

# 基于情感特征的新冠肺炎疫情舆情演化分析

甘宇祥<sup>1</sup>, 王亚博<sup>2</sup>, 薛均晓<sup>2</sup>, 张若琪<sup>3</sup>, 许书宁<sup>2</sup>, 郭毅博<sup>4</sup>

- (1. 郑州联大教育集团, 河南 郑州 450001;
2. 郑州大学软件学院, 河南 郑州 450002;
3. 河南师范大学软件学院, 河南 新乡 453007;
4. 郑州大学信息工程学院, 河南 郑州 450001)

**摘 要:** 针对突发事件的舆情演变态势进行分析, 发现社会舆情的演变规律, 提出了一种基于情感特征的舆情演化分析方法, 该方法包含舆论情感分析模块与舆情演化分析模块。舆论情感分析模块基于 BERT 预训练模型和 BiGRU 模型, 其中 BERT 作为词嵌入模型提取舆情文本特征向量, BiGRU 则用于提取文本特征向量的上下文联系实现对舆情数据情感极性的精准判别。在舆情演化分析模块中, 将舆情的情感特征在时间维度上进行动态可视化建模, 并基于其结果实现舆情数据的演化规律解析。在实验部分, 利用 2020 年 1 月 1 日到 2020 年 2 月 19 日的 100 万条新冠肺炎背景下的舆论数据进行了数值实验, 实验结果表明, 该方法能够有效地对疫情背景下的舆情数据进行演化分析。

**关 键 词:** 新冠肺炎; 舆情情感分析; 舆情演变分析

中图分类号: TP 391

DOI: 10.11996/JGJ.2095-302X.2021020222

文献标识码: A

文章编号: 2095-302X(2021)02-0222-08

## Public opinion evolution analysis of “COVID-19 epidemic” based on sentiment feature

GAN Yu-xiang<sup>1</sup>, WANG Ya-bo<sup>2</sup>, XUE Jun-xiao<sup>2</sup>, ZHANG Ruo-qi<sup>3</sup>, XU Shu-ning<sup>2</sup>, GUO Yi-bo<sup>4</sup>

- (1. Zhengzhou United Education Group, Zhengzhou Henan 450001, China;
2. School of Software, Zhengzhou University, Zhengzhou Henan 450002, China;
3. College of Software, Henan Normal University, Xinxiang Henan 453007, China;
4. School of Information Engineering, Zhengzhou University, Zhengzhou Henan 450001, China)

**Abstract:** In order to analyze the evolution of public opinion under emergencies and discover the law of the evolution of public opinion, a sentiment feature-based public opinion evolution analysis method was proposed, including a News Sentiment Analysis Module and a Public Opinion Evolution Analysis Module. The News Sentiment Analysis Module was based on the BERT pre-training model and the BiGRU model, where BERT was extracted as a word embedding, and BiGRU was employed to extract the contextual links of the textual feature vector to achieve a better understanding of the sentiment polarity of public opinion data. In the Public Opinion Evolution Analysis Module, this paper modeled the dynamic visualization of the sentiment features of public opinion in the time dimension, and then

收稿日期: 2020-09-16; 定稿日期: 2020-10-23

Received: 16 September, 2020; Finalized: 23 October, 2020

基金项目: 国家自然科学基金项目(6160051017); 国家重点研发计划项目; 河南省高等学校青年骨干教师培养计划

Foundation items: National Natural Science Foundation of China (6160051017); National Key R & D Plan; Plan for Young Backbone Teachers in Henan Province

第一作者: 甘宇祥(1967-), 男, 河南信阳人, 高级工程师, 硕士。主要研究方向为大数据与人工智能。E-mail: frencgan@126.com

First author: GAN Yu-xiang (1967-), male, senior engineer, master. His main research interests cover data and artificial intelligence.

E-mail: frencgan@126.com

通信作者: 薛均晓(1982-), 男, 河南南阳人, 副教授, 博士。主要研究方向为人工智能、网络空间安全等。E-mail: xuejx@zzu.edu.cn

Corresponding author: XUE Jun-xiao (1982-), male, associate professor, Ph.D. His main research interests cover artificial intelligence, cyberspace security, etc. E-mail: xuejx@zzu.edu.cn

based on the visualization results, enabled the resolution of evolutionary patterns of public opinion data. Finally, a numerical experiment was conducted using one million pieces of the COVID-19 news data from January 1, 2020 to February 19, 2020. The experimental results show that the method proposed in this paper can effectively analyze the sentiment polarity of public opinion data.

