

新冠肺炎突发事件下的公众行为分析——基于微博数据挖掘

摘要:【目的/意义】新冠肺炎病毒的全球蔓延给各国政治经济和日常生活带来剧烈的冲击,也给中国人民造成巨大影响,基于微博数据的挖掘有助于深入了解这一突发事件期间公众情绪、关注度变化以及区域差异。【方法/过程】本文利用机器学习有监督地识别得到 2019 年 12 月 31 日至 2020 年 2 月 23 日期间来自 345 个城市 244,880 条关于新冠肺炎的微博数据,从中提取公众情感指数和关注指数,讨论了疫情发展四个阶段全国范围内公众的情感和关注程度变化,进一步对比分析不同阶段不同地区的差异,最后对公众行为的阶段差异性,以及和经济发展水平、相距重灾区距离的相关性进行了检验。【结果/结论】研究发现:公众情绪经历了持续波动、上升,稳定正面三种状态;公众对新冠肺炎疫情的关注趋势表现为低关注、关注波动、逐渐上升和逐渐下降四个阶段;公众情绪在不同阶段均存在显著差异,对疫情的关注度在部分阶段存在差异;公众行为和区域位置相关,表现为公众所在地与重灾区接近程度和情感指数呈显著负相关,所在地经济发展水平和关注指数呈显著正相关。

关键词: 新冠肺炎; 微博数据; 情感分析; 公众行为; 机器学习

中图分类号: G206

Analysis of Public Behavior under COVID-19——Based on Weibo Data Mining

Abstract: 【Purpose/significance】 The global spread of the Novel coronavirus has brought tremendous impact on the political economy and daily life of various countries, and has also caused a huge impact on the Chinese people. Understanding public sentiment, changes in attention, and regional differences during this emergency has practical significance for the government to guide public behavior, grasp the development of public sentiment, and formulate corresponding prevention and control measures in time and area. 【Method/process】 This paper used machine learning to identify 244,880 microblog posts from 345 cities from December 31, 2019 to February 23, 2020, extracted public sentiment and focus index and discussed the nation's overall emotional trend and the degree of attention to this event in the four stages of the development of the epidemic. It further analyzed the differences in public behaviors in different regions at different stages, and finally the correlation between the public behavior, the level of economic development and the distance from the severely affected area was tested. 【Result/conclusion】 The study found that public sentiment experienced three states: continuous fluctuation, rising, and stable positive; the public's attention to the trend of the Novel coronavirus pneumonia showed four stages: low attention, fluctuation of concern, gradual rise, and gradual decline; there are significant differences in public sentiment in different stages, and the degree of attention to the epidemic differs in some stages; public behavior is related to regional location, which was shown by the close negative correlation between distance from the severely affected area and the sentiment index, and the level of economic development of the location and the focus index showed a significant positive correlation.

Keywords: COVID-19; Weibo data; sentiment analysis; public behavior; machine learning

1 引言

新型冠状病毒肺炎（COVID-19）简称新冠肺炎，是一种新的呼吸道传染病，主要通过呼吸道飞沫和接触传播，被证实可以在人与人之间传播^[1]。2020年1月11日，中国报告了41例武汉地区新型冠状病毒感染的肺炎病例，1月19日其他省份发现了新冠肺炎确诊病例，为减缓感染的传播，1月23日武汉市关停公共交通系统，并关闭离汉通道^[2]。尽管如此，新冠病毒仍在全国范围内蔓延，给中国公共卫生带来了巨大的挑战和威胁，并迅速成为全世界关注的焦点。2020年1月30日，世界卫生组织（WHO）将新型冠状病毒的流行列为国际关注的突发公共卫生事件。2月，伊朗、意大利和全球其他国家开始陆续爆发疫情。随后，该

流行病变成大流行病，到 3 月底，世界一半的人口处于居家隔离状态^[3]。截至 2020 年 5 月 30 日，全球新冠肺炎病例总数超过 600 万，死亡总数超过 35 万。

在中国的抗疫过程中，民众作为参与预防和控制传染病传播的最重要主体，积极响应政府的居家隔离措施，做出了巨大贡献和牺牲。由于这类突发事件易诱发公众情绪化行为和态度，因此对新冠肺炎疫情期间民众态度变化和不同区域民众的行为差异的研究，有助于政府及时回应民众意见，稳定社会和民众情绪；对舆情动态的准确掌握，有助于政府针对不同发展阶段、不同区域制定相应的疫情防控措施，提高疫情防控能力^[4]。

然而目前关于新冠肺炎疫情期间民众的行为研究还未广泛展开，Zhao 利用 2019 年 12 月 31 日-2020 年 2 月 20 日期间收集到的 3234 条微博热搜数据对公众关注热点和情感趋势进行了描述性分析^[5]。由于热搜数据是经过提炼后反映多数用户热点关注的文本，它仅说明大量博文中出现了某一文本，因此利用这一数据和用户的真实情感存在偏差。陈兴蜀根据关键词“肺炎”抓取了 2020 年 1 月 1 日至 2020 年 2 月 29 日之间共计 6 万条高热度话题博文，并研究了全国省际新冠肺炎疫情的时空演化过程^[6]。事实上在疫情发展过程中，关于新冠肺炎讨论的关键词在发生变化，以“肺炎”为关键词引入了部分标识错误。

