

我国公众对远程医疗的态度变化研究

——基于多时段微博文本数据分析

倪载健, 钱 颖

上海理工大学管理学院, 上海

收稿日期: 2024年9月27日; 录用日期: 2024年12月5日; 发布日期: 2024年12月12日

摘 要

目的/意义: 远程医疗在疫情期间发挥了重要作用, 避免了公众在医院聚集而引起交叉感染, 为隔离中的人们提供医疗服务, 得到了大量使用。然而公众对远程医疗的态度发生了何种变化, 尚未得到深入研究。方法/过程: 我们以“远程医疗”、“在线问诊”和“互联网医院”为关键词, 收集了2018年1月1日至2024年4月30日期间的微博数据。利用融合word2vec的LDA主题模型和Bi-LSTM-Attention情感分析方法, 探讨了疫情前、疫情中、疫情后公众对远程医疗的态度变化。结果/结论: 结果显示, 公众对远程医疗的态度由疫情前的“不信任”演化为“在疫情中有重要作用”, 进而到疫情后仍然“推荐使用”远程医疗。公众对远程医疗的积极情绪占比不断提升, 消极情绪占比稳步下降。经历新冠疫情, 公众普遍认识到远程医疗的便利性, 愿意将远程医疗作为问诊的辅助手段, 疫情提高了公众对远程医疗的信任度。此外, 公众隐私焦虑和医保报销是目前远程医疗面临的主要问题, 公众希望政府能出台相关政策法规, 支持远程医疗的进一步发展。

关键词

远程医疗, 文本数据, 主题挖掘, 情感分析, COVID-19

Research on the Evolution of Public Attitudes towards Telemedicine in China

—A Multi-Period Analysis of Weibo Text Data

Zaijian Ni, Ying Qian

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Sep. 27th, 2024; accepted: Dec. 5th, 2024; published: Dec. 12th, 2024

Abstract

Purpose/Significance: Telemedicine has been extensively utilized during the pandemic, preventing

cross-infections due to public gatherings in hospitals and providing medical services to isolated individuals. However, the evolution of public attitudes towards telemedicine has not been thoroughly investigated. Method/Process: We collected Weibo data from January 1, 2018, to April 30, 2024, using “telemedicine”, “online consultation”, and “Internet hospital” as keywords. Employing a LDA topic model combined with word2vec and a Bi-LSTM-Attention sentiment analysis technique, this study examines the changes in public attitudes towards telemedicine before, during, and after the pandemic. Result/Conclusion: The findings indicate that public attitudes towards telemedicine shifted from mistrust prior to the pandemic to recognition of its importance during the pandemic, and continued to favor its use post-pandemic. The proportion of positive sentiments has risen consistently, while negative sentiments have decreased steadily. The COVID-19 pandemic has led the public to appreciate the convenience of telemedicine, making it a preferred consultation method. The experience of using medicine during the pandemic has enhanced public trust in telemedicine. Moreover, concerns about privacy and medical insurance reimbursement are the primary challenges currently faced by telemedicine. The public calls for government policies and regulations to support its further development.

Keywords

Telemedicine, Text Data, Topic Mining, Sentiment Analysis, COVID-19

Copyright © 2024 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

远程医疗(Telemedicine, TM)是一种社会化医疗系统,它通过整合多个信息系统,利用视频技术和信息技术手段,增强医生及其他医疗人员对患者的诊断与治疗能力[1]。远程医疗对于提升欠发达及偏远地区的医疗资源可及性具有重要意义。在中国,远程医疗的标志性起点可追溯至1997年,当时国家卫生计生委成立了中国金卫远程医疗网[2]。但是远程医疗的发展到2018才进入快速发展的阶段,尤其是在新冠疫情期间,医院作为潜在的高风险区域,远程医疗有效减少了患者因就诊而感染病毒的风险,并为隔离人群提供了必要的医疗服务,因而得到了广泛应用[3]。2020年底,全国已有29个省份建立了省级远程医疗平台,远程医疗服务覆盖了88.46%的县级行政区[4]。《中国互联网络发展状况统计报告》数据显示,2020年6月我国在线医疗的用户规模达到2.7亿人,使用率为29.4%,截止至2023年6月份我国在线医疗的用户规模达到3.64亿人,较2020年6月增长了8814万人,使用率达到33.8% [5]。

新冠疫情大大推动了远程医疗在中国的使用量,公众对远程医疗的态度在疫情前后有什么变化是值得研究的问题。社交媒体上公众分享的关于远程医疗的文本为我们提供了丰富的数据。因此,本文以新浪微博为数据获取平台,采用LDA主题模型和情感分析方法,分析公众对远程医疗的关注点和情感态度在疫情前后的变化情况,为政府和远程医疗相关企业提供管理启示,为促进远程医疗的推广和普及提供实践指导。

2. 文献综述

2.1. 远程医疗的公众感知相关研究

随着远程医疗的发展,公众对其的态度也受到学者的关注[6]-[8],其中大部分研究是通过问卷调查方式开展的。如Sun等[9]通过问卷调查发现,隐私焦虑是阻碍公众使用远程医疗的重要因素。李旭丹等[10]通过问卷调查高校学生对在线健康咨询服务的使用意愿,指出增加对网上医生的信任和提高患者潜在利

益是提升使用意愿的关键。Herran 等[11]的在线问卷调查显示, 大多数患者对远程医疗持积极态度, 并希望其成为医疗服务的一部分。Jeraq 等[12]通过电话访谈发现, 远程医疗并未影响患者对医疗服务的满意度, 且线下问诊后选择远程医疗复诊的患者更倾向于对远程医疗表示满意。疫情期间, 远程医疗的使用更普遍。Massaad 等[13]发现疫情期间存在疾病焦虑, 呼吁政府推广远程医疗政策, 以应对心理健康问题。Holtz 等[14]在疫情期间的问卷调查研究发现, 公众对远程医疗服务表示满意。Malone 等[15]考察了疫情期间公众使用远程医疗平台的意愿, 发现互联网连接问题和老年人技术使用难度是阻碍远程医疗的因素。

