

# 突发公共卫生事件中公众健康信息需求主题与情感研究

钱小敏, 夏志杰\*

上海工程技术大学管理学院, 上海

收稿日期: 2022年10月28日; 录用日期: 2022年11月23日; 发布日期: 2022年11月30日

## 摘要

根据百度知网上用户针对新冠疫情提出的健康信息需求主题演变、情感波动及其相关关系, 为政府等部门针对性地发布健康信息和公众情绪疏导提供帮助。以百度知网上新冠肺炎相关问答记录文本为研究对象, 运用多维尺度分析方法可视化呈现公众健康信息需求主题, 基于时间视角分析健康信息需求演化模式及情感波动, 用相关热力图可视化呈现需求主题和公众情感间的相关关系。突发公共卫生事件中公众健康信息需求层次鲜明、范围广泛, 并且需求随时间推移波动明显但最终演化趋势相对集中; 公众情感倾向整体呈中性, 表现出先走向积极, 而后走向消极的趋势, 并且公众具体正负面情感随时间变化波动较大; 公众健康信息需求与其正负面情感间存在一定的相关关系。

## 关键词

突发公共卫生事件, 社会化问答平台, 健康信息需求, 情感分析, 多维尺度分析

# Research on the Theme and Emotion of Public Health Information Demand in Public Health Emergencies

Xiaomin Qian, Zhijie Xia\*

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Oct. 28<sup>th</sup>, 2022; accepted: Nov. 23<sup>rd</sup>, 2022; published: Nov. 30<sup>th</sup>, 2022

## Abstract

According to Baidu knows users' health information needs for COVID-19, the theme evolution, \*通讯作者。

文章引用: 钱小敏, 夏志杰. 突发公共卫生事件中公众健康信息需求主题与情感研究[J]. 应用数学进展, 2022, 11(11): 8377-8392. DOI: 10.12677/aam.2022.1111886

emotional fluctuations and their related relationships provide help for the government and other departments to release a health information and public emotional diversion in targeted. Taking Baidu know COVID-19 related Q & A; a recorded text as the research object, the multi-dimensional scale analysis method is used to visualize the theme of public health information needs, analyze the evolution mode of health information needs and emotional fluctuations from the perspective of time, and visualize the correlation between the theme of needs and public emotions with the relevant heat map. In public health emergencies, the demand for public health information is clear-cut and wide-ranging, and the demand fluctuates significantly over time, but the final evolution trend is relatively concentrated; the public emotional tendency is neutral as a whole, showing a tendency to move first to positive and then to negative, and the specific positive and negative emotions of the public fluctuate greatly with time; there is a certain correlation between public health information needs and their positive and negative emotions.

## Keywords

Public Health Emergencies, Social Q & A Platforms, Health Information Needs, Emotional Analysis, Multi-Dimensional Scale Analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着新冠疫情趋于常态化, 公众越来越倾向于在社会化问答平台上咨询或分享新冠疫情相关信息。社会化问答平台既能为公众提供信息支持, 又能提供情感支持[1]。对社会化问答平台上的健康信息进行需求分析, 能帮助政府和媒体机构精准投放健康信息以应对公众健康需求。此外, 由新冠疫情带来的舆情所引发的公众恐慌的次生危机, 严重阻碍了社会长足稳定地发展[2]。因此, 关注疫情中公众情绪变化, 有利于把握公众情感总体趋势, 及时调整疫情舆论整体走向。综上, 对社会化问答平台上的健康信息进行需求主题与情感研究是十分必要的。

目前, 已有研究主要通过人工编码和机器学习算法识别突发公共卫生事件中公众健康需求主题, 但较少有人对需求主题进行可视化分析; 也有部分学者从时间视角分析公众健康信息需求演变模式, 但大多数研究对突发公共卫生事件的时间划分依据不一; 此外, 针对突发公共卫生事件中公众情感分析也是一大热点, 但目前此类研究多为单一的情绪分析, 缺乏与主题结合的相关性分析。

鉴于此, 本文以特定时间段内社会化问答平台上与新冠疫情有关的问答记录为研究对象, 先根据突发公共卫生事件生命周期理论划分新冠疫情各个阶段, 再利用多维尺度分析方法可视化呈现公众健康信息需求主题分布和演变规律, 然后通过自定义情感词典的方法分析公众情感变化, 最后运用相关热力图可视化呈现需求主题与情感之间的相关性。本文的研究结果能增多需求主题关联词以使用户更方便快捷搜索出与突发公共卫生事件相关的信息, 并且能帮助相关部门有针对性地制定准确的政策以缓解公众面对新冠疫情的恐慌与焦虑。

## 2. 相关研究述评

### 2.1. 突发事件生命周期划分

基于时间视角分析用户健康信息需求演化规律是用户健康信息需求主题研究的一大趋势。已有研

研究表明用户健康信息需求随时间发生变化, 如石静等以一年为周期分析“慢友帮”和“Daily Strength”上 2013~2017 年的糖尿病群组用户信息需求的演变趋势[3]。此外, 不少学者针对突发事件生命周期划分依据分别提出了相应的生命周期理论。Fink S 提出了危机传播四阶段模型, 即潜伏期、暴发期、蔓延期和痊愈期[4]; Burkholder 等提出了紧急事件的三阶段模型, 即急性紧急事件阶段、晚期紧急事件阶段和后紧急事件阶段[5]; Robert H 提出了危机管理 5R 模式, 即缩减、预备、反应、恢复和韧性阶段[6]。基于以上生命周期理论, 董晨等将上海疫情发展周期划分为三个阶段, 即暴发、消退和常态化发展阶段[7]。

## 2.2. 用户健康信息需求主题

关于用户健康信息需求主题的研究一直是用户健康信息行为研究的主流方向。现有研究的数据采集方法有问卷法、访谈法、日志法和数据抓取等, 数据分析方法主要包括人工主题编码和利用机器学习算法进行自动化识别等。如周艳玫等利用问卷调查大学生信息获取能力, 得出大学生信息服务需求[8]; 陈静等通过移动经验取样法、关键事件技术和日志法采集大学生的健康信息搜索经历, 对大学生自述问题文本进行人工编码以识别健康信息需求主题[9]; 安璐等分别在微博、搜狗上爬取中东呼吸综合征(MERS)相关微博和微信公众号文章, 利用改进的潜在狄利克雷分配(LDA)模型对话题文本进行主题提取, 得到利益相关者在社交媒体中的话题关注点[10]。综上, 问卷法和访谈法主观性较强且收集到的数据量有限, 而网络爬虫能快速搜集到海量客观数据; 人工编码和机器学习算法可以较好识别需求主题, 但识别结果无法可视化呈现。