为降低关键词引起的标识错误，提高研究的可靠性，本研究引入机器学习分类技术剔除与新冠疫情无关的博文后，从中分析公众情绪变化和对这一事件的关注度变化，以揭示跨地区不同时期公众对新冠疫情的反应差异（情感变化和关注度变化），研究框架见图 1。本研究的时间范围为 2019 年 12 月 31 日至 2020 年 2 月 23 日，为了衡量公众行为的变化，首先通过情感分析得到情感指数以反映公众情感，定义了反映公众对新冠肺炎疫情关注程度，可用于空间、时间分析的关注指数，进一步结合新冠肺炎疫情发展的时间特征和词频统计结果，将研究范围划分为四个阶段，在此基础上讨论了研究期间全国范围内公众的整体情感和关注趋势变化以及跨区域的行为差异，最后本研究测试了三个假设：（1）新冠肺炎疫情的不同阶段，公众行为存在差异。（2）公众所在地与重灾区的接近程度和其情感值呈负相关，即离重灾区越接近的地区情绪更负面。（3）公众所在地的经济发展水平和其对事件的关注程度呈正相关，即经济发展水平越高的地区，关注程度更高。

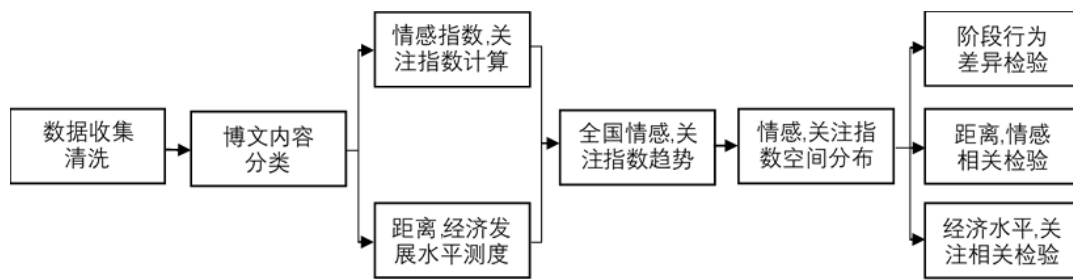


图 1 研究框架

2 研究方法和数据

2.1 微博数据收集

本研究数据来源于 github 开源项目^[7], 该项目建立了一个 2000 万活跃用户池, 根据与新冠疫情相关的 179 个关键词收集了 2019 年 12 月 1 日至 2020 年 4 月 30 日期间活跃用户共计 33,519,644 条博文信息 (用户 ID 经过脱敏处理)^[8]。本文选用了 2019 年 12 月 31 日至 2020 年 2 月 23 日期间共计 349,860 条博文, 并对博文进行必要的清洗。

2.2 博文内容分类

尽管数据集根据 179 个与新冠疫情相关的关键词收集得到, 但在研究时间范围内收集时间最早为 4 月 22 日, 这意味某些关键词在某些时间节点并未涉及新冠肺炎疫情, 并且一些地理信息如“武汉中心医院”仅表示用户的地理信息, 并未讨论有关疫情的信息。因此识别博文内容是否与新冠肺炎疫情相关是研究可靠性的基础。

现有的博文识别分类技术包括关键词匹配^[9]和机器学习分类^[10,11]。在关键词匹配中, 我们利用 jieba 分词工具获取了每日博文关键词的词频, 结合疫情发展的时间点, 发现新冠肺炎疫情的关键词在不同发展阶段存在差异。在此基础上本文将新冠肺炎疫情的发展分为武汉传播阶段 (12 月 31 日-1 月 19 日), 全国蔓延阶段 (1 月 20 日-1 月 25 日), 全国爆发阶段 (1 月 26 日-2 月 8 日) 和后疫情阶段 (2 月 9 日-2 月 23 日) 四个阶段, 并为每个阶段指定相应关键词作为判断博文是否与新冠疫情相关的依据, 表 1 显示了这些关键词。为对比两种方法的分类效果, 我们分别从新冠肺炎疫情的第二、三、四阶段中随机抽取 300 条博文, 由于第一阶段博文数量较少, 在该阶段仅抽取 100 条, 进而对 1000 条博文进行人工标记以区分是否与新冠肺炎疫情相关。

在机器学习分类中, 本研究比较了三种常用的分类器, 包括最广泛使用的支持向量机

SVC 和以随机森林 RF 为代表的 bagging 模型, XGBoost 为代表的 boosting 模型。在进行分类器判别前, Skip-Gram 词嵌入技术被用于博文文本和数字向量的转换, 最后通过 10 倍的交叉验证分别计算了准确度 (accuracy), F1 得分和召回率 (recall) 以比较不同分类器性能。

表 1 新冠肺炎疫情不同阶段的关键词

阶段 (闭区间)	关键词
12. 31-1. 19 武汉传播阶段	肺炎, 口罩, 不明, 冠状病毒, 新型冠状病毒, 新型病毒, 勤洗手, 病毒, n95, N95
1. 20-1. 25 全国蔓延阶段	武汉加油, 疫情, 出门, 在家, 疫苗, 钟南山, 感染, 疑似病例, 人传人, 人传染人
1. 26-2. 08 全国爆发阶段	封城, 特殊时期, 抗击, 防控, 捐款, 隔离, 封闭小区, 小区封闭, 健康码, 阻击战, 李文亮, 抗疫, 战疫, 潜伏期, 方舱, 小区出入证
2. 09-2. 23 后疫情阶段	COVID-19, covid-19, COVID19, covid19, 2019nCoV, cov-19, COV-19, COV19, cov-19, 复工