2.2. 利用社交媒体数据挖掘公众态度相关研究

随着 Web 2.0 技术的快速发展, 社交媒体已成为不同情境下获取和交换信息的重要渠道。社交媒体数据因其海量性、客观性以及跨越时空的特性, 成为研究公众情绪、态度和行为的宝贵资源[16]-[18]。在公共卫生领域的研究中, 这种研究方法也被大量应用。Jabalameli 等[19]通过自然语言处理技术对 Twitter 上的数据进行分析, 揭示了公众对新冠疫情的态度和看法, 发现 Twitter 上的舆论在很大程度上与主流媒体的报道相一致。Wang 等[20]对新浪微博上与 COVID-19 相关的帖子进行主题建模和情感倾向分析, 为疾病跟踪和公共卫生决策提供了有价值的参考。在远程医疗方面, 已有一些研究通过文本分析研究了公众对远程医疗的态度。高芳芳等[21]利用 LDA 主题模型和情感分析方法发现网络上针对在线问诊的态度偏负面。Pal 等[22]通过文本挖掘和机器学习方法探索了影响远程医疗用户体验的因素, 并发现时间与金钱、便利性、响应性以及易用性是提供积极用户体验的关键因素。Krittawong 等[23]利用文本数据挖掘技术, 分析了疫情期间社交媒体上与远程医疗相关的推文, 发现积极情绪与远程医疗的安全性、便利性相关, 而消极情绪则与成本、技术障碍等问题有关, 为远程医疗的优化提供了方向。这些研究表明, 社交媒体内容挖掘在理解公众对特定领域的态度和情绪方面具有重要作用, 为政策制定者和公共卫生管理者提供了科学的决策依据。Gait'an 等[24]基于刺激-生物体-反应模型, 利用 Twitter 上的推文分析疫情期间西班牙公众对远程医疗的认知演变, 发现了疫情影响了公众对远程医疗情绪与话题, 远程医疗由疫情前的积极形象演变为疫情后较差的形象。

3. 研究框架

3.1. 研究方法

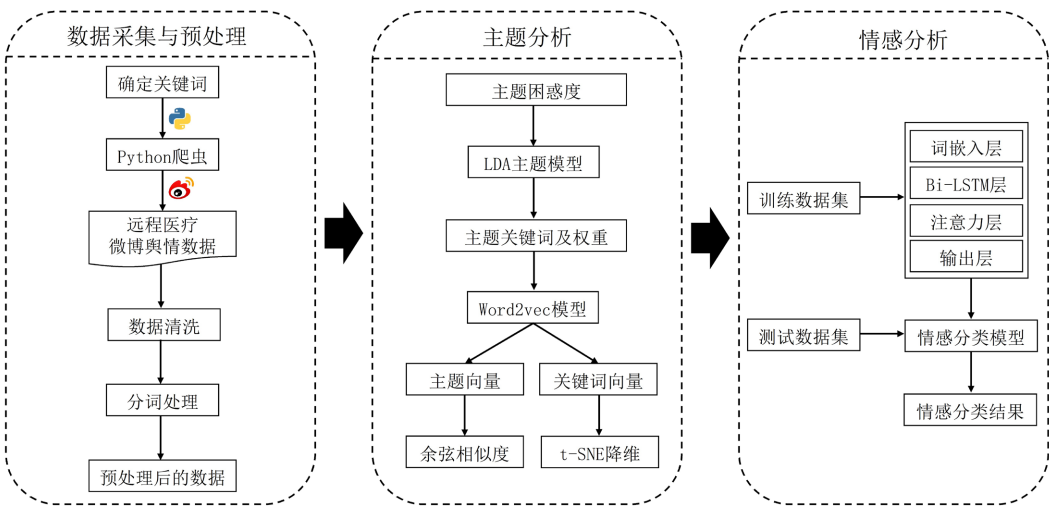


Figure 1. Analysis framework for public attitudes of the telemedicine
图 1. 远程医疗公众态度分析框架

本研究通过对疫情前、疫情中、疫情后的微博相关文本进行主题分析和情绪分析, 揭示了我国公众对远程医疗态度的变化, 总体研究框架如图 1 所示。

3.2. 研究方法数据收集和预处理

本研究以“远程医疗”、“在线问诊”和“互联网医院”为关键词, 采用 Python 编程语言开发网络爬虫, 爬取了 2018 年 1 月 1 日至 2024 年 4 月 30 日间微博数据。包括“用户 ID”、“用户昵称”、“微博正文”及“发布时间”等, 共获得微博文本数据 157,967 条。为有效提升数据质量, 我们对原始数据进行清洗, 剔除原始文本数据中的重复项、表情符号、网页链接和特殊字符等无效信息。经过上述数据清洗流程, 最终得到有效数据 116,864 条。此外, 本研究使用 jieba 分词工具对清洗后的文本进行了分词处理, 将完整句子拆分为独立词语, 并以哈工大提供的停用词表作为参考, 以提高分词的准确性和文本分析的质量。

3.3. 主题分析

LDA 主题模型, 是由 Blei 等人提出的一种生成式主题模型[25], 现在被广泛用于发现文档集合中的潜在主题。与传统的词频统计不同, LDA 考虑文本之间语义关联, 能识别短语背后的语义相关性。进行 LDA 建模的关键是确定主题数量 K 。为保证主题模型结果的质量, 选用困惑度[26]确定最佳的 topic 数目, 并采用基于 t-SNE 降维算法的数据降维可视化工具对 LDA 主题模型提取的主题进行可视化来[27]。

Word2vec 是 Tomas Mikolov 在 2013 年提出的非监督深度学习算法[28], 旨在将自然语言中的单词表示为具有理解上下文实际含义的低维向量。Word2vec 主要基于两种不同的训练模型, 即连续词袋模型 (Continuous Bag of Words, CBOW) 和跳字模型 (Skip-Gram)。Skip-Gram 模型的核心思想是根据当前词的概率去预测该词前后区间词的概率。CBOW 模型的训练方式恰好与 Skip-Gram 模型相反。为了更好地识别远程医疗舆情的主题及演化路径, 本文构造了融合 LDA 和 Word2vec 的 LDA-Word2vec 主题识别与演化模型。每个步骤的细节解释如下。