**Keywords:** COVID-19; analysis of public sentiment and emotion; analysis of public opinion evolution

随着微博、抖音等网络社交媒体的兴起,每个网民均可以作为舆论源进行信息的发表、阅读与传播。社交媒体的兴起一方面给人们带来巨大便利,另一方也带来了安全隐患。

当重大公共事件突发时,网民们往往不经过深思熟虑,便会在网上发表自己的见解,表达其对事件的情感态度,从而形成网络舆论<sup>[1]</sup>。自新冠肺炎疫情暴发以来,大众对疫情相关信息高度关注,在这样的背景下,社交媒体在信息聚合、舆论生成方面扮演的角色越来越重要。通过分析新冠肺炎疫情下的网络舆论情感极性,可以获得公众对新冠肺炎事件的主要情感观点和价值取向,从而使相关部门能够更加准确地把握特定事件的焦点舆论和发展方向,进一步辅助其进行分析、决策,从而更有效地干预和引导舆论方向来实现科学战“疫”。

文本情感分析指利用自然语言处理(natural language processing, NLP)和文本挖掘技术,对具有主观情感色彩的文本进行分析、处理、归纳和推理的过程<sup>[2]</sup>。就网络舆论情感分析而言,主要有基于情感维度模型的网络舆情信息分级等理论模型研究,情感倾向性分析、有如何构建网络舆情情感词典、用户情感和关系网络演化等技术应用研究,还有针对热门事件网络舆情情感分析的案例研究,形成了完善的网络舆情情感分析研究体系。所使用的情感分析技术包括基于机器学习和深度学习2种方法,其中机器学习方法使用 TF-IDF 等文本特征加传统的机器学习模型;深度学习方法使用卷积神经网络(convolutional neural network, CNN)和递归神经网络(recurrent neural network, RNN)模型去自动地提取文本特征。

本文基于情感分析和数据可视化方法,对新冠疫情期间的舆论情感态势进行研究。本文基于情感特征对新型冠状病毒肺炎疫情期间的网络舆论进行情感分析,构建了疫情期间的舆论情感极性分析模型。并结合相关舆论的情感特征对疫情期间的舆情信息进行了演变研究,从而更好地把握舆情演变的规律,提出针对性的方法应对舆情。为舆情系

统的治理提供帮助,加强相关部门应对舆情的管理能力。

## 1 相关的工作

### 1.1 自然语言处理

NLP 是用计算机来处理、理解人类语言,其属于人工智能的一个分支,是计算机科学与语言学的交叉学科,又称为计算语言学。NLP 技术可以将非结构化的文本转化为结构化信息<sup>[3]</sup>,并允许计算机通过算法来理解人类语言。从研究内容看,NLP 包括语法和语义分析、篇章理解等。从应用角度看,NLP 具有广泛的应用前景。特别是在信息时代,NLP 的应用包罗万象,本文所涉及到的便是 NLP 中的文本情感分析。

### 1.2 文本情感分析

文本将情感分析方法分为:基于情感词典、基于机器学习和基于深度学习的情感分析<sup>[3]</sup>3类。其中,基于情感词典的方法主要依赖的是人工搭建的情感词典,周咏梅等<sup>[4]</sup>提出了构建基于 HowNet 和 SentiWordNet 的中文情感词典的方法,该方法在微博文本情感分析任务取得了不错的效果。但其属于无监督学习,会导致构建情感词典困难,从而消耗了大量的时间和精力,但并非广泛适用。

PANG 等<sup>[5]</sup>2002 年第一次在电影评论数据集上将机器学习的方法应用在文本情感分类任务中。常用的机器学习方法包括朴素贝叶斯、K 最近邻、最大熵等。但这类方法需要人工构造大量特征从而获得更高、更好的分类效果,不仅如此,还需要大量的专业知识,此外仍存在泛化能力不足的问题,导致了此类方法只适用于特定场景。