2.3 情感指数计算

本文通过对分类后得到的与新冠肺炎相关的博文文本进行情感分析计算情感值(情感指数)以反映公众对新冠肺炎的态度和情绪信息^[12]。情感分析方法一般可以分为基于词典的分析和基于机器学习的分析两类, 与后者相比, 基于词典的分析方法不仅提供了一条博文的情感类别, 还提供了情感的程度(情感值)^[13], 因此本研究选择这种方法作为情感分析工具。

2.4 关注指数计算

灾害评估的研究中, Lachlan 将风险感知或关注程度作为公众对某一事件的关注程度的衡量, 其定义为某个区域中与事件相关的博文的数量除以同一时期内同一区域的背景博文的总数^[14]。在行为金融学研究中, 关注度被定义为某一事件被检索的次数, 如百度指数或谷歌指数。

在本研究中, 我们将某一地区某日的关注指数定义为该地区当日与新冠肺炎相关的发文数量和前一日武汉市与新冠肺炎相关的发文数量的比值, 通过公式 1 计算得到, 相应的阶段关注指数通过平均该阶段每日关注指数得到。

$$F_i = \frac{N_{i,t}}{N_{w,(t-1)}}$$

公式 1

其中, $N_{i,t}$ 表示城市 i 在 t 时刻的发文数量, $N_{w,(t-1)}$ 表示武汉市在 t-1 时刻的发文数量。该指数可以反映不同地区的关注度水平差异, 同时也可用于比较某一地区关注度的时间变

化。相反，如果仅用环比增速表示该地区的关注度，则不能反映空间维度上（不同地区）发文数量的差异；如果仅用发文数量代表该地区的关注度，则不能反映时间维度上关注度的变化程度。高关注度指数意味着该地区相对于武汉市产生了更多与新冠肺炎疫情相关的推文，反映了公众对疫情的高重视程度。

2.5 接近程度测度

为了检验公众所在地与重灾区接近程度和情感指数的关系，我们通过计算地理距离对接近程度进行衡量，其中重灾区定义为湖北武汉市。本研究采用 Haversin 公式（公式 2）计算了两地距离。

$$Haversin(d/R) = Haversin(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)Haversin(\lambda_2 - \lambda_1) \quad \text{公式 2}$$

其中 $Haversin(\theta) = \sin^2(\theta/2)$ ， R 为地球半径，取平均值 6371km， φ_2, φ_1 表示两地维度值， λ_2, λ_1 表示两地经度值， d 表示两地距离。

2.6 经济发展水平测度

为了检验公众所在地的经济发展水平和其对事件的关注程度的相关性，我们利用 2017 年该地的 GDP 数据对经济发展水平进行定义，数据来自中国城市年鉴 1995-2018，包括全国 4 个直辖市和 334 个地级市。在该假设的研究中，GDP 数据包含城市和关注指数包含城市匹配度达到 97.2%。

3 结果和讨论

在本节我们比较了两类分类技术的性能差异，进而对文本进行分类得到 244,880 条与新冠肺炎疫情相关的博文。在此基础上首先计算了全国每日情绪指数和关注指数并对每日关键词进行统计汇总得到 15 个疫情期间的高频词，通过展现新冠疫情期间公众反应的总体趋势，以增强对这一事件的认识和了解。其次，我们绘制了研究期间全国 234 个城市不同阶段两种指数的分布地图，以显示公众对新冠疫情的态度和关注的空间格局。最后我们对引言部分提到的三个假设做了检验。

3.1 博文内容判断结果

表 2 显示了不同分类方法三种性能指标的比较结果，其中不处理表示认为博文均与新冠肺炎疫情相关（相应的召回率无意义）。可以发现，不处理的准确度仅为 0.652，这意味着直接使用博文内容将引入大量的标识错误，相较而言，关键词分类和机器学习分类的性能均有

显著提升。更具体地说，支持向量机 SVC 在准确性和 F1 得分上均表现最佳，RF 在召回率指标上表现优异，其他两个指标较差与 SVC，关键词匹配的总体性能优于 XGBoost，这表明我们提出的分阶段匹配具有一定的竞争力。尽管如此，SVC 相较于关键词匹配的方式在三个指标均有所提升，分别为 2.6%，2.3%和 5.1%。本研究选择支持向量机 SVC 对博文内容的相关性进行识别分类，最终得到 244,880 条相关博文用于公众的情感指数和关注指数的提取，进而分析新冠肺炎研究期间跨区域不同阶段的公众行为。

表 2 不同分类方式的评估结果

	Accuracy	F1	Recall
SVC	0.795	0.850	0.893
RF	0.776	0.841	0.910
XGBoost	0.761	0.826	0.869
关键词匹配	0.775	0.831	0.850
不处理	0.652	0.790	-

