.
- [13] 张瑜, 李兵, 刘晨玥. 面向主题的微博热门话题舆情监测研究——以“北京单双号限行常态化”舆情分析为例 [J]. *中文信息学报*, 2015, 29(5): 143–151, 159.
- ZHANG Y, LI B, LIU C Y. Topic-oriented monitoring of public sentiment towards popular Weibo events—A case study on “Regular ‘Odd-even’ vehicle restriction in Beijing” [J]. *Journal of Chinese Information Processing*, 2015, 29(5): 143–151, 159. (in Chinese)
- [14] HU H J, WEI Y H, ZHOU Y. Product-harm crisis intelligent warning system design based on fine-grained sentiment analysis of automobile complaints [J]. *Complex & Intelligent Systems*, 2023, 9(3): 2313–2320.
- [15] 程晏. 基于 LDA 模型的地铁投诉文本挖掘及满意度评价研究 [D]. 北京: 北京交通大学, 2020.
- CHENG Y, Research on subway complaint text mining and satisfaction evaluation based on LDA model [D]. Beijing: Beijing Jiaotong University, 2020. (in Chinese)
- [16] ZHU R B, DING Q A, YU M, et al. Early warning scheme of COVID-19 related internet public opinion based on RVM-L model [J]. *Sustainable Cities and Society*, 2021, 74: 103141.
- [17] BRAVO-MARQUEZ F, MENDOZA M, POBLETE B. Combining strengths, emotions and polarities for boosting Twitter sentiment analysis [C]//Proceedings of the Second International Workshop On Issues of Sentiment Discovery and Opinion Mining. Chicago, USA: ACM, 2013: 2.
- [18] DAVIDOV D, TSUR O, RAPPOPORT A. Enhanced sentiment learning using Twitter hashtags and smileys [C]//Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Beijing, China: ACL, 2010: 241–249.
- [19] MOHAMMAD S, KIRITCHENKO S, ZHU X D. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets [C]//Proceedings of the Seventh International Workshop on Semantic Evaluation. Atlanta: ACL, 2013.
- [20] 王晰巍, 张柳, 文晴, 等. 基于贝叶斯模型的移动环境下网络舆情用户情感演化研究——以新浪微博“里约奥运会中国女排夺冠”话题为例 [J]. *情报学报*, 2018, 37(12): 1241–1248.
- WANG X W, ZHANG L, WEN Q, et al. Research on sentiment evaluation of online public opinion based on the Bayesian model in a mobile environment: The case of “China Women’s Volleyball Won the Championship in the Rio Olympics” in SinaWeibo [J]. *Journal of the China Society for Scientific and Technical Information*, 2018, 37(12): 1241–1248. (in Chinese)
- [21] 李卫疆, 漆芳, 余正涛. 基于多通道特征和自注意力的情感分类方法 [J]. *软件学报*, 2021, 32(9): 2783–2800.
- LI W J, QI F, YU Z T. Sentiment classification method based on multi-channel features and self-attention [J]. *Journal of Software*, 2021, 32(9): 2783–2800. (in Chinese)
- [22] CHOUDRIE J, PATIL S, KOTTECHA K, et al. Applying and understanding an advanced, novel deep learning approach: A Covid 19, text based, emotions analysis study [J]. *Information Systems Frontiers*, 2021, 23(6): 1431–1465.
- [23] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach [J]. *arXiv preprint arXiv: 1907.11692*, 2019.

[24] LAU J H, COLLIER N, BALDWIN T. On-line trend analysis with topic models: #Twitter trends detection topic model online [C]//Proceedings of COLING 2012. Mumbai, India, 2012: 1519-1534.

[25] 石磊, 杜军平, 梁美玉. 基于RNN和主题模型的社交网络突发话题发现 [J]. 通信学报, 2018, 39(4): 189-198.

SHI L, DU J P, LIANG M Y. Social network bursty topic discovery based on RNN and topic model [J]. Journal on Communications, 2018, 39(4): 189-198. (in Chinese)

[26] IFTENE A, GINSCA A L. Using opinion mining techniques for early crisis detection [J]. International Journal of Computers Communications & Control, 2014, 7(5): 857-864.

[27] HAN X H, WANG J L, ZHANG M, et al. Using social media to mine and analyze public opinion related to COVID-19 in China [J]. International Journal of Environmental Research and Public Health, 2020, 17(8): 2788.

[28] XU W, LIU L Y, SHANG W. Leveraging cross-media analytics to detect events and mine opinions for emergency management [J]. Online Information Review, 2017, 41(4): 487-506.

