

基于主题模型与图神经网络的突发公共卫生事件国际舆情演化分析研究

高境健¹, 桑国明¹, 刘智¹, 张益嘉^{*,1}, 林鸿飞²

1.大连海事大学, 信息科学技术学院, 辽宁大连, 116000

2.大连理工大学, 计算机科学与技术学院, 辽宁大连, 116000

gjj@dmlu.edu.cn, sangguoming@dmlu.edu.cn

liuzhi@dmlu.edu.cn, zhangyijia@dmlu.edu.cn, hflin@dlut.edu.cn

摘要

研究突发公共卫生事件国际舆情演变规律, 对国际舆情资源进行应急管理和舆论疏导有重要借鉴价值。本文使用谷歌新闻数据库以各国针对COVID-19的报道为对象, 构建国际舆情数据集。采用主题模型、图神经网络模型, 结合时间、空间维度与舆情生命周期探究全球舆论主题-情感的演化态势, 模型准确率为0.7973, F1值为0.7826, 性能优于其他基线模型。研究发现, 各国舆情呈现放射传播状态。国际媒体舆论的情感倾向和讨论主题存在正相关且随时间进行转变。

关键词: 舆情分析; 突发公共卫生事件; 主题模型; 图神经网络; 时空演化分析

International Public Opinion Evolution Analysis on Sudden Public Health Events using Topic Model and Graph Neural Network

Jingjian Gao¹, Guoming Sang¹, Zhi Liu¹, Yijia Zhang^{*,1}, Hongfei Lin²

1.Dalian Maritime University, Dalian, Liaoning 116000, China

2. Dalian University of Technology, Dalian, Liaoning 116000, China

gjj@dmlu.edu.cn, sangguoming@dmlu.edu.cn

liuzhi@dmlu.edu.cn, zhangyijia@dmlu.edu.cn, hflin@dlut.edu.cn

Abstract

Studying the evolution of international public opinion in sudden public health emergencies has important reference value for emergency management and public opinion guidance of international public opinion resources. This article uses Google News database to construct a international public opinion dataset based on reports from various countries on COVID-19. Using topic model and graph neural network model, combined with time and space dimensions and public opinion lifecycle, the evolution trend of global public opinion topic emotion was explored. The accuracy of the model was 0.7973, with an F1 value of 0.7826, which outperformed other baseline models. Research has found that public opinion in various countries is in a radiating state of spread. The emotional tendencies and discussion topics of international media public opinion are positively correlated and change over time.

Keywords: Public opinion analysis, Sudden public health emergencies, Topic model, Graph neural network, Spatiotemporal evolution analysis

1 引言

*张益嘉 (通讯作者): zhangyijia@dmlu.edu.cn

基金项目: 辽宁省社会科学规划基金项目“新冠病毒疫情下的国际舆情分析研究”(L20BTQ008)

©2024 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

舆情是指群体针对特定突发事件所表达的态度、意见和情绪的总和。面对复杂的国际形势，各国针对公共卫生事件和社会保障事件等突发事件的态度有所不同，研究国际舆情有助于了解国际舆论风向，针对性地调整国家战略以应对可能的风险和危机。突发公共卫生事件具有突发性、急剧性、不可预测性等特征，往往会对人民群众的生命健康造成严重威胁，并引发公众恐慌、不满、愤怒等复杂情感(An et al., 2018)。媒体报道往往会带动公众情绪，从报道的情绪中识别公众意见倾向并融合主题信息，判断舆情的大体走向，把握舆情演化规律才能有效引导舆情(曾子明and 陈思语, 2023)。

自2020年初全球大范围的突发公共卫生事件“新冠肺炎疫情”的爆发以来，各国政府、媒体及大众对疫情相关信息高度关注。突发的国际舆情事件在公众关注度、情感态度、舆论导向等方面随时间动态演变，不同地区用户情感受该事件影响程度也有较大差异，并存在明显的跨地域传播。这都使得单一围绕时间-地点的信息切片，或仅面向单一主题的舆情分析无法完整反映突发公共卫生事件背景下的国际舆情演化全貌(边晓慧and 徐童, 2022)。

高效、全面、准确地分析国际舆情演化态势，对于政府在突发公共事件下有效疏导公共舆论具有重要价值。本文以新冠肺炎的国际媒体报道为研究对象，构建了国际舆情数据集，基于主题模型、图神经网络及舆情周期理论提出了国际舆情演化分析框架。采用多主题视角和全局视角分析舆情时空演化态势，挖掘舆论主题及情感倾向的跨地域动态扩散特征，进而使各国政府把握全球疫情焦点舆论和发展方向，结合本国舆情态势制定相关干预和引导舆论方向战略。

2 相关工作

2.1 舆情演化分析

舆情事件的演化往往具有生命周期(史波, 2010)。国内外研究者均对舆情的传播过程进行了探索，这些研究根据不同角度将舆情按照事件发生序列和发展生命周期来划分阶段并构建模型(甘宇祥et al., 2021)。Burkholder and Toole (1995)提出三阶段模型：潜伏期、扩散期、消退期。Canton (2006)基于三阶段模型结合医学知识提出著名的四阶段F模型：前驱期(Prodromal)、发作期(Breakout)、延续期(Chronic)、痊愈期(Resolution)。潘崇霞 (2011)提出简化的网络舆情演化三阶段模型：初始传播、迅速扩散、消退。上述采用生命周期理论对公共事件进行阶段划分的研究，对于挖掘网络舆情演化规律具有重要意义。