- (1) 主题提取。对预处理后的语料进行主题建模, 生成疫情前、疫情期间和疫情后主题关键词及权重。
- (2) 词向量训练。使用 Word2vec 模型训练预处理后的语料得到词向量。
- (3) 主题向量化。输入主题关键词到词向量模型中, 得到主题关键词词向量。然后, 对各主题下的关键词词向量及其权重进行加权求和, 得到最终的主题向量。计算公式如下:

$$Topic_{vector} = \sum_{i=1}^n W_{mi} \cdot a_{mi} \quad (1)$$

- (4) 主题相关性测量。主题相关性通过余弦相似度[29]来计算:

$$Topic_{correlation} = similarity = \frac{L_{ij} \cdot L_{i+1j}}{\|L_{ij}\| \|L_{i+1j}\|} = \frac{\sum_{i=1}^n L_{ij} \cdot L_{i+1j}}{\sqrt{\sum_{i=1}^n (L_{ij})^2} \sqrt{\sum_{i=1}^n (L_{i+1j})^2}} \quad (2)$$

其中, L_{ij} 和 L_{i+1j} 分别表示全局主题和局部主题的两个不同的主题向量。

- (5) 可视化。对关键词词向量进行 t-SNE 降维可视化。同时, 利用主题余弦相似度绘制主题内容演化桑基图。

3.4. 情感分析

在本研究中, 我们利用 Bi-LSTM-Attention 情感分类方法对微博文本数据进行深入的情感分析。如图 2 所展示的模型架构图, 该模型构建了一个综合性的网络框架, 其设计包含了若干关键且不可或缺的组件, 分别是输入句子层、词嵌入层、Bi-LSTM 层、自注意力层和一个 Softmax 分类器。

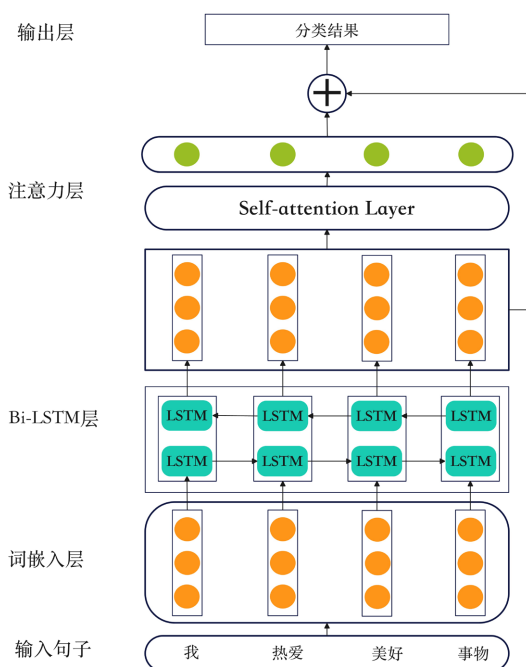


Figure 2. Overall architecture diagram of Bi-LSTM-Attention model

图 2. Bi-LSTM-Attention 模型的整体架构图

Bi-LSTM 利用前后两个方向的上下文特征信息, 有效解决了上下文信息的提取和利用问题。Bi-LSTM-Attention 模型进一步考虑了输入词向量对文本情感分类结果的具体影响。该模型通过计算词向量的概率权重, 为关键词汇分配更大的比重, 并以此优化情感分类的特征表示, 进一步提升了模型的分类精确度。模型具体步骤如下:

(1) 词嵌入层。采取 Word2Vec 模型对预处理好的数据进行预训练得到词向量。

(2) Bi-LSTM 层。通过 Bi-LSTM 模型将文本词汇映射到相同向量空间记 Bi-LSTM 中的前向 LSTM 和后向 LSTM 在时刻 i 的输入处理分别为:

$$\vec{h}_t = LSTM\left(x_t, \vec{h}_{t-1}\right) \quad (3)$$

$$\overleftarrow{h}_t = LSTM\left(x_t, \overleftarrow{h}_{t-1}\right) \quad (4)$$

$$h_t = \left[\vec{h}_t, \overleftarrow{h}_t \right] \quad (5)$$

Bi-LSTM 层的输出为:

$$h = (h_1, h_2, \dots, h_n) \quad (6)$$

(3) 自注意力层。通过融合 Attention 机制, 可以实现对当前输入进行权重调整, 对关键信息给予更多的关注。注意力权重的计算过程如下:

$$K = \tanh\left(\frac{W_h h}{W_a h_a \otimes e_k}\right) \quad (7)$$

$$\alpha = \text{Soft max}\left(w^T K\right) \quad (8)$$

$$r = h\alpha^T \tag{9}$$

其中, $K \in R^{(d+d) \times T}$, $\alpha \in R^T$, $r \in R^d$, $W_h \in R^{d \times d}$, $W_a \in R^{d_a \times d_a}$ 以及 $w \in R^{d+d_a}$ 都是参数矩阵。 α 是注意力权重矩阵, 而 r 是输入句子的权重表示。 $h_a \otimes e_k$ 表示向量 h 重复拼接了 k 次, 而 e_k 是大小为 k 的列向量。最终用于分类的句子表示为:

$$h^* = \tanh(r) \tag{10}$$

(4) Softmax 分类器。该文通过 Softmax 层来计算条件概率分布, 即预测输入句子对应的标签。条件概率的计算公式为:

$$\hat{y} = \text{Soft max}(W_{sm}h^* + b_{sm}) \tag{11}$$

式中, W_{sm} 为权重参数, b_{sm} 为偏置参数。

为更好地评估本文情感模型的分析质量, 本研究采用了第九届全国社会媒体处理大会(SMP 2020)公开竞赛所提供的标准化数据集进行模型训练。该数据集由两个主要部分组成: 第一部分是针对疫情相关主题的微博数据集, 第二部分则是一个广泛覆盖的通用微博数据集。通过这两个子集的融合, 本研究共收集了 49,374 条经过精确标注的微博文本, 其中测试数据集包含 41,374 条样本, 训练数据集则包含 8,000 条样本。这些样本覆盖了六种基本情绪类别: 乐观、愤怒、中立、悲伤、惊讶以及害怕。本研究中所提出的模型在测试数据集上进行了性能评估, 其评价指标如下: 模型的精确率为 86.11%, 精准率为 86.09%, F1 (H-mean)值为 86.08%, 召回率为 86.11%。此外, 模型的 AUC (Area Under roc Curve)值为 0.92, 表明模型在情感分类任务上具有良好的性能。