而基于深度学习的情感分析方法则不需要特点领域的专业人士进行分析并确定分类特征,该方法的扩展性得到大大地提升<sup>[6]</sup>。最初应用于计算机视觉和语音识别方向,然后扩展到了情感分析的领域,并成为该领域的热门技术。HEIKAL 等<sup>[7]</sup>通过构建 CNN 模型和 LSTM 模型进行情感分析。梁军等<sup>[8]</sup>基于情感极性转移模型的文本情感分析算法的

实验结果显示,其较普通 LSTM 和 RNN 模型有更好的效果。传统的深度学习模型是将特征设置为相同的权重进行训练,从而导致无法更好地关注短文本中对情感类别贡献较为突出的特征。随后 Google Mind<sup>[9]</sup>团队首次提出自注意力机制,由于该机制依赖参数少,且能够更好地帮助模型获得文本特征,表现出优秀的文本处理能力,也因此广泛应用于文本数据挖掘中。

### 1.3 舆情演化分析

舆情事件的演化过程往往具有生命周期<sup>[10]</sup>。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。其中比较突出的传播模型包括 BURKHOLDER 和 TOOLE<sup>[11]</sup>提出的三阶段模型、文献<sup>[12]</sup>在三阶段模型基础上提出的四阶段模型;方付建<sup>[13]</sup>将网络舆情发展历程划分为孕育、扩散、变换和衰减 4 个阶段。潘崇霞<sup>[14]</sup>将演化历程简化为初始传播、迅速扩散和消退 3 个阶段。同时学者们也对突发事件网络舆情各个阶段的特征进行了相应地分析。

### 1.4 门控循环单元

门控循环单元(gated recurrent unit, GRU)是循环神经网络的改进。其通过引入更新门和重置门机制,有效地解决了 RNN 在训练过程中出现的梯度爆炸和梯度弥散问题<sup>[15]</sup>。相较于长短期记忆网络(long short-term memory, LSTM),GRU 简化了神经网络结构,减少了模型的参数,并提高了训练速度(图 1)。在解决序列数据处理任务中,GRU 网络可以学习句子中单词的长期依赖性,还可以更好地对文本进行表征和建模<sup>[16]</sup>。GRU 网络不仅可以通过存储单元存储句子中的重要特征,同时还能选择性地忘记一些不重要的信息。在 GRU 网络中,每个神经元包括 1 个存储单元和 2 个门单元。

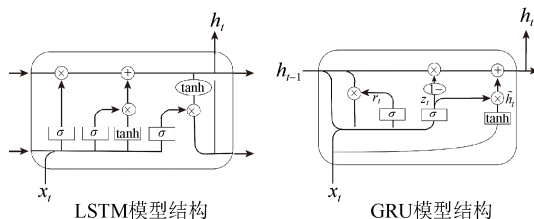


图 1 GRU 相比 LSTM 的改进

Fig. 1 Improvement of GRU compared to LSTM

### 1.5 BERT

目前,在 NLP 领域,现有模型大多使用 Word2Vec 或 Glove 等词向量训练工具,然而这些模型训练的

词向量本身就具有一定的自身缺陷,是属于静态编码的一种,同一单词在不同的上下文语义环境中会表达相同的含义,这使得模型对语义的理解会产生一定的偏差。

为了充分利用左右两侧的上下文信息,体现句子语义的空间相互关系,2018 年 10 月,Google 的 DEVLIN 等<sup>[17]</sup>提出了预训练模型 BERT (bidirectional encoder representation from transformers)(图 2),该语言表示模型采用了基于微调的多层双向 Transformer 编码器,并用于进行文本特征的提取,从而达到融合字左右两侧信息的效果<sup>[18]</sup>。自注意力机制是 BERT 编码器的核心,可以得到词的双向表示,并创新性地提出了 Masked 语言模型与下一个句子预测任务。BERT 模型输入的是 3 个向量的和,对于每一个输入的词,其表征包括词向量(token embeddings)、分段向量(segment embeddings)和位置向量(positions embeddings) 3 部分。BERT 向量相比于 Word2Vec 等模型词向量固定不变以及短距离单向的前后文信息融合,其融合了其他模型的优点,同时也替代了其缺点,能够充分结合上下文动态生成特征向量,表征字的歧义性,因此该模型在多种 NLP 的后续特定任务上取得了良好的效果<sup>[19]</sup>。