3.2 公众行为的时间变化趋势

图 2 展示了全国范围内公众情感指数和关注指数的整体变化趋势，分别以红色和橙色表示，其中粉色带宽表示情感指数均值的 95%置信区间。

第一阶段武汉传播期间（1 月 20 月前）由于缺乏新冠疫情的信息，对于不明原因的肺炎，公众情绪低落且波动较大，表现为情感指数均值在 0 值附近波动。另外根据图 3 的词频统计和关注度指数可以看出，这一时期公众对疫情的讨论数量较低，平均每日博文数量 38 条，最高 169 条（12 月 31 日），最低 10 条（1 月 12 日），由于样本量较少，本研究对这一阶段不进行过多分析。

在 1 月 20 日-1 月 25 日的第二阶段全国蔓延时期，公众积极讨论、分享新冠肺炎的消息，政府决定新建火神山、雷神山医院，“肺炎”、“口罩”、“武汉”、“加油”成为热度关键词。1 月 20 日钟南山院士回答记者问时肯定新冠肺炎病毒人传人后，当天情感指数达到阶段最低，表明了公众对人传人消息的负面情绪。1 月 23 日武汉关闭离汉通道后，情感指数迅速提高，当天“武汉”、“加油”关键词活跃，表明公众对武汉支持的声音提高了整体的情感值，1 月 25 日中国农历除夕当天，情感指数达到阶段最高值，并且关注指数在三个时间点分别达到峰值。

第三阶段全国爆发期包括 1 月 26 日至 2 月 8 日，作为防疫抗疫的重要阶段，中国政府采取了健康码、封闭式小区、隔离等措施，大多数中国民众在家隔离以防止疫情蔓延、病毒

感染。“希望”、“防控”、“结束”、“抗击”，“平安”，“过去”、“允悲”（表情符号，表示祈福）等表达正面情绪的词在该阶段高频出现，情感指数保持在较高水平波动，关注指数持续上升，两个指数在 2 月 2 日（被网友称为千年一遇的对称日）达到阶段峰值，并于 2 月 8 日元宵节达到整个研究阶段的峰值，情感指数和关注指数分别为 6.93 和 43.70。

2 月 9 日-2 月 23 日作为第四阶段（后疫情时期），该阶段各省陆续发布复工消息，复工工作提上政府日程，2 月 10 日“复工”作为高频词出现，反映了公众对复工的关切；尽管关于复工的讨论有较高的热度，“在家”、“隔离”等词的高频出现也反映出公众仍处于居家状态。在该阶段，总体情感指数稳定在 0 水平以上，关注指数呈下降趋势，关于新冠肺炎的讨论强度逐渐降低，两个指数同时在 2 月 14 日西方传统节日情人节出现峰值。