4. 模型结果与讨论

4.1. 词频统计结果及词云图分析

通过分析疫情前、疫情中和疫情后三个阶段的词频统计结果和词云图(表 1、图 3), 可以观察到公众对远程医疗态度的演变: 疫情前, 与“远程医疗”直接相关的词汇出现频率不高, 仅出现了“互联网”等。公众关注点集中在基础医疗服务上, 如“医疗”、“医生”、“患者”等, 这表明疫情前对远程医疗的认知度相对较低。“诈骗”、“骗子”和“隐私”表明网民对远程医疗的不信任和隐私焦虑。疫情中, 与远程医疗相关的词汇如“远程医疗”、“互联网”、“在线”等词频显著增加, 显示公共对远程医疗服务的认知提升。“疫情防控”、“隔离”等词汇说明在疫情期间, 远程医疗成为防控疫情、减少接触的重要手段。词频中“方便”、“快速”、“居家”等词汇的增加, 反映了疫情期间人们对远程医疗服务便利性和可及性的高度认可。疫情后, “科技”、“数字”、“智能”等词汇的词频上升, 反映出疫情后人们对远程医疗行业技术发展和数字化转型的期待。“监管”、“政策”等词汇的出现, 说明疫情后民众对规范远程医疗服务、加强政策监管的需求。

Table 1. Top 20 High-frequency words
表 1. 排名前 20 的高频词

时期	特征词
疫情前	医疗 医生 患者 互联网 时间 信息 专家 数据 技术 挂号 诈骗 咨询 远程 老人 骗子 药品 功能 隐私 宠物 焦虑
疫情中	医院 疫情 医疗 防控 互联网 时间 风险 优势 肺炎 咨询 居家 专家 交叉感染 挂号 核酸 隐私 免费 技术 数据
疫情后	医院 健康 患者 医疗 互联网 医生 门诊 时间 便捷 管理 技术 渠道 药品 疫情 咨询 数据 挂号 老人 专家 风险

综上所述, 疫情后公众远程医疗的认知度显著提高, 疫情成为推动远程医疗发展的重要契机。在疫情期间, 远程医疗的便利性和可及性得到了广泛认可, 满足了人们在特殊时期的就医需求。疫情后, 人们期望远程医疗行业能够更加智能化、数字化, 同时提出了加强监管的要求。



Figure 3. Word cloud map
图 3. 词云图

4.2. 主题结果分析

在本研究中, 为确定适宜的主题数量, 本文采用了困惑度作为评估指标。经过计算, 疫情前、中、后的微博文本分别聚类为 6 个、7 个和 6 个主题。利用 LDA 模型进行主题建模, 模型输出了每个主题下的关键词及其相应的权重。在每个主题中, 根据权重的高低, 本文筛选了前 9 个关键词, 以此代表该主题的核心内容。为了更直观地展示主题模型的聚类效果, 本文采用了 t-SNE 算法对 LDA 模型的提取结果进行数据降维和可视化处理。最终的可视化结果展示在图 4 中。

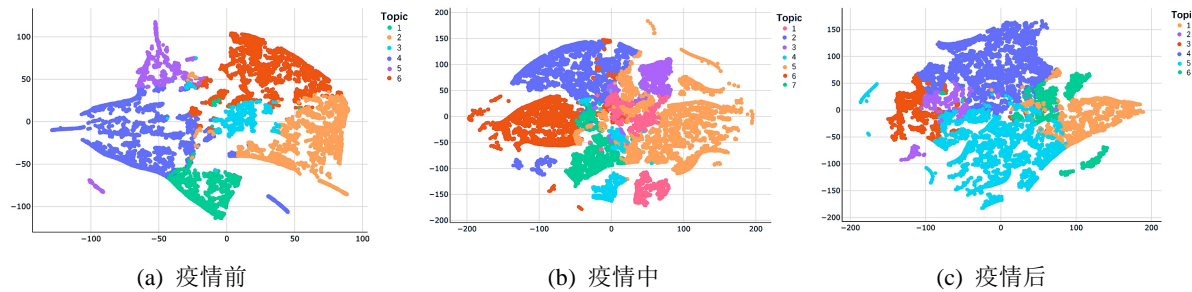


Figure 4. Visualization of topic extraction results
图 4. 主题提取结果可视化

疫情前微博文本可聚类为 6 个主题, 见表 2。主题 1 强调了远程医疗在节省时间和提高便利性方面的优势。关键词如“异地”“省时”和“预约”反映网民认为远程医疗能够减少就诊时间, 方便患者通过网络平台进行医疗咨询。主题 2 是用户体验分享。“担心”“不靠谱”和“骗子”等关键词揭示了公众对远程医疗服务在疫情前普遍持有的不信任态度。主题 3 围绕医疗服务的质量和网络诈骗问题展开。“莆田系”“监管”和“资质”等关键词显示了公众对于医疗服务提供者的合法性和专业性的关注, 以及对网络诈骗潜在风险的担忧。主题 4 着重于数据隐私和信息安全。“隐私”“安全”和“泄露”等关键词表明用户对个人信息保护深切关注, 担心远程医疗数据泄露和滥用。主题 5 讨论了医疗技术的进步和行业的发展。关键词如“技术”“互联网”和“智能”等关键词反映出人们对医疗行业技术创新和智能化发展的热切期待。主题 6 涉及政策和市场监管。“监管”“整治”和“认证”指出公众对加强医疗行业监管、保护患者权益和规范市场秩序的迫切需求。