[29] SARIRETE A. Sentiment analysis tracking of COVID-19 vaccine through tweets [J]. Journal of Ambient Intelligence and Humanized Computing, 2022: 1-9.

[30] DEL VICARIO M, QUATTROCIOCCI W, SCALA A, et al. Polarization and fake news: Early warning of potential misinformation targets [J]. ACM Transactions on the Web, 2019, 13(2): 10.

[31] SUN Y, WANG S H, FENG S K, et al. ERNIE 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation [J]. arXiv preprint arXiv: 2107.02137, 2021.

[32] World Health Organization. Coronavirus disease (COVID-19)—events as they happen [EB/OL]. [2020-03-22]. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>.

[33] 贾亚敏, 安璐, 李纲. 城市突发事件网络信息传播时序变化规律研究 [J]. 情报杂志, 2015, 34(4): 91-96, 90.

JIA Y M, AN L, LI G. On the online information dissemination pattern of city emergencies [J]. Journal of Intelligence, 2015, 34(4): 91-96, 90. (in Chinese)

Research on early warning of negative public opinion based on sentiment topic modeling

CUI Su<sup>1</sup>, HAN Yiliang<sup>1, 2</sup>, ZHU Shuaishuai<sup>1</sup>, LI Yu<sup>1</sup>, WU Xuguang<sup>1</sup>

(1. College of Cryptographic Engineering, Chinese People's Armed Police Force Engineering University, Xi'an 710086, China;

2. College of Cryptographic Engineering, Ministry of Education Key Laboratory of Counter-Terrorism Command Information Engineering, Xi'an 710086, China)

**Abstract:** [Objective] The effect of negative public opinion events on social networks is underestimated. To address the issue of sentiment-based methods not being able to directly achieve early warning of negative online public opinion, this study proposes a sentiment classification and topic extraction-based approach to public opinion topic modeling. Using negative emotional topics as an entry point, this study shifts from investigating negative public opinion events to examining negative public opinion topics, thus facilitating statistical and quantifiable analysis of such events. Additionally, to address the persistent shortcomings of methods for negative public opinion early warning, we construct a novel early warning evaluation metric, which is known as the public opinion topic arithmetic index (POI). This index comprehensively assesses the developmental trends of public opinion topics across three dimensions: explosion index (EI), sentiment index (SI), and dissemination index (DI). [Methods] This study employs the ERNIE 3.0 large-scale language model for sentiment classification. The annotated sentiment dataset is further trained and fine-tuned to obtain the required sentiment classifier. It performs sentiment classification on a COVID-19 Weibo emotional dataset, computing various post sentiments. The topic extraction module uses the TF-IDF algorithm to extract topics. Each noun tag is considered a potential topic, whereas each

Weibo post is treated as a document. The TF-IDF method captures frequently occurring words by calculating their frequencies and avoiding less important terms that appear in each document. The TF-IDF topic extraction algorithm extracts topics from negative emotional Weibo posts and identifies relevant topics associated with negative public opinion events. Finally, POI is employed for further analysis based on the extracted public opinion topics. Consequently, early warning is achieved by analyzing negative public opinion topics instead of events. Furthermore, POI comprehensively calculates the effect of negative public opinion topics by combining EI, SI, and DI. EI reflects the growth rate of the current number of textual instances related to negative emotional topics compared to the average number in a previous period; SI mainly reflects the public's emotions and sentiments triggered by public opinion topics; and DI mainly represents the scope and speed of dissemination of public opinion topics. Finally, a comprehensive negative emotional topic public opinion index is derived by calculating the EI, SI, and DI of emotional topics and postdata information, and the topics that exceeded the warning threshold are warned.

[Results] The experimental results reveal that the proposed early warning model effectively predicts social media public opinion events. Among the top ten negatively perceived topics ranked based on weight, the earliest warning time exceeds the average outbreak day by 161.01 hours, with an average of 2.1 early warnings. Additionally, the earliest warning time exceeds the average peak day by 261.81 hours, with an average of 5.8 early warnings.

[Conclusions] We establish a threshold for triggering the arithmetic index of public opinion topics by modeling and calculating the arithmetic index of negative public opinion topics in this study. This enables us to exclude negative topics and corresponding public opinion events that surpass the threshold, thereby achieving early warning for topic-related negative public opinion events. The proposed negative public opinion warning model accomplishes its intended objective by employing sentiment analysis methods for the early detection of online public opinions.

**Key words:** online public opinion; sentiment classification; topic extraction; public opinion index; early warning form

(责任编辑 刘婉琪)