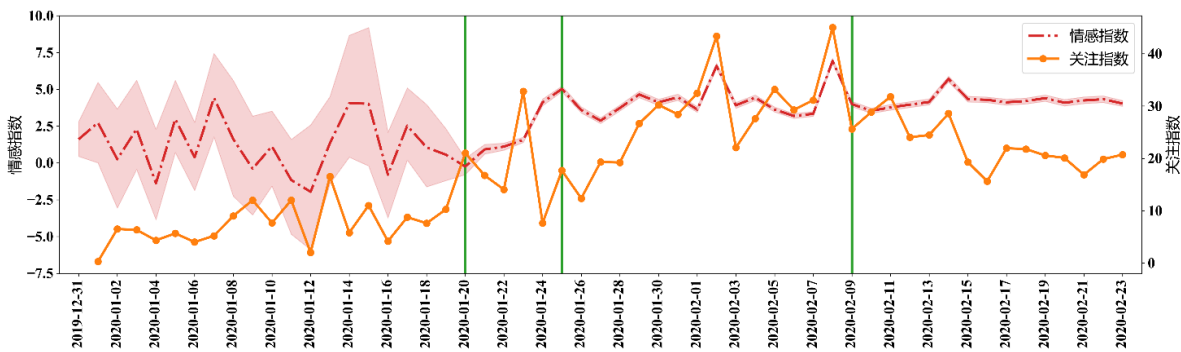


图 2 全国公众行为随时间变化的整体趋势

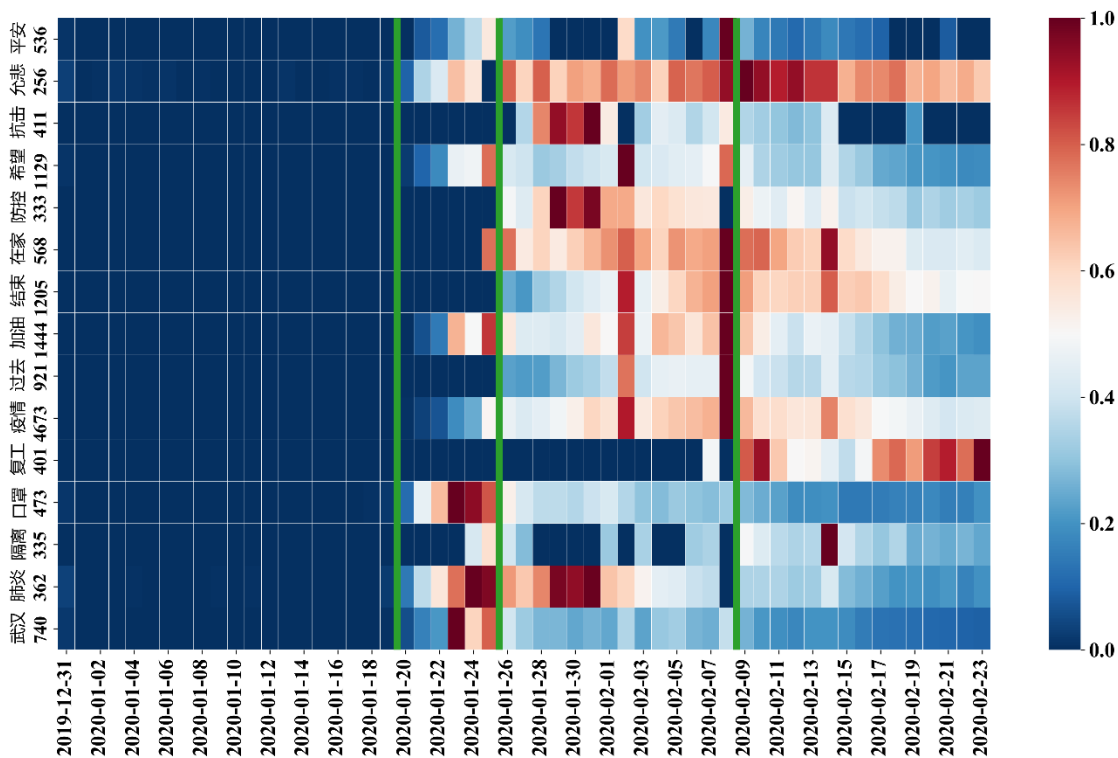


图 3 15 个关键词的词频变化趋势（左侧数值表示词频均值）

3.3 情感指数和关注指数的时空分布

本小节我们绘制了包括 4 个直辖市、2 个特别行政区、220 个地级市共计 226 个城市新冠肺炎疫情第二、三、四阶段情感指数和关注指数空间分布，以显示研究范围内公众反应的时空格局。

图 4 左侧显示了各个阶段和整体的情感指数箱线图，同时我们记录了相应阶段的均值、标准差。从图中可以看出，情感指数分布均匀，其中第二阶段全国蔓延期情感指数均值最低（3.26），远低于整体均值（4.03），且波动最大（1.15），第三阶段全国爆发期情感指数均值最高（4.37），波动较小（0.77），第四阶段后疫情时期情感指数波动最平稳（0.69），情感均值（4.22）大致介于前两者之间。为了表现情感指数的高低差异，我们根据整体情感指数的中位数（4.00）将情感指数分为两种水平。图 5 左侧显示了三个阶段的情感指数空间分布，其中蓝色和红色分别表示情感指数高、低两种水平，颜色深度表示两个极端。和前文的结论大致相同，随着时间推移，全国情感指数经历低落、高涨、稳定三种状态。其中湖北武汉市作为新冠肺炎疫情的重灾区，情感指数持续处于较低水平，并且以湖北为中心，红色增多且深度加深，这表明距离和情感指数可能存在反向关系。