多维尺度分析(MDS)作为信息可视化技术的典型代表, 可将高维空间中的信息投射到二维或三维空间, 从而根据空间聚类情况, 体现被调查对象间存在的联系或关系[11]。早期学者多将 MDS 用于文献共词分析[12]以发现某一学科领域的研究热点。近年来已有部分学者将其应用于信息需求主题识别研究中, 如 Cai H 等运用 MDS 对用户雅虎问答平台上使用的糖尿病术语进行聚类分析[13]; Qiang X L 等利用 MDS 对新型冠状病毒等人类冠状病毒的模式进行可视化聚类[14]。

## 2.3. 用户健康信息需求情感

突发事件往往会引发公众的恐慌和焦虑等情绪, 据此有研究人员针对突发事件下公众表达的情绪开展了相关研究。Zhang J 等运用 Lexalytics 对雅虎问答平台上用户创建的问答文本进行情感评分, 结果显示四个不同时间段的情绪得分均为负, 但第一阶段和第二阶段之间存在非常明显的情绪差异[15]; Zhao Y 等通过 ROST CM6.0 情感分析工具分析不同阶段公众在新冠肺炎疫情相关微博话题中表达的情感占比, 发现公众对新冠肺炎疫情相关话题的负面情绪不断减弱, 而正面情绪逐渐增强[16]。上述情绪分析研究主要为计算各阶段问答文本中用户情感得分, 以此得出用户总体情感变化趋势。然而, 有研究表明用户健康需求与其表达的情绪具有相关性, 如刘冰等对妈妈网论坛上处于身份转换期的女性的发帖进行需求主题与情感的相关性分析, 最终发现某一主题类型与某一情感类型间存在强相关关系[17]。因此, 目前有关突发公共卫生事件下用户情绪的研究亟需针对需求主题与情感间的相关分析进行探究。

综上, 目前突发公共卫生事件健康信息需求相关研究鲜有利用可视化数据分析方法将用户需求主题直观地呈现, 也很少有以生命周期理论为依据划分突发事件下用户健康信息阶段的研究, 并且很少考虑将情感分析与主题研究相结合开展研究。因此, 本文以典型中文社会化问答平台——百度知道上有关新冠肺炎疫情的问答文本为研究对象, 运用 MDS 分析可视化呈现用户健康信息需求主题, 基于生命周期理论划分时间阶段以探究用户健康需求主题与情感的演变规律, 最后结合情感与主题的相关性分析探索用户的情感驱动趋势。

### 3. 研究设计

#### 3.1. 数据来源与采集

百度知道作为全球领先中文互动问答平台, 截至 2021 年 4 月, 已累计解决用户提问超 5.9 亿, 每日活跃用户超 1.2 亿[18]。本文利用八爪鱼采集器对百度知道平台上以“新型冠状病毒”“新冠”和“COVID-19”为关键词检索到的相关问答文本进行爬取。其中, 爬取日期为百度知道上第一次出现相关问题的时间(2020 年 1 月 20 日), 到爬取日当天(2021 年 7 月 16 日), 爬取内容为相关问答文本与发布时间。本次共爬取 1202 条原始问答记录, 经过去除重复项与无关项等系列处理后, 最终保留 1143 条问答记录。

#### 3.2. 生命周期划分

突发事件的网络信息传播随着突发事件的演变具有一定的生命周期[7]。本文基于美国危机管理专家 Fink S 提出的突发事件生命周期四阶段划分法[4], 通过识别新冠疫情期间百度知道平台上提问量随时间变化的拐点以及重大事件节点, 将用户信息生命周期划分为三个阶段, 即暴发期(2020 年 1 月 20 日~2020 年 6 月 30 日)、平稳期(2020 年 7 月 1 日~2020 年 12 月 31 日)和衰退期(2021 年 1 月 1 日~2021 年 7 月 16 日), 具体如图 1 所示。2020 年 7 月 30 日是用户提问量从急剧下降向趋于平缓过渡的拐点, 并且当日全球新冠确诊病例超 1000 万例, 故将其作为划分暴发期与平稳期的时间点。而 2020 年 12 月 31 日是提问量从平缓向逐步衰退过渡的拐点, 也是中国新冠病毒疫苗上市的日子, 故将其作为划分平稳期与衰退期的时间点。

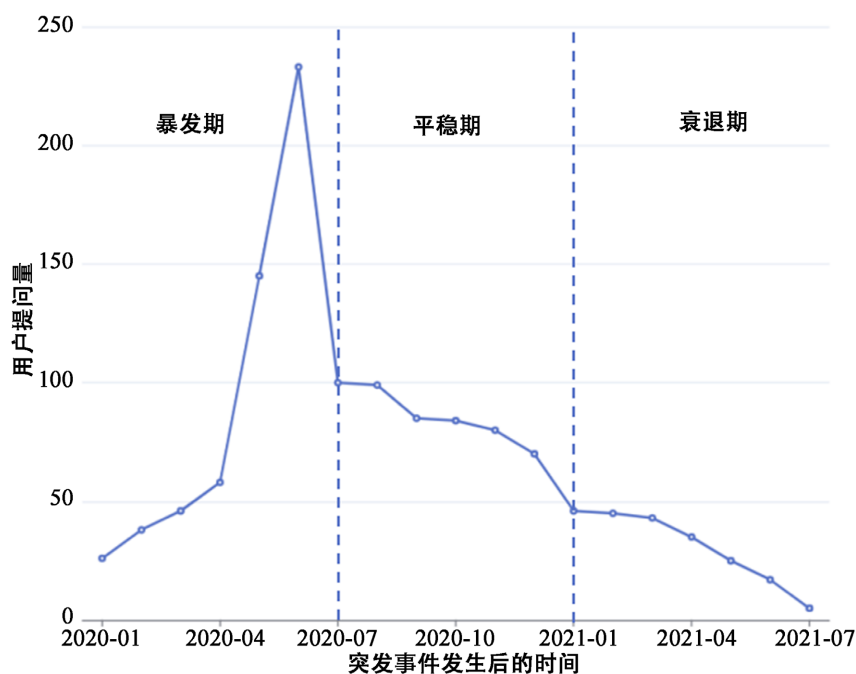


Figure 1. Life cycle division of COVID-19

图 1. 新冠疫情的生命周期划分

#### 3.3. 主题识别阶段

##### 3.3.1. 编码方案

本研究利用内容分析法对用户社会化问答平台上的问答文本进行主题编码[19]。主题编码的过程包括

以下六个步骤: 收集并熟悉数据、生成初始编码、识别编码的主题、回顾主题、定义并命名主题、形成最终结果[20]。考虑到 COVID-19 问答的独特性, 本文进一步简化 Zhang J 等得出的关于寨卡病毒的公众健康信息四大需求主题, 即关于寨卡病毒、自我表达、相关事件和其他[15], 并利用 A. Strauss 开放式编码[21]对新冠肺炎用户健康信息需求进行主题识别, 最终形成包括新冠病毒的相关信息、新冠病毒的社会影响、公众对新冠病毒的回应这三大主题, 具体主题分类结果如表 1 所示。整个编码过程由两名研究人员独立完成, 经过对比与讨论后, 对存在分歧的地方向该领域专家咨询意见后确定。Cohen's kappa 值可用于衡量数据集编码结果的内部一致性, 本研究 Cohen's kappa 值为 0.769, 表明本次编码结果是比较可靠的[22]。