疫情中微博文本可聚类为 7 个主题。主题 1 强调了远程医疗在疫情防控中的重要作用，关键词“感染”“防控”和“隔离”表明远程医疗可以防止感染扩散，有助于疫情防控。主题 2 着重于疫情中远程医疗为患者提供便捷和安全的就医途径。其中“看病”“远程”和“居家”等关键词反映了公众对远程医疗服务的依赖性增加。主题 3 围绕用户体验分享。关键词“推荐”“见效”和“安心”表明疫情期间患者对远程医疗服务的满意度信任度都很高。主题 4 聚焦于远程医疗隐私保护和数据安全问题。关键词“担忧”“曝光”和“泄露”反映了人们对于个人信息安全的担忧。主题 5 探讨了医保报销政策。关键词“医保”“报销”和“异地”显示了公众对远程医疗纳入医保报销的广泛期望，并反映了政府对此的积极回应。主题 6 关注了疫情期间医疗行业的技术进步和经济变化。关键词“发展”“投资”和“企业”表明了人们希望远程医疗行业能发展起来，吸引更多投资和企业。主题 7 涉及政府对医疗行业的监管和市场乱象的治理。关键词“权益”“打击”和“监管”指出了公众对加强医疗行业监管和规范市场秩序的需求。

Table 2. Topic identification results
表 2. 主题识别结果

主题	疫情前		疫情中		疫情后	
	主题标签	关键词	主题标签	关键词	主题标签	关键词
主题 1	远程医疗的优势	异地 省时 远程 预约 方便 快捷 老人 网络 访问	远程医疗与 疫情防控	疫情 感染 防控 核酸 隔离 社区 抗疫 措施 见效	推荐使用远 程医疗	推荐 复诊 咨询 方便 挂号 省心 异地 用药 小病
主题 2	远程医疗使用体验	感冒 担心 不靠谱 交流 用药 症状 指导 慢性病 骗子	远程医疗的 优势	看病 隔离 安全 疫情 远程 方便 快速 应急 居家	数据隐私与 信息安全	远程 资料 犹豫 曝光 怀疑 体检 态度 焦虑 平台
主题 3	服务质量与 网络诈骗	网络 诈骗 莆田系 法律 立案 百度 犯罪 资质 癌症	远程医疗使用 体验	疫情 隔离 感激 推荐 见效 感染 发烧 靠谱 安心	远程医疗 的优势	医院 服务 医疗 健康 便捷 医保 在线 咨询 网络
主题 4	数据隐私与 信息安全	信息 数据 隐私 安全 存储 泄露 焦虑 滥用 记录	数据隐私与 信息安全	医生 网络 数据 担忧 曝光 记录 隐私 泄露 信息	医保报销政 策	优惠 异地 报销 费用 比例 大病 保障 流程 线上
主题 5	远程医疗行业 前景	技术 数据 互联网 管理 前景 智能 发展 信息 设备	医保报销政 策	医保 服务 医疗 社保 老人 报销 政策 缴费 异地	政策监管与 市场规范	监管 打击 保障 假冒 安心 模式 咨询 门槛 平台
主题 6	政策监管与 市场规范	监管 服务 权益 整治 保护 注册 行业 政策 认证	技术发展与 行业前景	发展 疫情 平台 经济 技术 数据 健康 投资 企业	疫情促进行 业发展	信息 疫情 AI 前景 技术 产品 市场 行业 科技
主题 7			政策监管与 市场规范	权益 打击 罚款 牌照 监管 乱象 模式 规范 保障		

疫情后文本可为 6 个主题。主题 1 展现了对远程医疗推荐的倾向，关键词如“推荐”“复诊”和“咨询”突显了公众对远程医疗服务的接受度。随着疫情的扩散，远程医疗平台成为轻症患者的重要就诊渠道，相关话题标签如“#线上问诊发热人数一周增加超 100 万#”与主题一紧密相关。主题 2 集中于患者对远程医疗服务中个人隐私和数据安全的担忧，关键词“资料”“曝光”和“怀疑”揭示了患者使用服务时的顾虑和不安。主题 3 显示了互联网医疗在疫情期间的普及程度，以及其对提高医疗服务可及性的贡献。关键词“服务”“医疗”和“健康”显示了远程医疗服务的广泛认可。主题 4 着重于医保报销流

程的复杂性,特别是异地就医的报销难题,关键词“报销”“费用”和“流程”表达了患者对医保政策的关注及改进的期望。主题 5 探讨了政策监管对远程医疗市场信任的影响,关键词“监管”“打击”和“假冒”既体现了政府在建立市场信任和打击非法医疗行为中的职能,也反映了公众对医疗市场规范管理的持续关注。主题 6 关注了技术创新在疫情期间对医疗行业的推动作用,关键词“人工智能”“创新”和“技术”展示了科技进步如何助力医疗行业的发展和应对疫情挑战。

4.3. 主题演化分析

为进一步探究远程医疗相关帖子在疫情前后的主题演化趋势,结合 LDA 主题模型输出关键词及权重,计算出不同阶段主题直接的余弦相似度,绘制舆情演化桑基图,如图 5 所示。通过颜色和宽度的变化,桑基图可直观展示主题间的演化过程,辨别关键主题以及对其他主题的影响程度。从图 5 中可见,“远程医疗的便捷性和效率”、“数据隐私和信息安全”和“政策监管与市场规模”是疫情前、疫情期间和疫情后三个阶段共有的主题,“医疗技术发展与行业前景”是疫情前和疫情期间共有主题,“医保报销政策”是疫情期间和疫情后共有主题。“服务质量与网络诈骗”是疫情前独有的主题,“疫情防控与远程医疗”是疫情期间独有的主题。“推荐使用远程医疗”和“疫情促进远程医疗发展”是疫情后独有的主题。