2.2 文本挖掘

随着文本挖掘技术的不断成熟，使用高效的文本挖掘技术处理大量的文本数据成为了主流，弥补了文本数据在计量分析和内容分析中需要大量人工的缺点。Huang et al. (2012)将LDA (Latent Dirichlet Allocation) 模型应用到微博短文本主题检测。Cheng et al. (2021)基于LDA模型对国务政策进行主题提取，研究政策主题演化。当前，LDA模型已经广泛应用于主题挖掘研究中，但该模型存在无法融合上下文信息的问题，导致主题挖掘性能欠缺。

2.3 情感分析

情感分析方法是国际舆情分析研究的关键技术，目前主要基于三种方法，即情感词典、机器学习和深度学习(史伟et al., 2022)。情感词典作为早期重要的情感分类工具，词典的精细程度决定了对文本情绪分类的准确性。曾子明and 万品玉 (2018)基于情感词典提出一种多层次融合模型，对文本进行正负情感的分类。使用情感词典进行情感分类有着明显的领域针对性以及泛化性较低的缺点。机器学习作为使用大量语料进行模型训练的方法，在情感分类领域有着较好的效果。邓君et al. (2020)基于Word2Vec和SVM对舆论进行二分类任务，该方法并没有利用语法结构的信息。Machuca et al. (2021)NLP技术融合回归算法针对舆情背景下的Twitter舆论进行分类。机器学习方法未考虑上下文语义信息对情感分类效果的影响。深度学习有着更好的特征提取能力，通过融合文本的上下文信息增强情感分类准确性。传统的深度学习模型在长期依赖问题中依旧乏力，Vaswani et al. (2017)构建Transformer模型，使用自注意力机制进行词向量嵌入对长期依赖问题有着较好的效果。为充分利用左右两侧的上下文信息，Devlin et al. (2018)基于微调的多层双向Transformer模型编码器构建了BERT(Bidirectional Encoder Representation from Transformers) 预训练模型，并用于提取文本特征，从而达到融合两侧信息的效果(Burkholder and Toole, 1995)。

综上，一方面，现有研究缺乏对国际舆情在主题和情感维度的动态探索。另一方面，以新冠疫情为背景的舆情分析研究，缺乏国际主题-情感的动态演化分析。本文以突发公共卫生事件为背景，关注国际媒体对于该事件报道的主题和情感维度随时间的变化。提出了国际舆情演化分析框架，采用主题模型，生成更符合当下舆论风向的主题分类，证明了图神经网络模型在情感分类任务的优越性。结合报道的时间和空间信息进行国际舆情时空演化分析，探索世界各国舆情演化之间的内在关联，为相关部门的政策制定提供理论支撑。

3 研究方法

本文以突发公共卫生事件为背景，构建国际舆情演化的研究框架(见图1)。

- 1) 数据采集与预处理：获得国际媒体关于“COVID-19”的报道，进行数据预处理。
- 2) 主题-情感信息挖掘：采用BERT-LDA挖掘舆情数据主题信息，由图神经网络模型挖掘舆情数据的情感倾向，并证明模型的优越性。
- 3) 舆情周期划分及时空演化分析：根据舆情演化三阶段模型，划分为初始传播、迅速扩散、消退。结合报道时空信息和主题-情感信息进行国际舆情时空演化分析及现象挖掘。

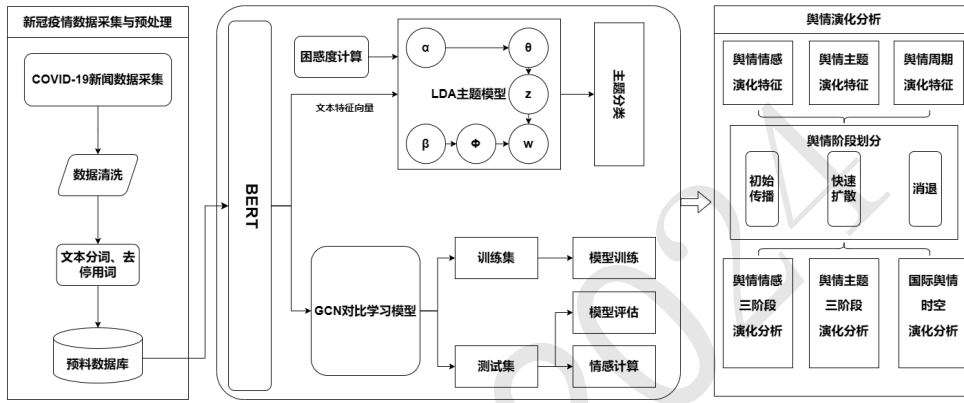


图 1: 国际舆情演化分析研究框架

3.1 图神经网络情感分类模型

本文采用图神经网络情感分析模型(见图2)对舆情文本进行情感分析。传统的词向量模型适合对短句和简单的句子进行分析。针对长文本情感分类，还应考虑上下文及单词与单词之间的关系的问题。首先，利用BERT预训练模型充分挖掘文本上下文特征信息，生成舆情文本的特征向量 V (见式1)，将所有文本的特征向量作为舆情图的节点集合：

$$\mathbf{V} = BERT(t) = [\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n] \quad (1)$$