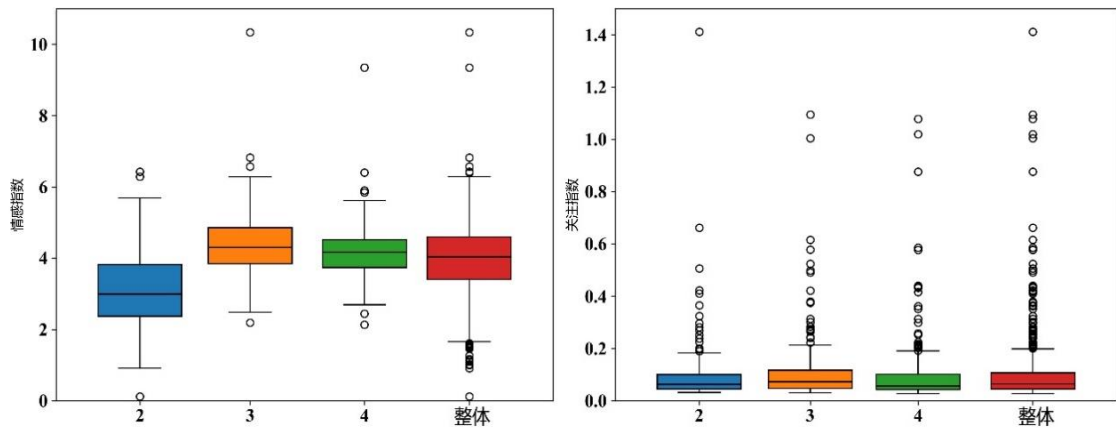


图 4 情感指数和关注指数的箱线图（左侧表示情感指数）

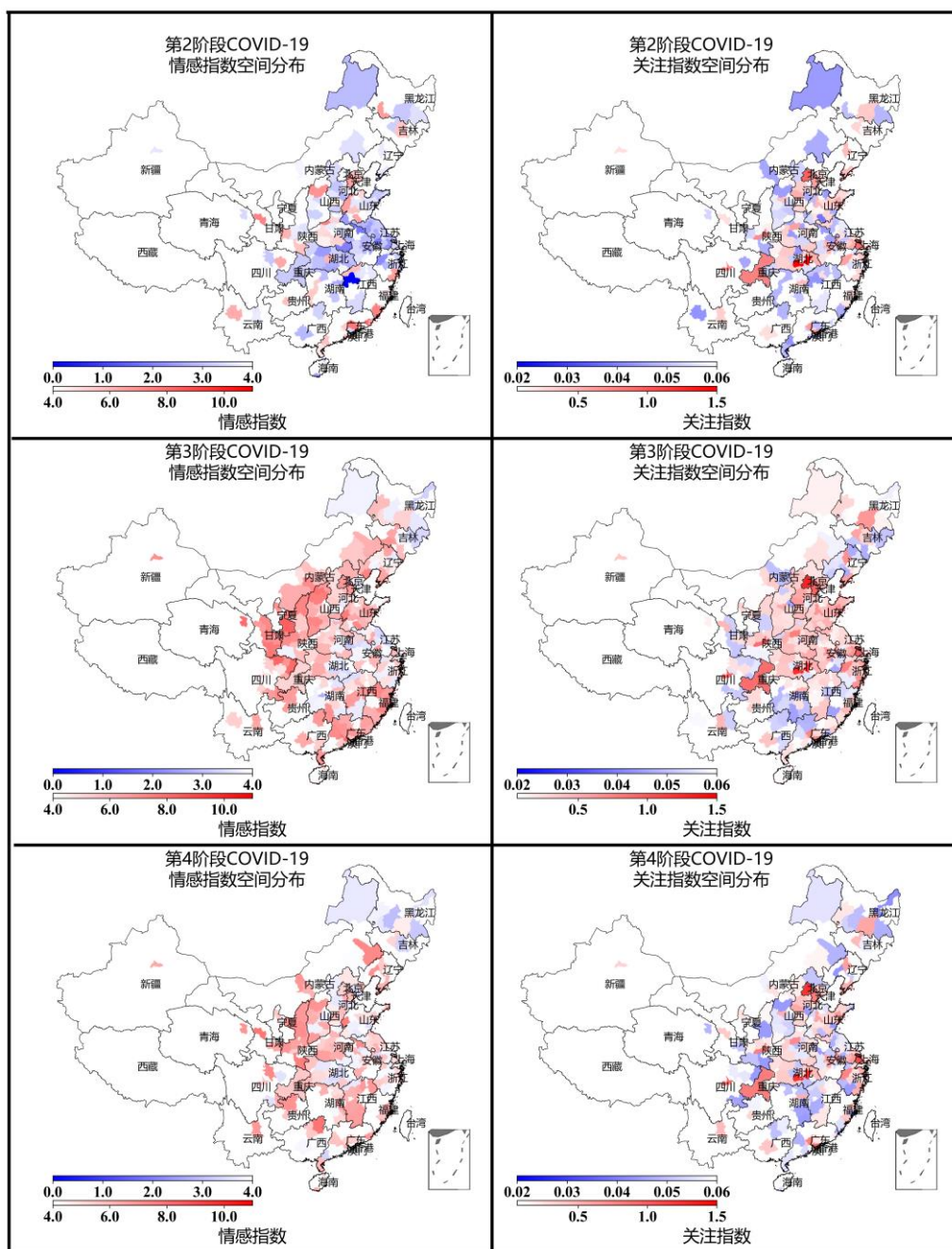


图 5 情感指数、关注指数各阶段的空间分布（左侧表示情感指数）

三个阶段中关注指数最低为 0.029，最高为 1.720，标准差 0.147，从图 4 右侧的关注指数的箱线图可以看出，关注指数在各阶段都存在较多的离群点，这意味着高关注指数和低关注指数的城市都占有一定比重。其中第二阶段全国蔓延期关注指数均值最低（0.105），第三阶段全国爆发期关注指数均值最高（0.122）且波动最小（0.137）。和情感指数的处理相同，我们根据整体关注指数的中位数将不同阶段所有城市的关注指数分为高低两种水平以展现不同城市公众关注度的层次差异。图 5 右侧显示了三个阶段关注指数的空间分布，由于第二阶段全国蔓延期新冠肺炎疫情还处于局部发展阶段，未形成全国性的大流行病，从空间分

布可以看出，该阶段对于其他两个阶段整体着色区域较少，反映了在这一阶段民众还未形成广泛的关注热度。第三阶段全国爆发期着色区域从 142 个城市扩大到 226 个，且红色区域比重最大，高水平关注城市大幅增加，新冠肺炎疫情被广泛讨论。第四阶段，整体关注热度下降，红色区域转变成蓝色区域，且区域面积减少。另外重灾区湖北武汉和北京、上海、天津、重庆四个直辖市的关注指数始终处于红色峰值区域，除此之外，长三角经济圈、珠三角经济圈、环渤海经济圈在整个疫情发展阶段的关注指数也处于高水平，这反映了该区域公众对新冠肺炎疫情的持续关注。相反，新疆、西藏、青海等西部地区由于博文数量不符合要求，在各个阶段始终没有形成有效的关注指数，而在云南、甘肃、内蒙地区，即使存在有效的关注指数，但整体水平较低，这表明经济水平和关注指数存在正向关联性的可能。