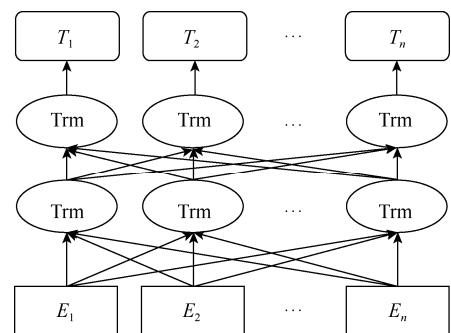


图 2 BERT 结构图

Fig. 2 BERT structure diagram

## 2 方法介绍

### 2.1 方法概述

为了能够更好地对疫情期间的舆论数据进行舆论情感极性分析与演化分析,如图 3 所示,本文提出了基于情感特征的疫情期间舆论情感极性分析算法来对疫情期间的舆论数据进行情感识别,在 BERT 模型后面加了一层 BiGRU,用来更好地捕获词向量之间的上下文联系,并对疫情期间的舆论数据进行精准的情感极性识别。接着联合疫情期间感染数据对舆情数据进行演化分析。通过对负面舆论

每日变化数据与疫情期间每日感染数据进行关联来探寻舆情演变规律,并通过计算词频的方式探究不同情感极性下的舆论热点。

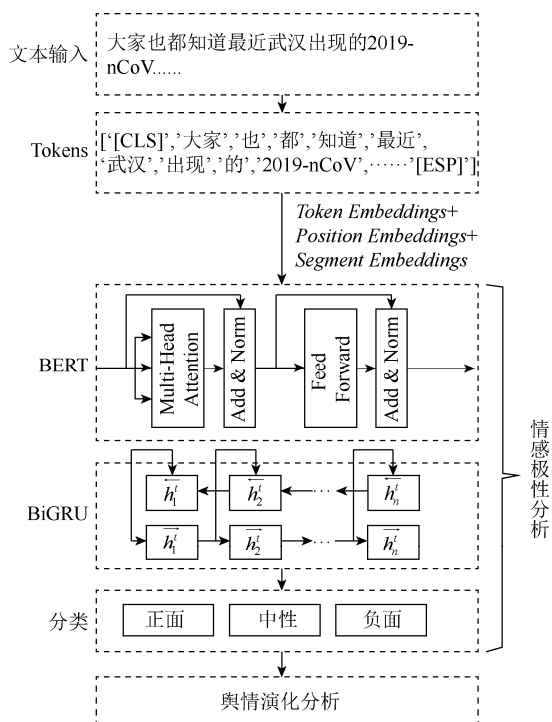


图3 方法结构图

Fig. 3 Method structure diagram

## 2.2 方法推导

传统的词向量模型适合对短句和简单的句子进行分析。为了解决一词多义的问题,还应考虑上下文文本及单词与单词之间的关系。BERT模型是Google提出的语言表示模型,与传统文本情感分析相比,BERT可以更好地覆盖上下文之间的联系,本文使用Google Research发布的预训练模型。

通过BERT Base提取文本特征。对于输入的文本可使用Bert进行特征提取,即

$$C_n = BERT(t_n) \quad (1)$$

模型取[CLS]标记在BERT训练的最后一层的输出 $C$ ,加上权重 $W$ 作为双向GRU模型的输入,即

$$a_i = g(W_a \times c_i + b) \quad (2)$$

其中,  $1 \leq i \leq n$ ,  $n$ 为BERT输出的特征维度; $b$ 是偏置量;激活函数 $g$ 为Sigmoid函数。

模型将输入向量送入BiGRU中,使用2个GRU从2个不同的方向计算向量序列,最后将2个方向的结果进行合并输出,即

$$h_i = BiGRU(a_i) \quad (3)$$

然后使用Softmax函数对BiGRU输出的特征向量进行分类,得到最后的情感极性识别结果。

## 3 实验

### 3.1 实验数据集介绍

本实验选取的数据集为中国计算机学会所举办的疫情期间网民情绪识别比赛中所提供的数据集,该数据集依据与“新冠肺炎”相关的230个主题关键词进行数据采集,抓取了2020年1月1日至2020年2月20日期间共计100万条微博数据,其中10万条为标注数据,被分为消极、积极与中性3类,见表1。在本文实验中,本文使用其中10万条标注数据作为训练集来训练疫情期间舆情情感极性分模型。然后利用训练得到的模型对另外90万条未标注数据进行情感极性的识别,并利用可视化技术研究舆情演化与疫情之间的潜在关系。