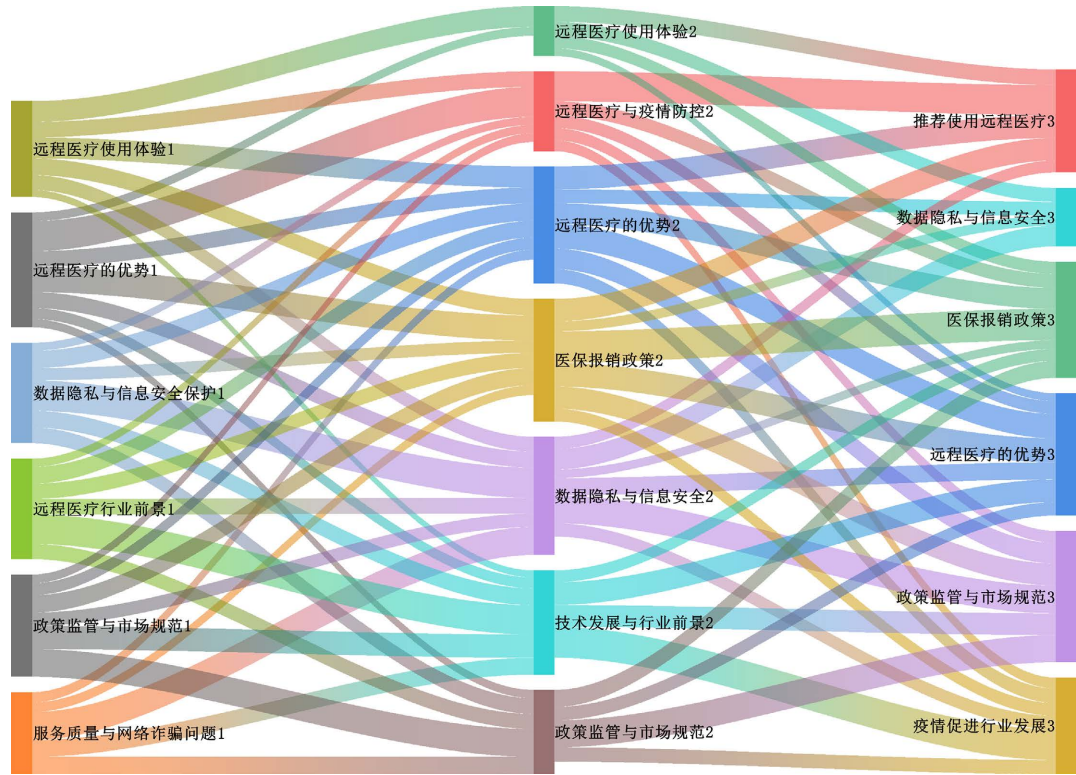


Figure 5. Topic evolution sankey diagram
图 5. 主题演化桑基图

通过桑基图我们发现两条不同阶段主题演化路径。首先,疫情前“远程医疗的优势”相关的讨论演变为疫情期间的“远程医疗与疫情防控”,最终演化为疫情后的“推荐使用远程医疗”。疫情前,公众已经认识到远程医疗因其省时便利、节约就医成本的优势。新冠疫情爆发后随着感染病例的增加,线下医疗资源紧缺,而且存在聚集风险。远程医疗由于是线上问诊,其便利性等优势使其在疫情中发挥了巨大作用。疫情后,基于疫情中对远程医疗的使用经验越来越多的人认可远程医疗的作用,开始推荐使用

远程医疗。另外，疫情前“服务质量和网络诈骗”相关的讨论演变为疫情期间“数据隐私和信息安全”相关的讨论，进而演变为疫情后的“政策监管与市场规范”。疫情前发生“魏则西事件”、“莆田系医院”等一系列事件使民众对远程医疗形成了负面态度，网络上出现“网络问诊，癌症起步”等舆论。公众普遍认为当前技术、法律法规、运营体系尚不成熟，公众对网络提供的在线医疗服务及信息不信任、对在线医疗医生专业能力的不信任和对平台保护个人隐私能力的不信任。疫情期间，受防疫政策的管制和远程医疗平台的推广，越来越多公众主动或被动的选择了远程医疗平台作为问诊的工具。这一时期公众担心个人的问诊信息会遭到泄露，导致不必要的麻烦。疫情后，隐私焦虑仍然是公众对远程医疗关注的热点之一，网民呼吁国家加大对远程医疗平台的监管力度，规范远程医疗的市场管理等，以更好地保护消费者的权益。

4.4. 情感分析

在本研究中，我们将情感分析模型应用于经过主题模型分类后的所有微博帖子。图 6(a)~(c)显示疫情前、疫情中和疫情后不同情绪种类的比例和不同主题下的情绪分布情况。

疫情前，与远程医疗相关帖子的负面情绪占比高于正面情绪。其中，“中性”情绪的帖子占比最高，达到 40.9%，而“愤怒”情绪以 22.4% 的占比位居其次。此阶段，公众对远程医疗的认知和接受度不够，他们的反感来自于对远程医疗的不信任和隐私焦虑。进一步分析发现，与正面情绪联系最紧密的话题是“远程医疗的优势”，这些帖子强调了远程医疗省时便捷、降低就医成本的优点。与负面情绪联系最紧密的话题是公众对远程医疗的“服务质量与网络诈骗”的讨论。在这些讨论中，常见问题如“远程医疗靠谱吗？”以及对远程医疗服务安全性和可靠性的担忧。此外，还有对在线医生资质真实性的质疑，如帖子所述：“网上的医生是真是假无法查证，诊断的结果不可信”。