3.4 假设检验结果

在本小节，我们检验了公众行为三个阶段的差异性，以及和经济发展水平、距离重灾区距离的关系。首先将公众的情感行为和对新冠肺炎疫情的关注行为表示为相应阶段的指数，我们检验了不同阶段，公众行为是否存在差异。通过图 4 的箱线图可以发现关注指数分布不满足正态性，且 KS 检验拒绝正态性假设，因此本文采用 Mann-Whitney 秩和检验进行差异检验。表 3 显示了检验结果，可以发现公众情感指数在两两阶段的检验中均表明存在显著差异。关注指数在不相邻的二、四阶段差异性不显著，而其他两阶段的对比均表明差异显著。这反映了阶段划分一定程度上区分了公众行为的不同特征。

表 3 公众行为差异的秩和检验结果

	2, 3 阶段	2, 4 阶段	3, 4 阶段
情感指数	6803.0***	5936.0***	21986.0**
关注指数	12276.0***	11736.0	23328.0***

注：数值表示统计量，*，**和***分别为在 10%，5%和 1%的水平上显著。原假设为样本间无显著差异。

Pearson 相关性检验用于第二个假设(即离重灾区越近的地区的情感指数更低)的测试。距离和情感指数在新冠肺炎疫情发展的不同阶段的分布关系见图 6，其中右上角表示相关系数和 Pearson 线性相关显著性，虚线表示回归线。可以发现，在第二阶段全国蔓延期和第三阶段全国爆发期，距离和情感指数呈显著正相关性，表明距离和情感指数同向变动，离重灾区越近，情感指数越低，公众情绪越负面。第四阶段后疫情时期新冠肺炎疫情各方面消息在全国范围内广泛传播，疫情得到有效控制，复工复产逐渐恢复，社会秩序、公众情绪稳定，

在该阶段两者的线性关系没有表现为显著相关性。

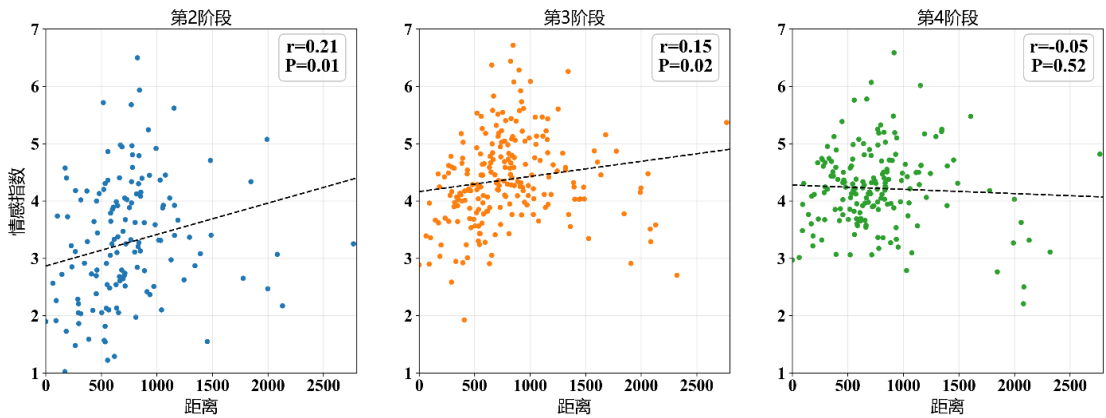


图 6 各阶段情感指数和距离的线性关系及检验

为了检验第三个假设，即经济发展水平越高的地区，关注指数更高，我们分别分析了在新冠肺炎疫情发展的不同阶段，经济发展水平和关注指数的相关性。图 7 显示了两者的线性关系，其中左上角表示相关系数和 Pearson 线性相关显著性，虚线表示回归线，三个阶段中经济发展和关注指数的相关系数分别为 0.60，0.84，0.87，且 P 值均小于 0.05，表明两者始终具有显著的相关性。这一结果说明，经济发展水平的差异对关注指数影响总是明显的，其原因在于发达地区的微博使用率较高，发文量较大，相应的关注指数相对于其他地区更高。

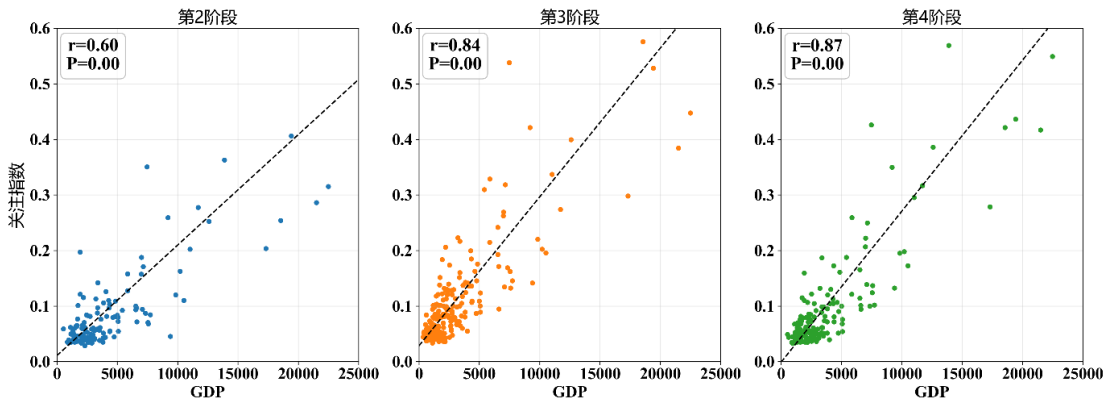


图 7 各阶段关注指数和经济发展水平的线性关系及检验

4 结论

本研究引入多种分类方式对博文是否关于新冠肺炎疫情进行判别，性能评估结果表明支持向量机 SVC 的分类性能优异，能够降低标识错误提高研究可靠性，在该模型分类基础上，