表1 部分数据集展示

Table 1 Partial data set presentation

| 时间              | 舆论数据内容                                                        | 标签 |
|-----------------|---------------------------------------------------------------|----|
| 01月08日<br>15:38 | 九旬老人拿放大镜写感谢信#: 病痛时你们护我周全,康复时我要说声谢谢                            | 正面 |
| 01月23日<br>22:44 | 现在有横扫附近药店高价出售口罩的小人;现在有生病还四处乱跑的病人;现在有很多制造谣言和跟风谣言的人;现在还有不劝劝的一些人 | 负面 |
| 02月18日<br>14:10 | 世卫组织在日内瓦召开发布会,介绍新冠肺炎疫情的最新情况。会上,世卫组织总干事谭德塞表示                   | 中性 |

同时为了验证本文方法的有效性,人们在开源的潭松波酒店评论数据集上进行了模型性能的测试。该数据集共6000条样本,其中情感积极样本3000条,负面情感样本3000条。

### 3.2 实验参数

本文使用Tensorflow 2.0搭建疫情期间舆情情感极性分析模型,使用的硬件设备为NVIDIA RTX2080Ti。模型中的词嵌入部分为Google开源的BERT预模型,其参数见表2。本文使用的损失函数为交叉熵损失函数,并使用Adam算法对损失函数进行优化。

表2 模型参数设置

Table 2 Model parameter setting

| 参数名称       | 具体设置 |
|------------|------|
| 学习率        | 1e-5 |
| Batch Size | 16   |
| 迭代次数       | 2    |
| 文本最大长度     | 140  |

### 3.3 情感极性分析实验

为了验证本文方法的有效性, 将其与几种主流的实验方法进行了实验结果对比, 并使用了准确率、召回率和 F1 Score 等指标对实验结果进行评测。召回率是某类判断正确的数目除以测试集中该类的数目。

(1) TF-IDF+LR: TF-IDF 用来评估字词对于一份文件集或一个语料库中的其中一份文件的重要程度。字词的重要性随着在文件中出现的次数成正比增加, 但同时会随着在语料库中出现的频率成反比下降<sup>[20]</sup>。本文先提取文本的 TF-IDF 特征, 然后使用 LR 进行分类。

(2) LSTM: 长短期记忆网络是一种特殊的 RNN, 主要目的是解决长序列训练过程中的梯度消失问题<sup>[21]</sup>。

(3) TextCNN: 将 CNN 应用到文本分类任务, 利用多个不同 size 的 kernel 来提取句子中的关键信息, 从而更好地捕捉局部相关性<sup>[22]</sup>。

(4) BERT-Base: BERT 是谷歌发布的大规模文本预训练模型, 这里使用的 BERT-Base 仅为 BERT 预训练模型提取文本特征后不做任何处理, 然后使用 Softmax 进行文本分类。

按照 1:4 的比例切分训练集和验证集, 其中训练集用于疫情期间舆论情感极性模型的训练, 验证集则用来验证本文方法的有效性。其对比实验结果见表 3。

表 3 不同实验方法的性能表现  
Table 3 The performance of different experimental methods

| 数据集         | 方法        | 准确率   | 召回率   | F1 Score |
|-------------|-----------|-------|-------|----------|
| 疫情期间网民情感数据集 | TF-IDF+LR | 0.608 | 0.601 | 0.599    |
|             | LSTM      | 0.652 | 0.638 | 0.640    |
|             | TextCNN   | 0.641 | 0.633 | 0.631    |
|             | BERT-Base | 0.749 | 0.709 | 0.713    |
|             | 本文        | 0.753 | 0.714 | 0.716    |
| 酒店评论数据集     | TF-IDF+LR | 0.813 | 0.806 | 0.801    |
|             | LSTM      | 0.865 | 0.860 | 0.858    |
|             | TextCNN   | 0.869 | 0.862 | 0.870    |
|             | BERT-Base | 0.933 | 0.929 | 0.930    |
|             | 本文        | 0.941 | 0.942 | 0.939    |