**Table 1.** Category coding table of public health information needs under public health emergencies

**表 1.** 突发公共卫生事件下公众健康信息需求类别编码表

类别	内涵	关键词
新冠病毒的相关信息 (A)	与病毒有关的医学信息、病毒的动态等	病因与症状、诊断、传播、传染、治疗、疫情等
新冠病毒的社会影响 (B)	政府对病毒的防控、病毒对衣食住行的影响等	防控、防护、监控、出行、快递、工作、学习等
公众对新冠病毒的回应 (C)	公众对病毒持有的态度与观点、对病毒的预防措施等	态度、观点、预防、措施等

### 3.3.2. 多维尺度分析

多维尺度分析(MDS 分析)能够将高维空间中的关键词投射到低维空间(二维或三维空间), 并且所有关键词间的原语义关系仍能以可视化的形式被如实表示[23]。本文利用 GooSeeker 软件分别对各阶段、各类别的问答文本进行分词、去停用词、合并同义词等文本预处理后, 根据高频关键词表构建关键词共现矩阵, 再利用 Excel 中的 Ochiai 系数将共词矩阵转换为相似矩阵, 进一步得到相异矩阵, 最后将相异矩阵导入 SPSS 软件, 用 INDSCAL 模型进行 MDS 分析[24]。经过 MDS 分析后, 语义相关的关键词在低维空间呈现聚合状态, 研究人员可根据实际需要不同颜色标记各子类别分布。MDS 分析的效果通常用应力值(Stress)和决定系数( $R^2$ )这两指标来衡量, 一般而言, Stress 值越低或  $R^2$  值越高, 则 MDS 分析结果越好。

### 3.4. 情感分析阶段

新冠病毒的暴发引发了全世界公众的恐慌, 情感分析能充分揭示人们对疫情所表达的情绪[15]。因此, 本文利用 GooSeeker 软件对各阶段用户问答记录文本进行情绪分析, 先进行分词、去停用词、同义词合并等操作, 再通过自定义情感词典对照并提取情感词以计算其情感得分(正面得 1 分, 负面得-1 分, 中性得 0 分), 最后根据汇总后的情感得分正负得出每个问答文本的情感倾向。此外, 本文借助徐琳宏构建的包括乐、好、怒、哀、惧、恶、惊七大分类的中文情感词本体库[25], 通过抽取问答文本中的正面情感(赞扬、感谢、期望、感动、安心)和负面情感(恐惧、憎恶、怀疑、哀伤、愤怒)进行分析[26]。

有研究表明问答与情感之间具有相关关系[21], 故本文利用 Pearson 相关系数, 通过计算正负面问答记录的主题占比及相应情感得分, 确定上述文本的主题和情感值分布, 并采用 HeatMap 的形式呈现问答文本中主题与情感间的相关关系[17]。

## 4. 结果分析

### 4.1. 社会化问答平台用户健康需求主题分析

在突发事件生命周期划分和初步主题编码的基础上, 本文利用 SPSS 软件对各阶段、各主类别的问答记



录文本依次进行 MDS 分析以识别用户健康需求子类别, 并对所属不同子类别的关键词用不同颜色的圆圈标记加以区分。本次 MDS 分析结果的评价指标 Stress 值均小于 0.2,  $R^2$  值均大于 0.9, 表明结果是合理的。

暴发期(2020.1.20~2020.6.30)各主类别的可视化分析结果如图 2(a)-(c)所示, 各子类别所包含的具体关键词、关键词个数、占比和结果评价指标如表 2 所示。暴发期新冠病毒的相关信息(A)中的子类别症状与诊断(A1) (40%)和传播与传染(A3)占比(31.4%)分别位居第一、二位, 这也许是因为暴发期病毒的不确定性, 公众更易产生与病毒自身属性相关的初始需求。暴发期新冠病毒的社会影响(B)中的子类别防控(B1)的占比(59.1%)略高于疫情影响对象(B2) (40.9%), 这可能是由于疫情肆虐影响人们的日常生活, 公众希望政府等相关部门出台防控政策以缓解生活压力。暴发期公众对新冠病毒的回应(C)中的子类别预防措施(C2)的占比(74.5%)明显高于态度观点(C1) (25.5%), 这可能是由于病毒在暴发期肆意猖獗, 人们害怕感染病毒从而更关心自身防疫措施。

平稳期(2020.7.1~2020.12.31)各主类别 A、B、C 的可视化分析结果如图 3(a)-(c)所示, 各子类别所包含的具体关键词、关键词个数、占比和结果评价指标如表 2 所示。平稳期新冠病毒的相关信息(A)中的子类别症状与诊断(A1) (34.1%)和传播与传染(A3)占比(26.8%)仍然分别位居第一、二位, 并且以上两子类别关键词还出现外国国家名, 说明随着国内疫情逐步平稳, 人们的注意点开始转移到疫情还未得以控制的一些外国国家。同时, 症状与诊断(A1)关键词还涉及到后遗症, 这表明随着疫情进一步推移, 人们更加深入地思考新冠肺炎对人体的危害。子类别药物研发(A5)出现了“变异”“上市”“接种”等关键词, 原因可能是随着新冠病毒的变异, 人们希望接种疫苗以免招病毒侵害的需求愈发强烈。平稳期新冠病毒的社会影响(B)中的子类别疫情影响对象(B2)的占比(55.3%)略高于防控(B1) (44.7%), 这表明虽然本次疫情在政府部门干预下得到有效控制, 但人们不得不开始担心疫情给整个社会带来的前所未有的巨大影响。平稳期公众对新冠病毒的回应(C)中的子类别态度观点(C1)的占比(84.8%)显著高于预防措施(C2) (15.2%), 并且上述两类别关键词中出现外国国家名, 这可能是因为国内疫情得以控制, 人们对相关病毒预防措施早已了然于心, 而国外疫情仍然凶猛, 公众纷纷针对国外疫情展开议论。此外, “态度观点”关键词中还新增“谣言”一词, 这意味着随着公众对疫情的深入了解, 公众不再盲目相信某些与新冠肺炎有关的言论, 而开始以批判性的眼光审视此类缺乏依据的信息。