通过余弦距离计算舆情图节点集合中不同节点之间的语义相似度(见式2)，当两节点之间距离在特定范围内时，构建其两点语义关系的边(见式3)。在构建的初始舆情图的基础上引入三层图神经网络模块学习节点嵌入进行特征融合，输出考虑相似文本节点之间相互作用的特征信息(见式4)(Zeng et al., 2023):

$$\text{CosineDistance} = \frac{\mathbf{V} \cdot \mathbf{V}'}{|\mathbf{V}| \times |\mathbf{V}'|} = \frac{\sum_{i=1}^n (\mathbf{V}_i \times \mathbf{V}'_i)}{\sqrt{\sum_{i=1}^n (\mathbf{V}_i)^2} \times \sqrt{\sum_{i=1}^n (\mathbf{V}'_i)^2}} \quad (2)$$

$$E_{ij} = \begin{cases} 1, 0 \leq \text{CosineDistance}_{ij} \leq 0.2 \\ 0, \text{ other} \end{cases} \quad (3)$$

$$v_i^{(l+1)} = \sigma \left(\sum_{j \in N_i} \frac{1}{c_{ij}} v_j^{(l)} w^{(l)} + b^{(l)} \right) \quad (4)$$

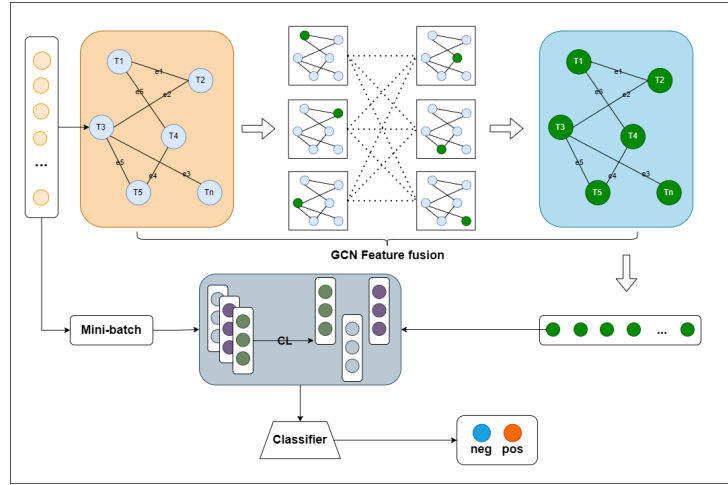


图 2: 基于图神经网络的情感分类模型

其中, t 为舆情文本数据, V 为舆情文本的特征向量, n 为向量维度, V'_i 为GCN生成的融合节点特征, E_{ij} 为语义关系边, $v_j^{(l)}$ 表示在第 l 层节点 j 的特征, σ 表示非线性变换, c_{ij} 为归一化因子, N_i 表示节点 i 和节点 i 的邻居, $W^{(l)}$ 表示第 l 层的权值, $b^{(l)}$ 表示第 l 层的截距。

为了提高模型在情感分析任务中的泛化能力, 本文引入对比损失函数, 基于源数据集的情感标签信息, 对隐藏向量进行对比学习。给定小批量文本的隐藏向量 $H = \{V_i\}_{i=1}^{N_b}$, 对于一个 $V_i \in H$, 如果 V_i 和 V_t 有相同的情感标签, 即 $y_t = y_i$ 则 V_i 被认为是 V_t 的正例, 而其他例子 $V_x \in H$ 被认为是负例。计算所有正例(V_t, V_i)和(V_i, V_t)的对比损失(见式5, 6)(Liang et al., 2022)。将最终的特征向量输入到分类器中进行分类, 使用交叉熵损失函数作为分类器损失函数:

$$\mathcal{L}_{cl} = \frac{1}{N_b} \sum_{V_i \in H} l(V_i) \quad (5)$$

$$l(V_i) = \log \left(\frac{\sum_{t=1}^{N_b} 1_{[t \neq i]} 1_{[y_t = y_i]} \exp \left(\frac{\text{sim}(V_t, V_i)}{\tau_s} \right)}{\sum_{x=1}^{N_b} 1_{[x \neq i]} \exp \left(\frac{\text{sim}(V_x, V_i)}{\tau_s} \right)} \right)^{-1} \quad (6)$$

其中 $1_{[t \neq i]} \in (0, 1)$ 为指示函数, $\text{sim}(V_t, V_i)$ 计算向量 V_t 和 V_i 的余弦相似度。

图神经网络模型利用BERT预训练模型进行文本的特征编码, 增大模型提取特征和融合上下文特征的能力。由于舆论文本的情感分类任务相同类别的不同文本之间存在一定的关联, 通过引入图神经网络融合不同文本的特征表示, 设置融合模块通过学习相邻节点的信息来更新节点自身的特征表示, 生成考虑相似文本节点之间相互作用的新特征, 可以有效的增加舆情情感分类的准确性。为增加模型的泛化能力充分利用文本相似性, 引入对比学习模块更好的将模型学习到的类别特征延伸至不同的目标数据, 进一步提高了分类的准确性。本文利用BERT预训练模型, 引入图神经网络特征融合及对比学习模块, 有效的利用舆情文本的相似性增强舆论情感特征提取能力, 在舆论情感分类任务中具有较好的性能。