从表 3 可看出, 本文方法在疫情期间的舆论情感分析中相比其他方法, 在 4 个指标上均有较大程

度的提高。其原因: ①在大规模数据上, 训练得到的 BERT 预训练模型能够涵盖到更多的信息来应对突发的疫情期间舆论数据; ②本文方法在引入 BiGRU 之后可以更加有效地提取舆情数据中词与词之间的联系。

### 3.4 疫情期间舆情数据情感演化分析

使用前文训练所得到的疫情期间舆论情感分析模型对另外 90 万未标注的舆论数据进行预测, 然后将该数据和新冠疫情期间每日新增人数与患者总数目进行关联与可视化分析。各类数据占比如图 4 所示。其部分预测结果见表 4, 可以看出对于这 90 万的未标注数据, 本文模型也能实现较为精确地识别结果。其中, 0 代表负面数据、1 代表中性数据、2 代表积极数据。

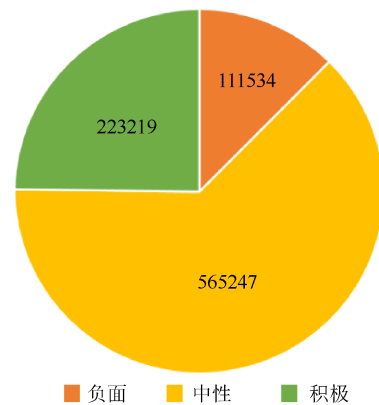


图 4 各类情感极性舆论数据的占比

Fig. 4 Percentage of various emotional polar news

表 4 部分预测数据展示  
Table 4 Part of the forecast data display

| 时间                 | 舆论内容                                        | 标签 |
|--------------------|---------------------------------------------|----|
| 01 月 21 日<br>23:10 | 武汉出现的 2019-nCoV, 建议不要恐慌, 学会去预防              | 中性 |
| 01 月 21 日<br>23:10 | 没有什么困难是可以打倒伟大的中华民族的, 为奋斗在一线的各位医护人员和工作者祈福    | 积极 |
| 01 月 21 日<br>23:11 | 感觉自己鼻塞, 呼吸困难, 喉咙有点疼, 不确定有没有发烧, 但确定的是没有接触武汉人 | 负面 |

考虑疫情的变化波动会影响公众发表舆论的情感, 例如在疫情高发期间, 不断增长的感染者数目会造成社会恐慌, 也会直接影响负面情绪舆论的数目变化, 所以将每日新增人数与每日感染总人数与负面舆论每日变化数目进行关联与可视化分析, 并将其结果绘制成图, 图 5 中左边的纵坐标为全国每天新增确诊人数, 右边的纵坐标代表每天的



负面舆论的数目,其中负面舆论数目越多,越可近似认为当天网民负面情绪积累较多;图6左边的纵坐标是全国每天确诊总人数,右边的纵坐标代表每天的负面舆论的数目。

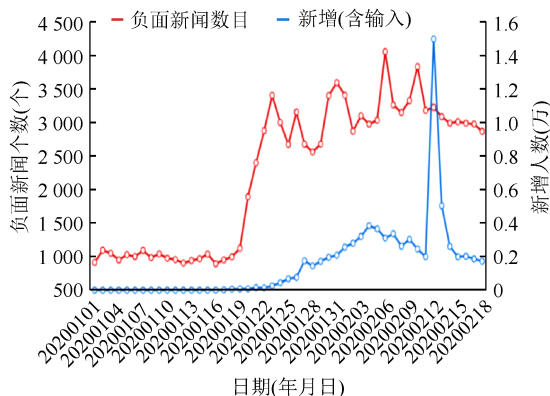


图5 负面舆论数目与每日新增患者数目变化

Fig. 5 Changes in the number of negative news and the number of new patients per day