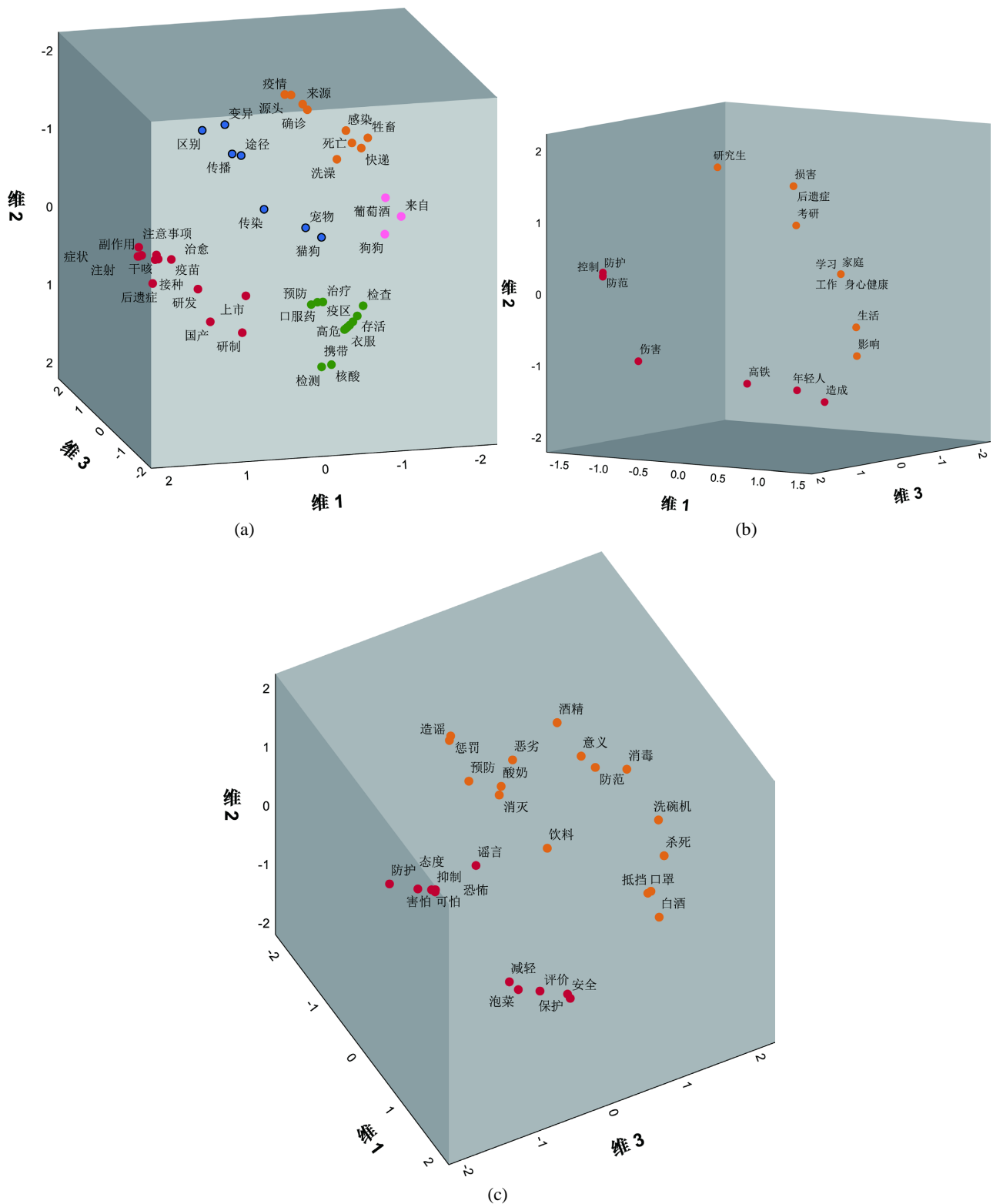
衰退期(2021.1.1~2021.7.16)各主类别 A、B、C 的可视化分析结果如图 4(a)-(c)所示, 各子类别所包含的具体关键词、关键词个数、占比和结果评价指标如表 2 所示。衰退期新冠病毒的相关信息(A)中的子类别传播与传染(A3) (36.4%)占比位居首位, 合理解释是国内整体疫情逐渐衰退, 而局部地区疫情反弹导致人们始终密切关注病毒的传播与传染。药物研发(A5)占比(27.3%)居于次位, 并且关键词中还新增“副作用”“注意事项”等词, 原因可能是公众响应国家号召开始接种疫苗, 理所应当关心疫苗接种后的注意事项, 而新冠疫苗刚上市不久, 因此人们会担心疫苗是否有副作用。衰退期新冠病毒的社会影响(B)中的子类别疫情影响对象(B2)的占比(77.8%)明显高于防控(B1) (22.2%), 这说明尽管国内疫情明显好转, 但本次疫情给各行各业带来的影响仍然不容小觑。衰退期公众对新冠病毒的回应(C)中的子类别预防措施(C2)的占比(53.6%)稍高于态度观点(C1) (46.4%), 这可能是国内疫情局部波动使得公众又开始重视自身防疫并积极寻求有效的预防措施。此外, “态度观点”中有出现“造谣”“谣言”关键词, 这说明社会上流传的某些新冠肺炎言论缺乏一定的真实性, 如接种新冠疫苗致人死亡等。

综上, 本文归纳出突发公共卫生事件下公众健康信息需求主题架构, 具体如图 5 所示。主题架构由上下两层类别构成, 上层包括新冠疫情的相关信息、新冠病毒的社会影响、公众对新冠病毒的回应这三个主类别, 下层包括症状与诊断、治疗、传播与传染、疫情动态、药物研发; 防控、疫情影响对象; 态度观点、预防措施这九个子类别。









**Figure 4.** (a) MDS analysis of related information of COVID-19 in the declining period; (b) MDS analysis of the social impact of COVID-19 in the declining period; (c) MDS analysis of the public response to COVID-19 in the decline period  
**图 4.** (a)衰退期新冠病毒的相关信息 MDS 分析; (b)衰退期新冠病毒的社会影响 MDS 分析; (c)衰退期公众对新冠病毒的回应 MDS 分析