3.2 舆论主题聚类

为了剖析全球新闻媒体的舆情演化趋势, 探究研究期间新闻媒体讨论的主要话题, 本文构建BERT-LDA模型进行主题细粒度分析。针对LDA模型无法有效提取上下文信息的问题, 本文通过BERT预训练模型提取文本特征, 并与LDA模型融合进行主题聚类分析, 实现报道主题与报道上下文结合。困惑度计算常用来评价LDA模型的预测和泛化能力, 困惑度越低表示模型性能越好(蔡政群 et al., 2022)。实验中, 困惑度计算(见式7)确定主题最优数 k 。

$$\text{Perplexity}(D) = \exp \left[\frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right] \quad (7)$$

其中 D 为文档所有词的集， M 为文档数量， W_d 为文档 D 中的词， N_d 为每个文档中 D 的词数， $p(w_d)$ 为文档中词出现的概率。

4 实证分析

4.1 数据收集与预处理

在深入挖掘新闻文本内部蕴含的主题-情感特征信息之前，本文选取谷歌新闻数据库GDELT (The Global Database of Events, Language, and Tone)设置关键检索字段“V1 THEMES”为“COVID-19”，本文认为在该事件爆发初期存在着较多的因无知所造成的影响，这其中缺乏人们对待这个事件理性的认知以至于存在较大的不确定性，最终选取实验时间区间为2022年7月1日-2022年8月31日，获取新闻来源、新闻获取链接、发布位置信息等属性信息，利用爬虫爬取指定链接的新闻报道。经数据预处理获取新闻事件120,679条，使用大规模语言模型对数据集进行情感倾向标注，其中消极为0积极为1，具体来说，本文在大模型的标注阶段使用的为通义千问作为数据标注的大模型，在标注阶段设置的prompt是以ICIO框架作为基础框架对其中的提示条件例如描述具体任务、任务的背景信息引导模型做出更好的回复、输入数据的格式以及输出数据的格式进行设置。为了验证大模型标签的有效性，我们采用人工验证选取部分数据一组为500条，邀请两位了解新冠肺炎事件的团队成员对两组数据进行标注，将两组人工标记数据分别与大模型标记数据进行对照试验，度量标准为kappa (k) 系数，其中结果分别为0.69、0.73，整体一致性较高，证明数据集标注的有效性。

4.2 舆论情感演化分析

4.2.1 情感分类模型评估

运用图神经网络模型对舆论进行情感分类，结合自身数据集特性模型参数设置见表1。为验证提出的情感分类模型在该任务上的优越性，选取四个对照实验证明模型的有效性，选取精确率P(Precision)、召回率R(Recall)、F1值(F1-Score)和准确率(Accuracy)作为评估指标。实验对比基本模型选择TF-SVM(王新宇, 2016)、LSTM(王楠and 吕欣隆, 2021)、BERT-SVM(严佩敏and 唐婉琪, 2020)、BERT-AAD(Ryu et al., 2022)，对比实验结果见表2。

参数名称	参数说明
GCN学习率	0.001
GCN隐藏层数	3
Batchsize	256
预训练次数	5
损失函数	交叉熵损失函数
函数优化	Adam算法

表 1: 模型参数设置

实验模型	模型对比侧重	P	R	F1	Acc
TF-SVM	词袋模型不能更好的提取文本特征	0.5312	0.5518	0.5366	0.5523
LSTM	无预训练模型且为单向记忆网络	0.6183	0.6337	0.6119	0.6350
BERT-SVM	未考虑相似文本之间的相互影响	0.6512	0.6734	0.6672	0.6833
BERT-AAD	未考虑相似文本之间的相互影响	0.7026	0.7219	0.7013	0.7230
BGCN	模型泛化性较差	0.7405	0.7463	0.7311	0.7387
BGCN-CL		0.7857	0.7880	0.7826	0.7973

表 2: 模型评估对比

实验结果表明，图神经网络情感分类模型在性能上优于其他模型。对比传统的词袋模型以及深度学习模型，引入BERT预训练模型可以更好的提取文本的特征用于分类任务。BGCN-CL模型在F1指数上比BERT-AAD模型有着较明显的提高，说明考虑相似文本之间的相互影响

是必要的。在去除CL对比学习模块后模型性能有所下降，证明了对比学习模块在模型中的有效性，可以有效的提升模型学习不同类别的特征，从而获得更好的模型泛化能力。综上，本文提出的模型可以有效的对舆情进行情感分析，情感倾向预测结果样例见表3

文本ID	文本部分内容	标签结果
20220706073000-544	From digital payments to video banking, the way consumers conduct basic financial transactions has changed substantially throughout the past decade, and the COVID-19 pandemic has only accelerated the trend toward digital banking.	NEG
20220803133000-750	CBC News LoadedBritish ColumbiaB.C.opens up immunizations to children 4 and under in next phase of COVID-19 fightInfants, toddlers and pre-schoolers can now get appointments for the first of two recommended Moderna Spikevax COVID-19 vaccine doses. Social SharingChildren between six months and four years old can now get appointments for the Moderna vaccine David P.	POS
20220831134500-741	Get our free mobile app According to their website, the program which launched this past January has not been funded by the Federal Government to continue. Because Congress hasn't provided additional funding to replenish the nation's stockpile,The good news is that you still have time to place your order for the free kits before Friday amp; apos; s final day. HOW TO GET FREE COVID-19 TESTING KITSEach household can order 8 free COVID-19 at-home testing kits for free.	POS