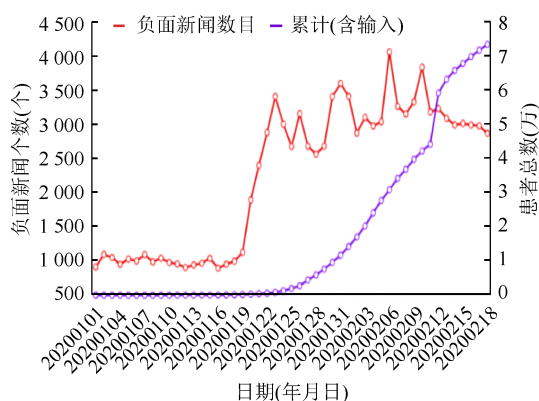


图6 负面舆论数目与每日确诊患者总数目变化

Fig. 6 Changes in the number of negative news and the total number of confirmed patients per day

从图5和图6可以看出,随着疫情感染的加剧,公众情绪也会越来越趋于负面。同时对图中几个负面舆论总数量的关键拐点进行了分析,分析结果如下:

(1) 1月1日至1月19日:全国各省份并未对每日新增感染者数量进行汇总与公示,尽管疫情已经开始持续了一段时间,但并未引起全国范围内广大人民群众的一些负面情绪。

(2) 1月20日至1月23日:从1月20日起,各省份均开始对每日的疫情信息进行统计上报,国家卫生健康委在国家卫生健康委网站每日汇总发布全国各个省份确诊病例和疑似病例的数量,每日的负

面舆论数目也随着新增确诊病人数量逐渐攀升,由此可见,在这段时间内大家开始逐步意识到这场疫情所造成的危害。这种增长情况直至1月23日武汉封城之后得到了缓解。党和政府做出的强有力的积极有效抗疫措施导致疫情得到了进一步的遏制,为此可以看出国家的有效举措就是民众的定心针。

(3) 2月3日至2月5日:疫情持续时期,发生了“武汉红十字会在捐赠物资调度、配置等方面出现的问题”等恶性事件引起负面舆论,致使短时间内公众负面情绪的积累达到了最高点。在红十字会专职副会长失职失责被免职后,舆论报道数量也出现了明显的下降,后期逐步进入舆情的平稳期。

(4) 2月12日:在2月12日,全国对新冠肺炎患者的统计口径发生了改变,《新型冠状病毒感染的肺炎诊疗方案(试行第五版)》在湖北省的病例诊断分类中,增加了“临床诊断病例”,即疑似病例具有肺炎影像学特征者。但主流媒体及时对这一修改做出了报道,在2月12日新冠肺炎患者激增的情况下,负面新闻数目在后续几天呈下降趋势。

从上述结果可以看出,在疫情高发期间,负面舆论的数目也随之增加。为了更好地缓解社会恐慌,政府及有关部分应该在疫情爆发的节点做好舆情的防控。

### 3.5 疫情期间舆情数据热点分析

探究疫情期间各类情感极性的热点信息,对各类情感极性的舆论新热点的把控有助于了解疫情的趋势与走向,本实验步骤如下:

(1) 分词。探寻舆情数据热点,将舆论数据进行分词是基础。这里使用Python所带的Jieba库对90万条预测的数据进行分词处理。

(2) 去除停用词。停用词是指在信息检索中,为节省存储空间和提高搜索效率,在处理自然语言数据(或文本)之前或之后会自动过滤掉某些字或词。这里使用哈尔滨工业大学所提供的中文停用词表对上一部分分好的词进行去除停用词处理。