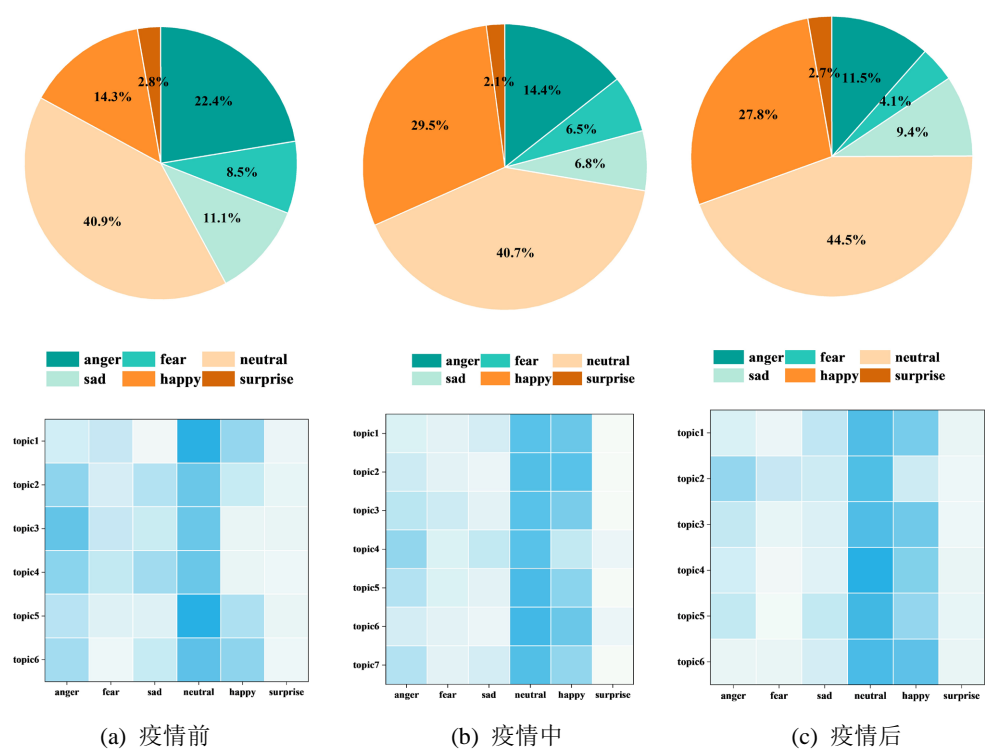


Figure 6. Three-stage sentiment evolution
图 6. 三阶段情绪演化

疫情中, 公众对远程医疗的态度发生了改变。尽管表达情感为“中性”的帖子依然占据最高比例, 但“快乐”情绪的表达迅速上升至 29.5%, 与此同时, “生气”和“悲伤”情绪的表达分别降至 14.4% 和 6.8%。这一变化反映了公众对远程医疗在疫情防控中作用的认可和重视。情绪分析显示, 与正面情绪密切相关的话题是“远程医疗与疫情防控”“远程医疗的优势”和“技术发展与行业前景”, 这些讨论强调了疫情期间远程医疗带来的便利性和实用性, 以及其对个人生活的积极影响。然而, 与负面情绪联系最紧密的话题是公众使用远程医疗的隐私焦虑, 一些网民表达了对个人信息和问诊记录泄露的担心, 特别是对部分平台未经同意匿名公开问诊记录的行为表示了强烈的不满和困扰。

疫情后, 社交媒体上与远程医疗相关的帖子中正面情绪的占比持续上升。情绪分析揭示, “远程医疗使用体验”、“远程医疗的优势”以及“疫情促进行业发展”等话题与正面情绪表达最为紧密相关。这些讨论表明, 远程医疗被视为一种实用的工具, 尤其适用于那些因各种原因无法进行线下问诊的个体。与负面情绪联系最紧密的话题是“数据隐私与信息安全”和“政策监管与市场规范”, 部分网民表示体格检查是问诊重要的部分, 只通过语言交流和照片无法确保问诊的质量。此外, 公众对远程医疗服务提供者的资质持有疑虑, 并呼吁政府加强对远程医疗平台的监管, 以确保服务的专业性与合规性。

5. 结论与启示

在这项研究中, 通过 LDA 主题模型和情感分析技术对微博上远程医疗相关的帖子进行分析, 以探讨疫情影响下公众对远程医疗态度的动态变化。我们的研究发现, 公众对远程医疗的态度在疫情前后发生了反转, 疫情前的不信任和使用意愿低转变为疫情中肯定远程医疗的重要作用和看好远程医疗的发展。疫情提高了公众对远程医疗服务的认识和接受度。在疫情结束后正向的情绪持续增加, 公众依然希望远程医疗能够继续发展, 但是对于个人信息泄露的担忧一直存在, 公众广泛的呼吁加强对远程医疗的监管, 保护用户信息安全。医保报销是另一个关注的焦点, 公众通过社交媒体平台呼吁政府将远程医疗纳入医保报销范围得到了回应, 国家医保局宣布正式将远程医疗纳入医保报销范围内。

本研究运用文本分析、情绪分析等理论和技术, 借助长时段微博数据, 探索了公众对远程医疗的态度动态变化, 研究可以从几个方向扩展。首先, 本研究的结论是基于单一社交平台。单一来源的数据无法代表具有不同社会人口统计特征组合的所有用户。未来的研究可以将数据来源扩展至包括抖音等其他社交媒体平台。其次, 本研究采用的方法是文本挖掘, 它提供了对远程医疗观念变化的定性描述。在未来的研究中, 可以结合文本挖掘应用定量方法, 以识别与观念变化相关的因素并量化其影响。

基金项目

国家社会科学基金一般项目“新冠疫情中‘信息疫情’的形成机理及应对策略研究”(项目编号: 22BGL240)。

参考文献

- [1] Combi, C., Pozzani, G. and Pozzi, G. (2016) Telemedicine for Developing Countries. *Applied Clinical Informatics*, 7, 1025-1050. <https://doi.org/10.4338/aci-2016-06-r-0089>
- [2] Shen, Y., Chen, L., Yue, W. and Xu, H. (2021) Digital Technology-Based Telemedicine for the COVID-19 Pandemic. *Frontiers in Medicine*, 8, Article 646506. <https://doi.org/10.3389/fmed.2021.646506>
- [3] 蒋帅, 孙东旭, 翟运开, 等. 远程医疗在新冠肺炎疫情防控中的实践与探索[J]. 中国数字医学, 2021, 16(3): 109-113.
- [4] 李陈晨, 王振博, 黄国书, 等. 我国医疗机构远程医疗运营管理模式现状研究[J]. 中国医院统计, 2023, 30(4): 248-252.
- [5] CNNIC. CNNIC 发布第 53 次《中国互联网络发展状况统计报告》[EB/OL]. <https://cnnic.cn/6/86/88/index.html>, 2023-03-23.
- [6] 金智旻. 互联网医院用户满意度影响因素研究[D]: [博士学位论文]. 武汉: 华中科技大学, 2020.