**Table 2.** MDS analysis results of the three stages  
**表 2.** 三阶段的 MDS 分析结果

阶段	主类别	子类别	关键词	关键词 个数	关键词 占比	Stress	R <sup>2</sup>
暴发期	新冠病毒 的相关信 息(A)	症状与诊断(A1)	咳嗽、发烧、症状、检查、鼻涕、通气、头疼、报告、 发热、头晕、乏力、潜伏期、检测、高烧	14	40%	0.151	0.929
		治疗手段(A2)	治愈、特效药、治疗	3	8.6%		
		传播与传染(A3)	感染、接触、传播、传染、途径、快递、传染源、纸币、 蚊子、溶胶、携带	11	31.4%		
		疫情动态(A4)	疫情、病例	2	5.7%		
		药物研发(A5)	疫苗、临床、投入使用、实验、研发	5	14.3%		
合计				35	100%		
新冠病毒 的社会影 响(B)		疫情防控(B1)	防控、卫生、管理、政府、行政部门、场所、监督、公 共场所、办公、人民政府、工业、预防、防护、措施、 消毒、商场、隔离、超市、紧急措施、交通运输、防疫、 出租车、病人、医院、防治、公交	26	59.1%	0.162	0.907
		疫情影响对象(B2)	影响、开工、延期、工厂、企业、公司、放假、快递、 复工、上班、封路、发货、市场、汽车行业、农村、外 贸出口、口罩、停工	18	40.9%		
合计				44	100%		
公众对新 冠病毒的 回应(C)		态度观点(C1)	希望、害怕、压力、保护、担忧、担心、可怕、恐慌、 威胁、焦虑、恐怖、捐款	12	25.5%	0.150	0.946
		预防措施(C2)	预防、口罩、消毒、消灭、方法、方式、净化器、防御、 隔离、防护、居家、白酒、效果、药物、酒精、防控、 措施、大蒜、喷洒、公众、喝酒、防治、金银花、抗生 素、紫外线、板蓝根、防疫、双黄连、通风、防范、疫 苗、杀菌、防毒面具、消毒剂	35	74.5%		
合计				47	100%		
平稳期	新冠病毒 的相关信 息(A)	症状与诊断(A1)	检测、确诊、阳性、判断、症状、日本、变异、法国、 发烧、方法、阴性、感冒、后遗症、核酸	14	34.1%	0.149	0.936
		治疗手段(A2)	美国、治疗、治愈、患者、疗法	5	12.2%		
		传播与传染(A3)	感染、传播、通过、蚊子、苍蝇、英国、途径、携带、 来源、传染、动物	11	26.8%		
		疫情动态(A4)	疫情、病例、韩国	3	7.3%		
		药物研发(A5)	疫苗、变异、上市、研究、中国、注射、研制、接种	8	19.5%		
合计		41	100%				
新冠病毒 的社会影 响(B)		疫情防控(B1)	防控、采取、外出、措施、公共关系、防范、活动、政 府、办公楼、防疫、管控、举措、防护、封城、幼儿园、 部门、飞机	17	44.7%	0.138	0.957
		疫情影响对象(B2)	影响、地产、危害、行业、金融、工作、生活、企业、 心理、开工、上班、感染、中学生、采购、工伤、大学 生、学习、封锁、楼市、开学、酒店	21	55.3%		
合计		38	100%				

Continued

公众对新冠病毒的回应(C)	态度观点(C1)	疫情、确诊、感染、印度、巴西、结束、惶恐、严重、灾祸、小学生、启示、解决、抵抗、恐怖、重要、希望、疑问、减缓、谣言、消失、评价、怀疑、危险、丢人、想法、可怕、厉害、西班牙	28	84.8%	0.156	0.948
	预防措施(C2)	美国、预防、口罩、举措、措施	5	15.2%		
合计	33	100%				
衰退期新冠病毒的相关信息(A)	症状与诊断(A1)	症状、确诊、变异、核酸、干咳、检测、检查、判断	8	18.2%	0.174	0.923
	治疗手段(A2)	预防、治疗、后遗症、治愈	4	9.1%		
	传播与传染(A3)	感染、源头、传染、来源、传播、葡萄酒、狗狗、来自、牲畜、途径、快递、宠物、猫狗、携带、衣服、存活	16	36.4%		
	疫情动态(A4)	疫情、死亡、疫区、高危	4	9.1%		
	药物研发(A5)	疫苗、上市、注射、研发、口服液、接种、副作用、注意事项、国产、区别、洗澡、研制	12	27.3%		
合计	44	100%				
新冠病毒的社会影响(B)	疫情防控(B1)	高铁、防范、控制、防护	4	22.2%	0.077	0.977
	疫情影响对象(B2)	影响、生活、工作、家庭、学习、身心健康、人际交往、造成、损害、后遗症、年轻人、考研、研究生、伤害	14	77.8%		
合计	18	100%				
公众对新冠病毒的回应(C)	态度观点(C1)	安全、保护、惩罚、意义、造谣、谣言、评价、恶劣、态度、可怕、恐怖、害怕、严重	13	46.4%	0.168	0.897
	预防措施(C2)	杀死、消毒、预防、洗碗机、酒精、白酒、酸奶、口罩、消灭、减轻、抵挡、饮料、防范、防护、抑制	15	53.6%		
合计	28	100%				

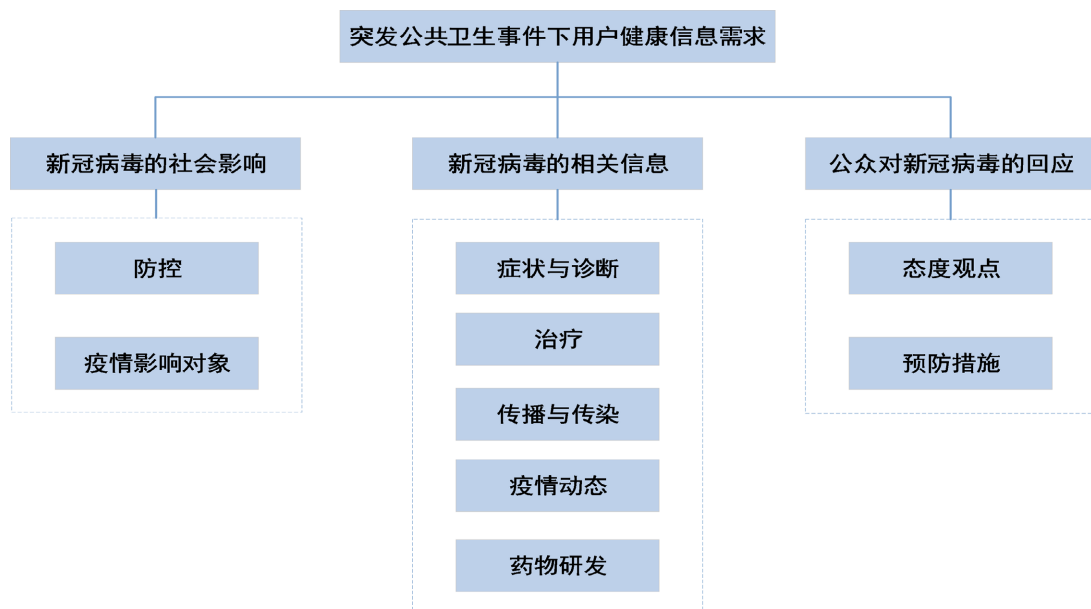


Figure 5. User health requirements topic distribution

图 5. 用户健康需求主题分布

为进一步验证不同时期公众健康信息需求是否存在显著差异, 本文利用卡方检验对各阶段用户健康需求主题分布进行显著性检验。结果显示  $\chi^2 = 41.573$ ,  $df = 4$ ,  $p < 0.001$ , 说明不同时期公众健康信息需求存在显著差异。Cramer's V 系数为 0.135, 说明公众健康需求与时间之间存在弱强度相关关系。