表 3: 预测结果样例

4.2.2 舆论情感态势分析

针对舆情新闻数据及同一时间新增病例数目进行计量分析(见图3)，由图可知舆情报道数量和感染病例数量有着直接关系，即随着当日确诊人数的增多，

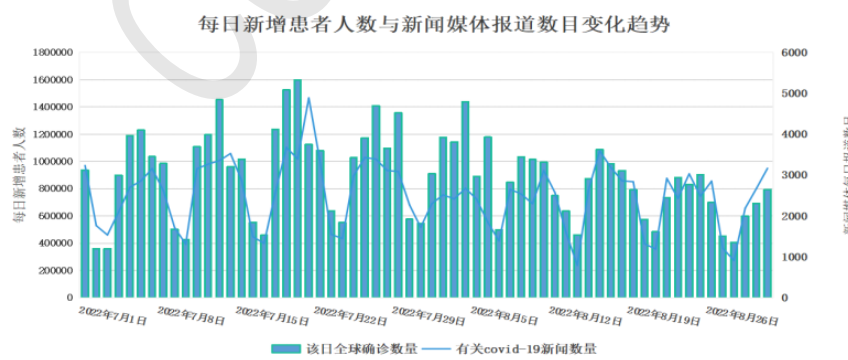


图 3: 新闻报道与新增病例数目统计

有关舆情的新闻报道数量也会有明显的增长趋势。因此，感染数目的变化会导致新闻媒体上关于舆情的报道数量变化(Burkholder and Toole, 1995)。图中曲线变化体现了从开始到爆发再到消退的三阶段现象，这也得益于舆情信息公开及在爆发初期政府进行策略优化，进一步消除公众的不知情恐慌与无端猜测，遏制负面情绪的增长(张冬and 魏俊斌, 2021)。

将新闻中的消极词汇和积极词汇进行统计，计算在报道中的情绪占比即Positive Score、Negative Score以日平均为单位进行可视化分析(见图4)。由图可知Positive Score与Negative Score变化趋势呈现明显的负相关关系，当新闻媒体报道中出现的带有积极情绪的词汇占比增加，报道所用的消极词汇的数量就会减少，反之亦然。

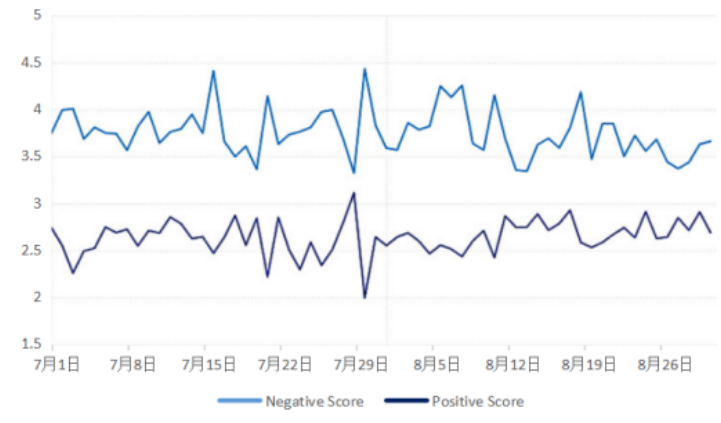


图 4: 消极词汇和积极词汇占比时序变化

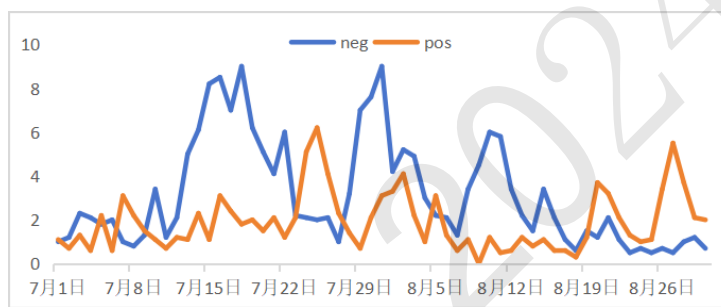


图 5: 情感倾向时序演化图

采用情感分析模型研究国际媒体情感演化过程，获得情感随时间变化趋势(见图5)。由图可知全球新闻媒体的新闻报道消极情感高于积极情感，且在2022年7月18日和2022年7月31日新闻报道的消极情绪达到顶峰。媒体的积极情感整体表现为局部波动，积极情感与消极情感大体呈现负相关的特点。舆情生命周期划分及特点见表4。

阶段	时间段	特点
初始传播	7月1日-7月13日	消极情绪呈上升趋势，积极情绪平稳而消极情绪有小幅震荡
快速扩散	7月13日-8月12日	消极和积极情绪均处于大幅度震动，消极情绪波动大于积极情绪，并存在负相关
消退	8月12日-8月31日	情绪态度趋于稳定，积极情绪逐渐增加