- [7] 翟运开, 路薇, 赵杰, 等. 基于结构方程模型的远程会诊患者满意度研究[J]. 中国卫生政策研究, 2018, 11(9): 64-70.
- [8] 张利萍. 远程医疗对患者满意度的影响因素调查[J]. 中国药物与临床, 2020, 20(1): 37-39.
- [9] Sun, S., Zhang, J., Zhu, Y., Jiang, M. and Chen, S. (2022) Exploring Users' Willingness to Disclose Personal Information in Online Healthcare Communities: The Role of Satisfaction. *Technological Forecasting and Social Change*, **178**, Article 121596. <https://doi.org/10.1016/j.techfore.2022.121596>
- [10] 李旭丹, 龚泽鹏, Reinhardt, J.D. 基于收益-风险模型分析高校学生在线健康咨询服务使用意愿的影响因素[J]. 暨南大学学报(自然科学与医学版), 2020, 41(3): 253-259.
- [11] Herran, M., Cullere, M., Dezotti Atty, N.R. and Sarmiento, C.S. (2022) Patient's Opinion on the Maintenance of the Telemedicine Modality in the Post-Pandemic Time. *Fertility and Sterility*, **118**, e326. <https://doi.org/10.1016/j.fertnstert.2022.09.105>
- [12] Jeraq, M.W., Mulder, M.B., Kaplan, D., Lew, J.I. and Farra, J.C. (2022) Telemedicine during COVID-19 Pandemic: Endocrine Surgery Patient Perspective. *Journal of Surgical Research*, **274**, 125-135. <https://doi.org/10.1016/j.jss.2021.12.014>
- [13] Massaad, E. and Cherfan, P. (2020) Social Media Data Analytics on Telehealth During the COVID-19 Pandemic. *Cureus*, **12**, e7838.
- [14] Holtz, B.E. (2021) Patients Perceptions of Telemedicine Visits before and after the Coronavirus Disease 2019 Pandemic. *Telemedicine and e-Health*, **27**, 107-112. <https://doi.org/10.1089/tmj.2020.0168>
- [15] Viteri Malone, M.A., Cabrera Chien, L., Pergolotti, M., Canin, B., Battisti, N.M.L., Krok-Schoen, J.L., et al. (2023) Evolving Oncology Care for Older Adults: Trends in Telemedicine Use after One Year of Caring for Older Adults with Cancer during Covid-19. *Journal of Geriatric Oncology*, **14**, Article 101497. <https://doi.org/10.1016/j.jgo.2023.101497>
- [16] 娄岩, 杨嘉林, 黄鲁成, 等. 基于网络问答社区的老年科技公众关注热点及情感分析——以“知乎”为例[J]. 情报杂志, 2020, 39(3): 115-122.
- [17] 范昊, 庄逸彤. 基于知乎平台内容挖掘的元宇宙公众感知研究[J]. 现代情报, 2024, 44(2): 65-80.
- [18] 秦琴, 柯青, 谢雨杉. 全球健康危机下公众的情感和认知——信息搜寻和加工行为视角下的探讨[J]. 现代情报, 2022, 42(4): 62-76.
- [19] Jabalameli, S., Xu, Y. and Shetty, S. (2022) Spatial and Sentiment Analysis of Public Opinion toward COVID-19 Pandemic Using Twitter Data: At the Early Stage of Vaccination. *International Journal of Disaster Risk Reduction*, **80**, Article 103204. <https://doi.org/10.1016/j.ijdrr.2022.103204>
- [20] Wang, J., Zhou, Y., Zhang, W., Evans, R. and Zhu, C. (2020) Concerns Expressed by Chinese Social Media Users during the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data. *Journal of Medical Internet Research*, **22**, e22152. <https://doi.org/10.2196/22152>
- [21] 高芳芳, 林心婕. 围绕在线问诊的舆论焦点与网络情绪研究: 基于微博舆情的分析[J]. 未来传播, 2022, 29(2): 32-40.
- [22] Pal, S., Biswas, B., Gupta, R., Kumar, A. and Gupta, S. (2023) Exploring the Factors That Affect User Experience in Mobile-Health Applications: A Text-Mining and Machine-Learning Approach. *Journal of Business Research*, **156**, Article 113484. <https://doi.org/10.1016/j.jbusres.2022.113484>
- [23] Krittanawong, C., Narasimhan, B., Hahn, J., Narasimhan, H., Jneid, H., Virani, S.S., et al. (2022) Individual Sentiments on Telehealth in the COVID-19 Era: Insights from Twitter. *Progress in Cardiovascular Diseases*, **71**, 100-102. <https://doi.org/10.1016/j.pcad.2022.04.011>
- [24] Arenas Gaitán, J. and Ramírez-Correa, P.E. (2023) COVID-19 and Telemedicine: A Netnography Approach. *Technological Forecasting and Social Change*, **190**, Article 122420. <https://doi.org/10.1016/j.techfore.2023.122420>
- [25] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.
- [26] 赵凯, 王鸿源. LDA 最优主题数选取方法研究: 以 CNKI 文献为例[J]. 统计与决策, 2020, 36(16): 175-179.
- [27] 李廷进. 基于主题模型的文本聚类研究与应用[D]: [硕士学位论文]. 太原: 山西大学, 2020.
- [28] Mikolov, T., Sutskever, I., Chen, K., et al. (2013) Distributed Representations of Words and Phrases and Their Compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, New York, 3-6 December 2012, 3111-3119.
- [29] Ma, J., Wang, L., Zhang, Y., Yuan, W. and Guo, W. (2023) An Integrated Latent Dirichlet Allocation and Word2vec Method for Generating the Topic Evolution of Mental Models from Global to Local. *Expert Systems with Applications*, **212**, Article 118695. <https://doi.org/10.1016/j.eswa.2022.118695>