## 4.2. 社会化问答平台用户情感分析

本文基于自定义情感词典对问答记录文本进行情感分析, 不同时期情感倾向占比分布如图 6 所示。

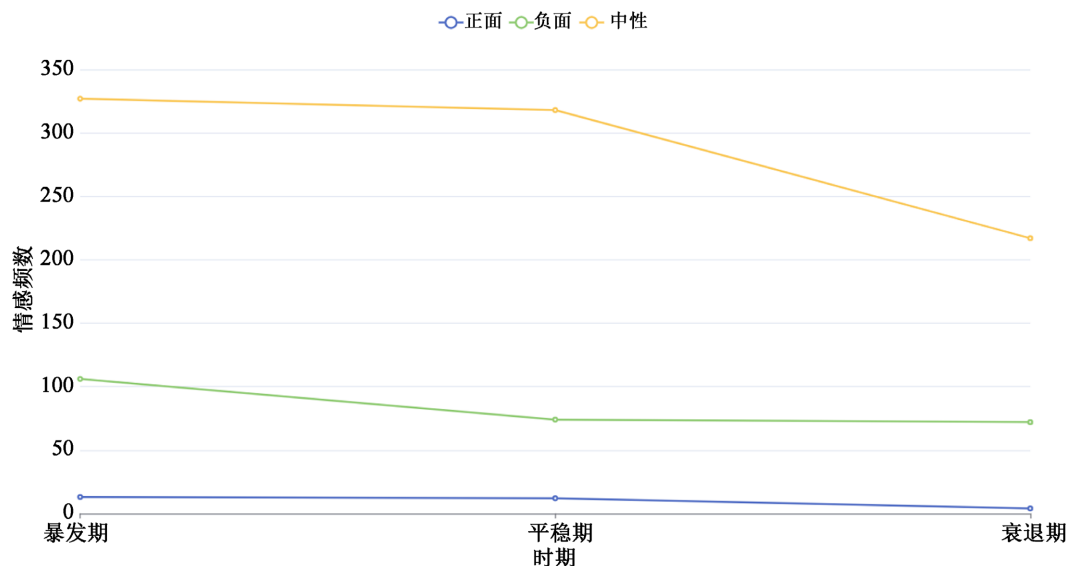


Figure 6. Distribution of the percentage of public emotional tendency in different periods

图 6. 不同时期公众情感倾向占比分布

由图 6 可知, 从横向看, 三个时期的用户情感倾向一致, 具体表现为以中性情绪为主, 负面情绪次之, 正面情绪最少。可能是因为公众在平台上就新冠病毒的相关信息作出的问答具有客观性, 并且在面对新冠疫情时普遍情绪不高涨, 因此他们所表达出的情绪往往呈现中性。从纵向看, 平稳期负面情绪比重相较于暴发期呈现下降趋势, 正面情绪占比相对持平, 而中性情绪比重加大, 这表明整体疫情好转导致公众总体情感走向积极; 衰退期负面情感占比相较于平稳期有所回升, 正面情感比重有所下降, 而中性情感占比则略微下降, 这可能是局部疫情反复导致公众情感整体上走向消极。

此外, 本文根据徐琳宏创建的中文情感词本体库, 抽取相关问答文本进行正面(赞扬、感谢、期望、感动、安心)和负面(恐惧、憎恶、怀疑、哀伤、愤怒)情感分析。不同时期具体情感倾向占比分布如图 7 所示。

由图 7 可知, 从横向看, 负面情感以恐惧和哀伤居多, 新冠病毒具有较强的传染性, 并且此病毒目前还不能被彻底消灭, 公众可能因害怕被病毒感染而表现出惴惴不安的状态以及对自身的无能为力而感到悲伤; 而正面情感以赞扬和期望为主, 可能原因是国内新冠疫情的有效控制得力于中国政府的积极防控, 公众理所当然的对政府所作所为表示赞扬, 同时, 随着疫情逐步深入, 公众更期望科研人员成功研制疫苗, 并尽早将疫苗上市以及投入使用以增强自身防疫能力。从纵向看, 负面情感中恐惧、怀疑、哀伤的比重随时间变化有所下降, 而憎恶、愤怒的比重有所上升, 这也许是因为公众在深入了解新冠病毒相关信息后渐渐消除了心中的恐惧, 也因疫情得到有力控制而缓解了内心的哀伤, 同时因疫苗的大量投入使用而减少了对疫苗副作用的怀疑态度, 但局部疫情频发又引起了公众内心对新冠病毒深深的痛恨; 而正面情感占比变化幅度较大, 其中赞扬占比逐步减少可能是暴发期政府及时干预使疫情得到稳定控制

所致, 感谢只出现在暴发期可能是因为疫情暴发期公众在问答平台上紧急提问并对应答者作出答谢, 期望只存在于暴发期和平稳期说明公众注射疫苗的心愿已落实到位, 感动和安心只出现在平稳期也许是因为疫情趋于稳定, 公众对为疫情贡献力量的行为表示感动, 同时也对日渐平息的疫情状况表示很安心。

根据正负情感对应的问答记录文本中主题分布及相应情感得分, 本文通过计算主题与情感间的 Pearson 相关系数来确定主题与情感的相关关系, 并利用颜色表示相关性强弱, 颜色越趋近红色代表相关性越强, 越趋近蓝色代表相关性越弱。具体结果如图 8 所示。