表 4: 舆情生命周期划分及特点

初始传播阶段，在突发公共卫生事件的背景下，全球经济形势不容乐观“印度洋岛国斯里兰卡总理维克勒马辛哈宣称斯里兰卡已经破产。他告诉国会，国家已经陷入深度衰退，食品、燃

料和药品的严重短缺将继续存在。”此事件的发生得到了全球世界各地的高度关注，使得群众对于突发公共卫生事件平稳后的经济状况感到担忧，但因信息传播速度影响，公众在该阶段对该事件的态度相对平稳。

迅速扩散阶段，突发公共卫生事件的出现使得全球经济出现新的挑战，中国作为全球第二大经济体深受影响，使得中国在国际上的政治外交变得困难。中美双边关系影响，导致欧美国家对中国的报道相对负面。美国众议院议长佩洛西不顾中方坚决反对，执意在时局动荡的情况下“访台”，严重践踏一个中国原则，损害我国的利益。中国始终坚持主权和领土完整原则，由于中美关系的影响可能导致欧洲和美国对中国的报道相对负面。在该时段内，国际关于突发公共卫生事件的讨论多以中国为主，即中国主流媒体为主要来源。随着国家及主流媒体引导在短时间里出现积极情绪主导的状态。以美为首的西方发达国家因新冠病毒限制中国外交，强行干预国家内政及大肆宣扬突发公共卫生事件为中国病毒导致国际舆论恶化。由于外交和病毒等多方原因造成围观热度和舆论压力，使得公众情绪在恐慌和抱怨中逐渐转向以负面情绪为主导。

消退阶段，由于国家增加对突发公共卫生事件舆情的管控以及猴痘疫情逐渐蔓延，人们将关注的重心逐渐转移到其他相关症状上，虽然有关该事件的舆论热度逐渐消退，公众情绪呈现积极向上的状态，但负面情绪仍有波动性起伏。

4.3 舆论主题演化分析

为直观分析舆情热点制作新闻热点词云图(见图6)。词云图展现国际更多关注卫生防控及病毒疫苗的研发，带有明显情感倾向的词语出现较少，“epidemic”、“N95”、“prevention”、“vaccine”、“control”等关键词表明国际讨论重心着眼于如何预防病毒以及治疗病毒、如何抗击病毒等卫生防控角度，反映该事件舆情态势目前较于稳定。本文意在研究舆论主题随时间的转变，探究各个阶段主题的动态变化获取国际舆论风

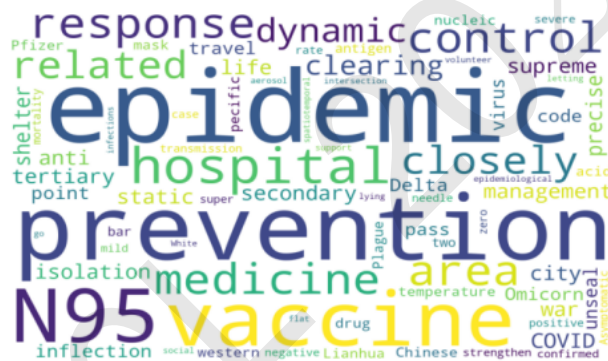


图 6: 热点词云图

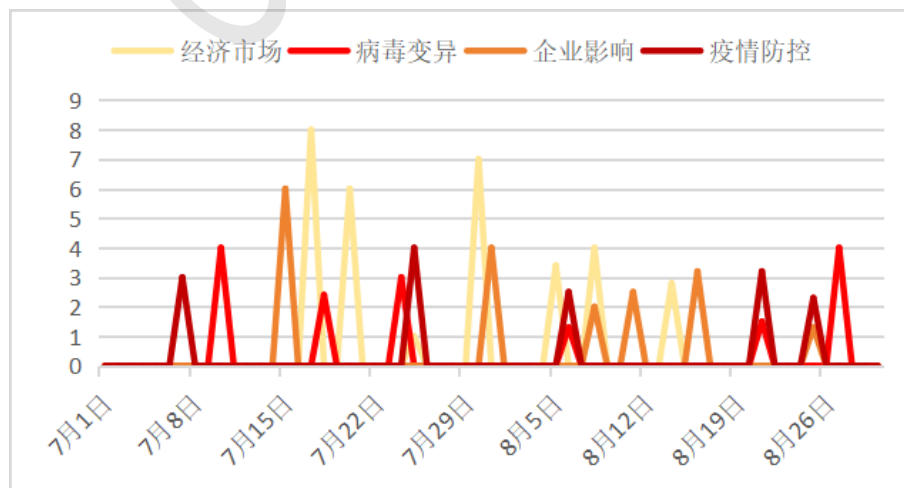


图 7: 主题时序演化图

向。基于困惑度计算确定最优主题数并进行主题分类得到频率均值最高的四组主题分别为“经济市场”、“病毒变异”、“企业影响”、“疫情防控”，由潜在主题中出现词汇概率排位较高的十个特征词来表示舆论主题，见表5-7。计算不同阶段下新闻报道主题与上述四组主题的相关性，统计每个主题出现频率随时间的动态演变(见图7)。