(3) 计算词频。按照消极、中性、积极3个标签对文本归类,并统计高频词汇的次数。以此判断疫情期间数据的热点。

通过上述实验步骤,可以获得各类情感舆论中的高频热点词汇见表5。为了更加直观地分析舆情热点,将其绘制成词云,如图7和图8所示。



- 络的情感分析[J]. 中文信息学报, 2015, 29(5): 152-159.
- LIANG J, CHAI Y M, YUAN H B, et al. Polarity shifting and LSTM based recursive networks for sentiment analysis[J]. Journal of Chinese Information Processing, 2015, 29(5): 152-159 (in Chinese).
- [9] MNH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//2001 MIT Neural Information Processing Systems (NIPS) Conference. Cambridge: MIT Press, 2014: 2204-2212.
- [10] 史波. 公共危机事件网络舆情应对机制及策略研究[J]. 情报理论与实践, 2010, 33(7): 93-96.
- SHI B. Research on the coping mechanism and strategies for network public opinion on public crisis event[J]. Information Studies: Theory & Application, 2010, 33(7): 93-96 (in Chinese).
- [11] BURKHOLDER B T, TOOLE M J. Evolution of complex disasters[J]. The Lancet, 1995, 346(8981): 1012-1015.
- [12] CANTON L G. Emergency management[M]. Hoboken: Wiley-Interscience, 2007: 22-23.
- [13] 方付建. 突发事件网络舆情演变研究[D]. 武汉: 华中科技大学, 2011.
- FANG F J. Study on the evolution of public opinion on network of unexpected event[D]. Wuhan: Huazhong University of Science and Technology, 2011 (in Chinese).
- [14] 潘崇霞. 网络舆情演化的阶段分析[J]. 计算机与现代化, 2011(10): 203-206.
- PAN C X. Analysis of evolution phases of network public opinion[J]. Computer and Modernization, 2011(10): 203-206 (in Chinese).
- [15] 李骁, 黄征. 基于 GRU 网络的互联网信息挖掘[J]. 信息技术, 2018, 42(3): 1-5, 9.
- LI X, HUANG Z. A gated recurrent unit neural network for Web information extraction[J]. Information Technology, 2018, 42(3): 1-5, 9 (in Chinese).
- [16] 杨东, 王移芝. 基于 Attention-based C-GRU 神经网络的文本分类[J]. 计算机与现代化, 2018(2): 96-100.
- YANG D, WANG Y Z. An attention-based C-GRU neural network for text classification[J]. Computer and Modernization, 2018(2): 96-100 (in Chinese).
- [17] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2020-09-01]. [https://xueshu.baidu.com/usercenter/paper/show?paperid=147v0rh04e5c0a70qy4u0mc03q394989&site=xueshu\\_se](https://xueshu.baidu.com/usercenter/paper/show?paperid=147v0rh04e5c0a70qy4u0mc03q394989&site=xueshu_se).
- [18] 杨飘, 董文永. 基于 BERT 嵌入的中文命名实体识别方法[J]. 计算机工程, 2020, 46(4): 40-45, 52.
- YANG P, DONG W Y. Chinese named entity recognition method based on BERT embedding[J]. Computer Engineering, 2020, 46(4): 40-45, 52 (in Chinese).
- [19] 王子牛, 姜猛, 高建瓴, 等. 基于 BERT 的中文命名实体识别方法[J]. 计算机科学, 2019, 46(S2): 138-142.
- WANG Z N, JIANG M, GAO J L, et al. Chinese named entity recognition method based on BERT[J]. Computer Science, 2019, 46(S2): 138-142 (in Chinese).
- [20] 施聪莺, 徐朝军, 杨晓江. TFIDF 算法研究综述[J]. 计算机应用, 2009, 29(S1): 167-170, 180.
- SHI C Y, XU C J, YANG X J. Study of TFIDF algorithm[J]. Journal of Computer Applications, 2009, 29(S1): 167-170, 180 (in Chinese).
- [21] 任智慧, 徐浩煜, 封松林, 等. 基于 LSTM 网络的序列标注中文分词法[J]. 计算机应用研究, 2017, 34(5): 1321-1324, 1341.
- REN Z H, XU H Y, FENG S L, et al. Sequence labeling Chinese word segmentation method based on LSTM networks[J]. Application Research of Computers, 2017, 34(5): 1321-1324, 1341 (in Chinese).
- [22] 刘春磊, 武佳琪, 檀亚宁. 基于 TextCNN 的用户评论情感极性判别[J]. 电子世界, 2019(3): 48, 50.
- LIU C L, WU J Q, TAN Y N. Polarity discrimination of user comments based on TextCNN[J]. Electronics World, 2019(3): 48, 50 (in Chinese).