利用 2019 年 12 月 31 日至 2020 年 2 月 23 日期间 244,880 条微博文本, 本文提取了公众的情感指数和关注指数, 并将新冠肺炎疫情发展分为四个阶段分别讨论了全国范围内公众整体的情感变化趋势和对这一事件的关注程度变化, 进一步对比分析了不同地区公众不同阶段的行为差异, 最后检验了三个假设。本文的研究发现和结论包括: (1) 公众在新冠肺炎疫情第一阶段情绪值较低且波动较大, 在第二阶段情绪逐渐上升为正面情感, 并在第三、四阶段保持稳定。(2) 公众对新冠肺炎的关注趋势表现为低关注、关注波动、逐渐上升和逐渐下降四种状态。(3) 公众的情绪变化在不同阶段存在显著差异。(4) 公众在新冠肺炎疫情的第二、三阶段, 所在地与重灾区接近程度和情感指数表现为显著负相关。(5) 社会经济发展水平和公众对新冠肺炎疫情的关注程度呈显著正相关。本研究的发现可以帮助政府和卫生部门更好地把握公众态度和关注状态, 能够及时的调整安排和公众进行沟通疏导, 另外通过了解指数的时空分布为及时调整、制定区域性政策提供参考, 从而形成有针对性的措施来提高公众应对新冠肺炎疫情的能力以预防和控制疫情的传播。

本文的主要贡献在于: (1) 通过机器学习分类器对收集到的微博博文进行识别, 筛选了与这一事件相关的博文, 一定程度上避免了标识错误, 提高了研究的可靠性。(2) 设计了关注指数以反映公众对新冠肺炎疫情的关注程度, 该指标可用于不同阶段、不同地区的关注度比较。(3) 采用个人博文数据直接反映了公众对新冠肺炎态度和关注。(4) 检验了三个假设, 结果表明疫情期间经济发展水平、距重灾区距离和公众行为存在相关性, 该结果为进一步实证研究提供了参考。(5) 本文的研究成果表明新冠肺炎疫情期间公众行为空间分布的差异, 这为更多详细的区域公众行为研究提供了有用信息, 进一步拓展研究包括局部地区的行为差异研究、不同受灾程度地区和公众行为的关系、受灾损失评估等。

5 参考文献

- [1] ZHI Z L X B X Z. Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine Association [J]. 2020, 41(2)(139-44.
- [2] WILDER-SMITH A, FREEDMAN D O. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak [J]. Journal of Travel Medicine, 2020, 27(2):
- [3] TOSEPU R, GUNAWAN J, EFFENDY D S, et al. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia [J]. Science of The Total Environment, 2020, 725(
- [4] HUA G, BIN C, HONGHONG Z, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. %J Journal of medical Internet research [J]. 2014, 1(16):

- [5] ZHAO Y, CHENG S, YU X, et al. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study [J]. J Med Internet Res, 2020, 22(5): e18825.
- [6] 陈兴蜀, 常天祐, 王海舟, et al. 基于微博数据的“新冠肺炎疫情”舆情演化时空分析 [J]. 四川大学学报(自然科学版), 2020, 57(02): 409-16.
- [7] HU Y. weibo-public-opinion-dataset [M]. \url{<https://github.com/nghuyong/weibo-public-opinion-dataset>}. 2020.
- [8] HU Y, HUANG H, CHEN A, et al. Weibo-COV: A Large-Scale COVID-19 Social Media Dataset from Weibo [J]. URL:<https://arxiv.org/abs/200509174v5>, 2020,
- [9] GUAN X, CHEN C. Using social media data to understand and assess disasters [J]. Natural Hazards, 2014, 74(2): 837-50.
- [10] SEBASTIANI F. Machine learning in automated text categorization [J]. 2002, 34(1 %J ACM Comput. Surv.): 1-47.
- [11] YUAN F, LIU R. Mining Social Media Data for Rapid Damage Assessment during Hurricane Matthew: Feasibility Study [J]. Journal of Computing in Civil Engineering, 2020, 34(3):
- [12] CARAGEA C, SQUICCIARINI A, STEHLE S, et al. Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. In ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management (pp. 642-651). The Pennsylvania State University. [J]. 2014,
- [13] DANDANNAVAR P, MANGALWEDE S R. Sentiment Analysis of Real World Big Data—A Review of General Approaches [J]. IJCTA, 2017, 10(8)(185-92.
- [14] LACHLAN K A, SPENCE P R, LIN X. Expressions of risk awareness and concern through Twitter: On the utility of using the medium as an indication of audience needs [J]. Computers in Human Behavior, 2014, 35(554-9.