初始传播阶段，舆论倾向病毒变异和病毒防控两个部分。病毒并没有大规模爆发，讨论如何应对新一轮的卫生防控成为首要问题。结合情感倾向结果分析出媒体舆论情感相对稳定，正负情感均有波动但幅度较小。政府应加强病毒知识科普，完善公众防护措施控制舆论风向。

主题	关键词
病毒变异	'covid', 'virus', 'health', 'hospital', 'epidemic', 'N95', 'delta', 'omicron', 'policy', 'public'
疫情防控	'epidemic', 'prevention', 'N95', 'covid', 'health', 'sports', 'vaccine', 'quarantine', 'public', 'positive'

表 5: “初始传播阶段主题-词”

快速扩散阶段，随着病毒的大规模扩散对社会所造成的经济损失不断增加，股票下跌、企业盈亏的现象增加导致国际对突发公共卫生事件下的经济市场及企业影响的讨论增加。病毒变异及防控的讨论随着扩散规模的严重也显得尤为重要。主题词中的“negative”，“decline”反映出世界经济形势较差，结合情感倾向结果分析媒体舆论倾向较为消极且情感波动较大。政府应积极制定经济相关政策，引导主流媒体对该事件积极信息的报道传达主流价值观，减少公众因消极信息而带来的恐慌提升政府公信力。

主题	关键词
经济市场	'pneumonia', 'amp', 'covid', 'business', 'market', 'china', 'area', 'policy', 'republic', 'negative'
病毒变异	'covid', 'virus', 'health', 'hospital', 'epidemic', 'N95', 'delta', 'omicron', 'policy', 'public'
企业影响	'enterprise', 'Biden', 'business', 'market', 'policy', 'response', 'decline', 'amp', 'pandemic', 'evolution'
疫情防控	'epidemic', 'prevention', 'N95', 'covid', 'health', 'sports', 'vaccine', 'quarantine', 'public', 'positive'

表 6: “快速扩散阶段主题-词”

消退阶段，媒体更加关注大规模卫生健康问题对企业经济的影响以及防控措施和病毒知识的科普。结合情感倾向结果分析媒体舆论情感整体较为积极，源于政府积极普及防疫措施和疫苗注射并取得了较好的效果，有效的减小该事件所造成的危害。通过主流媒体宣传防疫成果引导社会舆论倾向，使得公众对待该事件的态度出现由消极到积极的转变。

主题	关键词
病毒变异	'covid', 'virus', 'health', 'hospital', 'epidemic', 'N95', 'delta', 'omicron', 'policy', 'public'
企业影响	'enterprise', 'Biden', 'business', 'market', 'policy', 'response', 'decline', 'amp', 'pandemic', 'evolution'
疫情防控	'epidemic', 'prevention', 'N95', 'covid', 'health', 'sports', 'vaccine', 'quarantine', 'public', 'positive'

表 7: “消退阶段主题-词”

4.4 舆情主题-情感时空演化分析

为动态观察各国舆论主题随时间变化规律，根据上文划分的阶段，对不同阶段各个国家的主题向量进行聚类按照时间序列进行可视化展示，绘制全球地图其中国家颜色与图7主题颜色对应(见图8)。

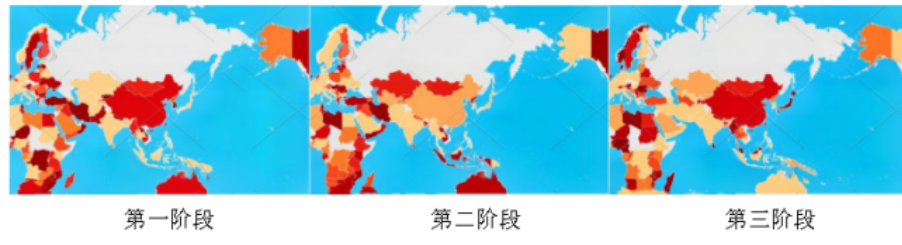


图 8: 主题-情感时空演化图

由图8可知，突发公共卫生事件舆情在国际上具有以舆情处于初始爆发阶段国家为中心，向周围国家放射性传播舆情态势的特点。以中国即周边国家为例，在中国周边国家处于舆情的迅速扩散阶段时，中国境内正在处于初始传播阶段。随着舆情在周边国家的发展，中国由初始传播阶段进入到迅速扩散阶段，政府的积极治理使得中国舆情由迅速扩散到消退的转变。对比多个国家发现中国在由初期传播到消退的转变明显，反映出中国虽作为发展中国家，但对于突发卫生事件的防控及认真态度站在世界前列，中国各级政府积极响应中央号召，果断实施相应政策，极大缓解因疫情造成的经济损失及公众恐慌，为世界其他国家的卫生防控提供了一套重要的管理模式，为世界公共卫生安全做出了重要的贡献。