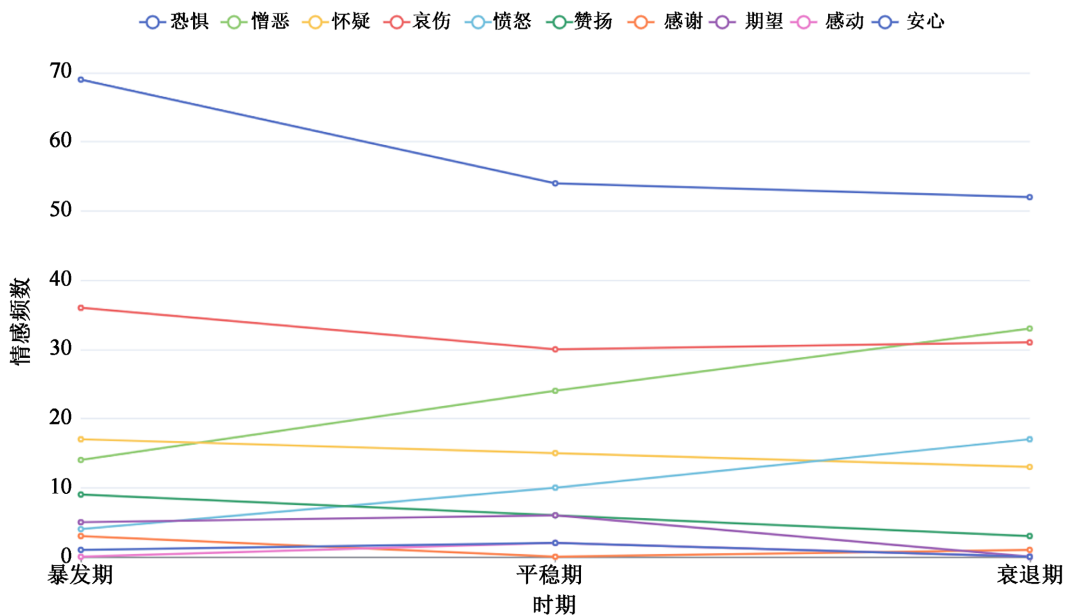


Figure 7. Distribution of positive and negative emotions of users in different periods  
图 7. 不同时期用户正负情感分布

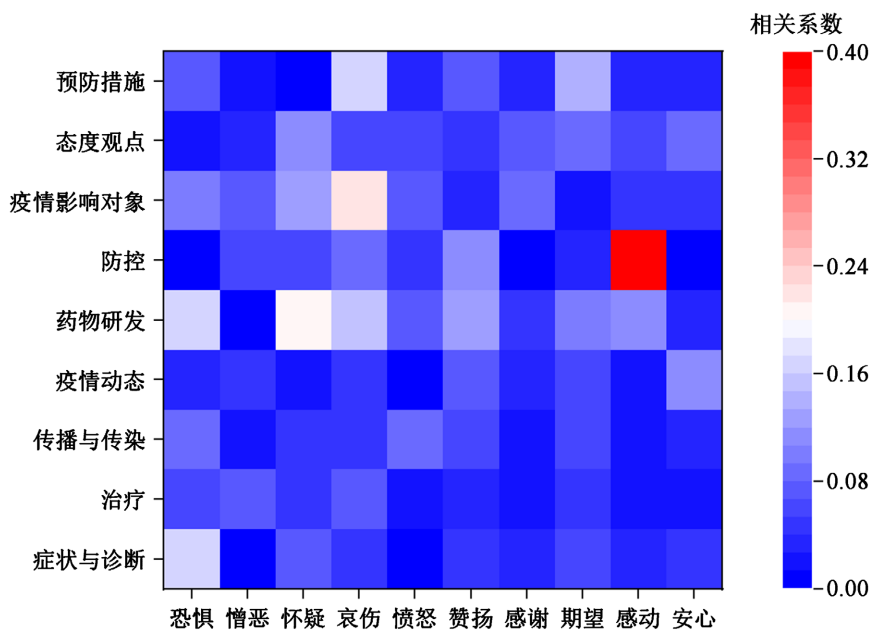


Figure 8. Correlative heat map between theme and emotion  
图 8. 主题与情感的相关热力图



由图 8 可知, 某一主题在很大程度上与某一正负面情绪呈现较强的相关关系。如与“症状与诊断”最为相关的是恐惧, 与“治疗”“疫情影响对象”“预防措施”突出相关的是哀伤, 与“传播与传染”最为相关的是愤怒, 与“药物研发”“态度观点”最为相关的是怀疑, 人们可能因出现某种症状而害怕自己被诊断为新冠肺炎, 可能担心新冠肺炎患者被治疗后存在后遗症, 也许会惋惜因新冠疫情遭受损失的各行各业, 会因为应对疫情的预防措施比较局限而感到些许无奈, 也会因为新冠病毒传播广、传染性强而表现出怨恨, 同时对特效药的研发和某些不实言论持质疑态度。而与“疫情动态”最为相关的是安心, 与“防控”最为相关的是感动, 这表明公众对目前互联网上实时更新的疫情动态表示肯定, 也非常认可政府及有关部门为控制疫情所做的防控措施。

## 5. 结论与建议

本文以百度知道上新新冠肺炎相关问答记录文本为例, 主要探讨了突发公共卫生事件下公众健康信息需求及演变模式、情感变化及其与主题相关性。经分析, 主要有如下三点发现:

公众健康信息需求类型具有层次鲜明、范围广泛的特点, 同时需求随时间推移波动明显但最终演化趋势相对集中。公众健康信息需求包括 3 大主类别, 其中各主类别又可分为 2~5 个子类别。公众健康信息需求随疫情进展变化明显, 暴发期、平稳期和衰退期公众关注主题不一。随着疫情的发展, 公众关注点从国内转向国外, 再由国外转向国内局部地区; 从疫苗研发上市转向疫苗副作用; 从相信认可网络健康信息到质疑揭穿网络健康谣言。在国内疫情处于衰退期的形势下, 公众主要集中在因局部地区疫情反弹引发的对传播与传染、预防措施、疫情影响对象的关注, 以及对疫苗等药物研发的持续关注。

公众情感倾向整体呈中性, 表现出走向积极又走向消极的趋势, 并且公众具体正负面情感随时间变化波动较大。公众在社会化问答平台上大多提问或回答与新冠疫情相关的客观问题, 所以公众总体情感偏中性。同时, 随着国内疫情整体缓和, 公众情感总体上走向积极; 但国内局部地区疫情反复, 公众情感有走向消极的趋势。具体而言, 负面情绪以恐惧和哀伤为主, 正面情感以赞扬和期望居多。整体疫情得到控制使得公众负面情感中的恐惧、哀伤、怀疑占比有所下降, 但局部疫情突发又使得憎恨、愤怒比重有所上升; 而公众的正面情感主要出现在暴发期和平稳期, 大多是对政府等有关部门及工作人员的正面回应。

公众健康信息需求与其正负面情感间存在一定的相关关系。就负面情感而言, 公众害怕感染新冠肺炎, 对新冠肺炎治疗可能产生后遗症、自身预防措施有限以及新冠疫情对整个社会带来的损失感到悲伤, 也因新冠病毒传染性强使得局部地区疫情频发而感到愤怒, 同时对诸如新冠特效药研制成功等药物研发方面的不实言论持怀疑态度。此外, 公众对政府实时公布疫情动态的行为感到安心, 并且十分肯定政府及相关部门为新冠疫情得以控制所做出的贡献。

通过以上总结, 本文针对性地对相关部门提出以下三点建议:

以多渠道、多样化的方式全面发布各类健康信息, 同时根据疫情进展合理调整信息发布。有关部门可通过微博、微信公众号、知乎、百度知道等平台发布诸如新冠病毒基本信息、社会影响、公众预防措施等信息。在疫情的不同发展阶段, 健康信息的发布应有所侧重, 比如当前处于衰退期, 整体疫情好转而局部疫情反复, 相关部门应及时提醒公众做好个人防护措施, 此外, 还应大力宣传接种疫苗的必要性。

采取有效措施来稳定公众情绪, 并根据疫情变化提供公众心理疏导的渠道。随着病毒变异引发局部地区疫情暴发, 相关部门应及时安抚公众情绪并提醒公众要做好长期抗疫的准备, 有关网站也可推出心理疏导方面的科普文章或视频以缓解公众对疫情的恐惧和悲伤。此外, 针对网络上新冠疫情相关信息, 政府要严格把控其质量并及时公开辟谣, 以防谣言扰乱视听导致公众情绪紊乱。同时, 公众也应做到不信谣、不传谣。

根据公众关切的需求及时做出合理回应以便更好地为公众提供服务。比如相关部门可明示公众做好个人防护工作是避免感染病毒的重要前提以减缓公众对感染新冠病毒的恐惧, 也可通过政务微博、公众号等官方渠道发布疫苗注射方面的真实信息来消除公众内心的疑惑, 还可给公众树立信心以重振他们因疫情而受挫的各行各业的士气。

此外, 本文还存在一些不足之处。本文只选取百度知道这一社会化问答平台作为研究对象, 因而得出的公众健康信息需求可能受制于平台属性而表现出单一、不全面的特点。因此, 未来可考虑多个平台对比研究, 以发现不同平台属性对用户健康信息需求的影响。

## 基金项目

国家社会科学基金一般项目(21BGL243); 上海市哲学社会科学规划一般项目(2020BGL005)。

## 参考文献

- [1] 张宁, 李华, 李培植, 等. 突发公共卫生事件中的互联网公众反应——以知乎平台新冠肺炎相关话题为例[J]. 现代情报, 2021, 41(2): 11.
- [2] 邝良锋, 罗昱夫. 网络舆情的实质与治理——基于新冠肺炎的舆情分析[J]. 华南师范大学学报(社会科学版), 2021(2): 193-204+208.
- [3] 石静, 厉臣璐, 钱宇星, 周利琴, 张斌. 国内外健康问答社区用户信息需求对比研究——基于主题和时间视角的实证分析[J]. 数据分析与知识发现, 2019, 3(5): 1-10.
- [4] Fink, S. (1986) *Crisis Management: Planning for the Inevitable*. American Management Association, New York, 20.
- [5] Burkholder, B.T. and Toole, M.J. (1995) Evolution of Complex Disasters. *The Lancet*, **346**, 1012-1015. [https://doi.org/10.1016/S0140-6736\(95\)91694-6](https://doi.org/10.1016/S0140-6736(95)91694-6)
- [6] Robert, H. (2004) *Emergency Management*. China Citric Press, Beijing, 22.
- [7] 董晨, 申琦, 郭颖, 等. 新冠肺炎疫情期间上海市公众疫情相关信息需求情况分析[J]. 中国公共卫生, 2022, 38(6): 800-803.
- [8] 周艳玫, 刘东苏, 王衍喜, 郝丹. 大学生信息行为调查分析与信息服务对策[J]. 图书情报工作, 2015, 59(6): 61-67.
- [9] 陈静, 张璐, 陆泉. 突发公共卫生事件中大学生健康信息需求动因与主题研究[J]. 图书情报工作, 2021, 65(6): 82-92.
- [10] 安璐, 杜廷尧, 李纲, 余传明. 突发公共卫生事件利益相关者在社交媒体中的关注点及演化模式[J]. 情报学报, 2018, 37(4): 394-405.
- [11] 赵一鸣. 基于多维尺度分析的潜在主题可视化研究[D]: [博士学位论文]. 武汉: 华中师范大学, 2013.
- [12] 郭婷, 郑颖. 数据挖掘在国内图书情报领域的应用现状分析——基于文献计量分析和共词分析[J]. 情报科学, 2015, 33(10): 91-98.
- [13] Cai, H., Li, Z., Yan, C., et al. (2018) A Shallow Neural Network Based Short Text Classifier for Medical Community Question Answering System. 2018 *IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Tianjin, 19-23 July 2018, 1537-1541. <https://doi.org/10.1109/CYBER.2018.8688090>
- [14] Qiang, X.L., Xu, P., Fang, G., et al. (2020) Using the Spike Protein Feature to Predict Infection Risk and Monitor the Evolutionary Dynamic of Coronavirus. *Infectious Diseases of Poverty*, **9**, 30-37. <https://doi.org/10.1186/s40249-020-00649-8>
- [15] Zhang, J., Chen, Y., Zhao, Y., et al. (2020) Public Health and Social Media: A Study of Zika Virus-Related Posts on Yahoo! Answers. *Journal of the Association for Information Science and Technology*, **71**, 282-299. <https://doi.org/10.1002/asi.24245>
- [16] Zhao, Y., Cheng, S., Yu, X., et al. (2020) Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. *Journal of Medical Internet Research*, **22**, e18825. <https://doi.org/10.2196/18825>
- [17] 刘冰, 厉鑫, 张赫钊, 翟羽佳. 网络健康社区中身份转换期女性信息需求主题特征及情感因素研究——以“妈妈网”中“备孕版块”为例[J]. 情报理论与实践, 2019, 42(5): 87-92.
- [18] 新京报. 加码泛知识内容赛道, 百度将为优质答主提供千亿流量扶持[EB/OL].

<https://baijiahao.baidu.com/s?id=1686401331437810777&wfr=spider&for=pc>

- [19] 肖雪, 周静. 老龄化背景下我国公共图书馆老年服务状况的调查与分析——基于内容分析法的实证研究[J]. 图书情报知识, 2013(3): 16-27.
- [20] Braun, V. and Clarke, V. (2008) Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, **3**, 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- [21] Strauss, A. (1987) *Qualitative Analysis for Social Scientists*. Cambridge University Press, Cambridge, 5. <https://doi.org/10.1017/CBO9780511557842>
- [22] Viera, A.J. and Garrett, J.M. (2005) Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, **37**, 360-363. <https://doi.org/10.1007/978-3-540-75148-9>
- [23] Zhang, J. (2008) *Visualization for Information Retrieval*. Springer, Berlin.
- [24] 吴恒, 陈燕翎. 基于 UGC 文本挖掘的游客目的地选择信息研究——以携程蜜月游记为例[J]. 情报科学, 2017, 35(1): 101-105.
- [25] 徐琳宏, 林鸿飞, 赵晶. 情感语料库的构建和分析[J]. 中文信息学报, 2008(1): 116-122.
- [26] 翟羽佳, 过南杉, 阎嘉琪. 突发公共卫生事件中虚假信息的时滞性扩散与情感关联分析[J]. 情报科学, 2021, 39(5): 62-69.