5 总结与展望

本文针对与“COVID-19”相关新闻报道构建了从“国际舆情数据收集及预处理”、“舆情情感分类”、“舆情主题聚类”、“国际舆情周期演化分析”的全过程研究框架，揭示了突发性公共卫生事件在全球各时间段的发展态势，通过对比实验证实了模型的优越性。以“初始传播、迅速扩散、消退”三个阶段在时间维度上进行研究，发现在初始传播和消退阶段的新闻报道在情感和讨论主题上处于相对稳定的状态，而在快速扩散阶段由于确诊数量的不断增加，也增大了各国媒体对该事件的担忧程度，从而使得各国新闻的情感和讨论主题更加偏向于消极的态度。在全球空间维度研究中，针对各国媒体的舆论主题和情感倾向，结合三阶段时间维度动态观察每个阶段国家之间的相互影响。分析实验结果得出结论，突发性公共卫生事件背景下各国之间存在舆情态势放射性传播的现象，当目标国家出现并爆发大规模疫情后随着时间的推移，邻国也会慢慢的由相对平稳的初始传播阶段转入快速扩散阶段。该发现有利于各国政府在邻国出现疫情问题后，及时地制定相关外来人员入境政策以防止国内大规模疫情的爆发。

本文存在一些不足，如选取的数据时间跨度相对较短，无法更加多样、更加深层的挖掘舆情演化态势。数据类型局限于报道文本，较为单一。今后的研究会着重考虑增加数据类型，引入多个模态的数据更加深层的挖掘舆情演化。扩大研究数据集深层挖掘舆情时间维度信息，为相关部门提供更全面的舆情现象分析及理论支撑。

参考文献

- Lu An, Chuanming Yu, Xia Lin, Tingyao Du, Liqin Zhou, and Gang Li. 2018. Topical evolution patterns and temporal trends of microblogs on public health emergencies: an exploratory study of ebola on twitter and weibo. *Online information review*, 42(6):821–846.
- Brent T Burkholder and Michael J Toole. 1995. Evolution of complex disasters. *The Lancet*, 346(8981):1012–1015.
- Lucien G Canton. 2006. San francisco 1906 and 2006: An emergency management perspective. *Earthquake spectra*, 22(2_suppl):159–182.

- Quan Cheng, Jianhua Kang, and Minwang Lin. 2021. Understanding the evolution of government attention in response to covid-19 in china: A topic modeling approach. In *Healthcare*, volume 9, page 898. MDPI.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Bo Huang, Yan Yang, Amjad Mahmood, and Hongjun Wang. 2012. Microblog topic detection based on lda model and single-pass clustering. In *International Conference on Rough Sets and Current Trends in Computing*, pages 166–171. Springer.
- Bin Liang, Qinlin Zhu, Xiang Li, Min Yang, Lin Gui, Yulan He, and Ruifeng Xu. 2022. Jointcl: A joint contrastive learning framework for zero-shot stance detection. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 81–91. Association for Computational Linguistics.
- Cristian R Machuca, Cristian Gallardo, and Renato M Toasa. 2021. Twitter sentiment analysis on coronavirus: Machine learning approach. In *Journal of Physics: Conference Series*, volume 1828, page 012104. IOP Publishing.
- Minho Ryu, Geonseok Lee, and Kichun Lee. 2022. Knowledge distillation for bert unsupervised domain adaptation. *Knowledge and Information Systems*, 64(11):3113–3128.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Ziming Zeng, Shouqiang Sun, and Qingqing Li. 2023. Multimodal negative sentiment recognition of online public opinion on public health emergencies based on graph convolutional networks and ensemble learning. *Information Processing & Management*, 60(4):103378.
- 严佩敏 and 唐婉琪. 2020. 基于改进bert的中文文本分类. *工业控制计算机*, 33(7):4.
- 史伟, 薛广聪, and 何绍义. 2022. 情感视角下的网络舆情研究综述. *图书情报知识*, 39(1):105–118.
- 史波. 2010. 公共危机事件网络舆情应对机制及策略研究. *情报理论与实践*, (7):4.
- 张冬 and 魏俊斌. 2021. 情感驱动下主流媒体疫情信息数据分析与话语引导策略. *图书情报工作*, 14.
- 曾子明 and 万品玉. 2018. 融合演化特征的公共安全事件微博情感分析. *情报科学*, 36(12):3–8.
- 曾子明 and 陈思语. 2023. 基于lda与bert-bilstm-attention模型的突发公共卫生事件网络舆情演化分析. *情报理论与实践*, 46(9):158–166.
- 潘崇霞. 2011. 网络舆情演化的阶段分析. *计算机与现代化*, (10):203–206.
- 王新宇. 2016. 基于情感词典与机器学习的旅游网络评价情感分析研究. *计算机与数字工程*, 44(4):578–582.
- 王楠 and 吕欣隆. 2021. Covid-19 疫情下政务媒体情感传播特征及用户情感体验研究. *情报探索*, 1(2):1.
- 甘宇祥, 王亚博, 薛均晓, 张若琪, 许书宁, and 郭毅博. 2021. 基于情感特征的新冠肺炎疫情舆情演化分析. *图学学报*.
- 蔡政群, 李诗轩, 吴玉敏, and 杨斯涵. 2022. 基于lda模型的学术次生衍生舆情传播的演化分析. *武汉理工大学学报:信息与管理工程版*, 44(3):395–401.
- 边晓慧 and 徐童. 2022. 重大突发公共卫生事件下的公众情感演进分析:基于新冠肺炎疫情的考察. *数据分析与知识发现*, (7):128–140.
- 邓君, 孙绍丹, 王阮, 宋先智, and 李贺. 2020. 基于word2vec 和svm 的微博舆情情感演化分析. *情报理论与实践*, 43(8):112–119.