doi:10.3772/j.issn.2095-915x.2026.02.006

基于在线健康社区患者评论的主题识别和情感分析研究

开放科学（资源服务）标识码（OSID）

高烨　李松　袁安琪

河北大学管理学院 保定 071002

摘要：[目的 / 意义] 为探究在线健康社区中患者评论数据背后的医患关系失谐机理，构建医患信任修复机制，提出一种基于在线健康社区患者评论主题识别和情感分析协同研究框架。[方法 / 过程] 首先构建逆主题频率优化 LDA 模型的 ITFLDA 模型用于挖掘初始主题特征词，然后引入初始主题词项间的正点互信息 PPMI 重新建立特征词筛选规则，实现患者评论主题分类；最后基于主题分类结果，设计 ALBERT-BiGRU 融合主题情感双重注意力机制的情感分析模型，聚焦细粒度情感判别特征，完成主题导向下的情感分类优化。[结果 / 结论] 通过对“39 健康网”平台患者评论数据开展实证研究，提炼出患者就诊时重点关注的主题维度：治疗效果、医术医德和诊疗态度，并深入挖掘出各主题维度下引发患者负面情绪的医患核心矛盾，据此提出针对性改进策略，为数字化医疗生态治理提供可操作性路径。

关键词：在线健康社区；主题识别；情感分析；医患关系

中图分类号：G203；G35

Research on Topic Identification and Sentiment Analysis Based on Online Health Communities Patient Comments

GAO Ye　LI Song　YUAN Anqi

School of Management, Hebei University, Baoding 071002, China

Abstract: [Objective/Significance] In order to explore the dissonance mechanism of doctor-patient relationship behind patient review data in online health communities and build a doctor-patient trust repair mechanism, a collaborative research framework based on topic recognition and sentiment analysis of patient comments in online health communities is proposed. [Methods/Processes] Firstly, the ITFLDA model that optimizes the LDA model using inverse topic frequency is constructed to mine the initial topic feature words, and then the positive point mutual information between the initial topic words is introduced to re-establish the feature word screening rules and achieved the topic classification of patient comments. Finally, based on the results

作者简介　高烨（1996-），博士研究生，主要研究方向为信息管理与数据挖掘；李松（1965-），通信作者，博士，教授，主要研究方向为系统工程理论与方法、信息管理与数据挖掘，E-mail：lees3432@163.com；袁安琪（1988-），博士研究生，主要研究方向为信息管理与数据挖掘。

引用格式　高烨 , 李松 , 袁安琪 . 基于在线健康社区患者评论的主题识别和情感分析研究 [J]. 情报工程 , 2026, 12(2): 74-86.

of topic classification, a sentiment analysis model integrating the dual attention mechanism of topic and sentiment with ALBERT-BiGRU is designed to focus on the fine-grained sentiment discrimination features, completing the optimization of sentiment classification under topic orientation. [Results/Conclusions] Through empirical research on the patient comments of "39 Health Network" platform, extract the thematic dimensions that patients focus on in the process of treatment: treatment effect, medical skill and medical ethics, and attitude towards diagnosis and treatment. The core contradictions between doctors and patients that cause negative emotions in patients under each thematic dimension are deeply explored. On this basis, targeted improvement strategies are proposed to provide an operational path for the governance of the digital medical ecosystem.

Keywords: Online Health Communities; Topic Recognition; Sentiment Analysis; Doctor-patient Relationship

引言

随着公众健康管理意识提升，在“互联网+医疗健康”等国家政策[1]的支持和引导下，在线健康社区（Online Health Communities，OHCs）已发展成为公众寻求医疗咨询和健康管理的新途径。OHCs是以互联网为媒介，通过整合健康信息为患者构建信息共享、情感支持和问答咨询的平台[2]。当前主流OHCs均开通患者评论功能，通过深入挖掘OHCs衍生的海量评论数据获取患者核心诉求及不同医疗场景下的情感特征，可为医疗服务质量优化提供改进方向，有助于缓和医患关系、增进医患互信，推动医疗行业良性发展。

本文以“39健康网”平台患者评论数据为样本，聚焦在线健康社区患者评论文本的主题识别和情感分析研究。在主题识别方面，LDA模型虽被广泛应用，但高频噪声词会干扰主题特征词提取；基于词项概率的主题表征方式也容易遗漏概率表现不太突出但与主题语义紧密相关的关键词项。为克服上述不足，提出ITFLDA-PPMI（Inverse Topic Frequency Latent Dirichlet Allocation-Positive Pointwise Mutual Information）模型挖掘患者重点关注的主题方向。通过构建逆主题频率优化LDA的ITFLDA（Inverse Topic Frequency Latent Dirichlet Allocation）模型减弱主题间高频噪声词的影响，并重新构建主题特征词筛选规则，不再以词项概率为唯一标准，而是基于正点互信息PPMI（Positive Pointwise Mutual Information）引入与主题词项语义紧密关联的相关词，增强主题词的语义连贯性，使主题刻画丰富细致且贴合语境，精准描述出患者就诊过程中关注的核心主题。

在情感分析方面，随着自然语言处理技术的蓬勃发展，BERT及其衍生模型、BiLSTM、BiGRU等模型在相关任务中表现出较强优势[3-5]。同时，情感表达往往与特定场景或话题密切相关，同一词汇在不同主题语境下的情感色彩可能存在差异，且文本中的某些关键词蕴含着丰富情感信息。基于此，构建ALBERT-BiGRU融合主题特征和全局情感特征协同引导的双重注意力机制模型用于分析不同主题下患者评论的情感倾向。首先结合ALBERT和BiGRU获取文本初始表征，再利用主题特征向量动态调整词级注意力分布，获得主题增强表征，同时引入全局情感特征辅助向量引导情感关键词获得更高的注意力权重，形成主题情感双增强的文本特征表示，最后通过Softmax

doi:10.3772/j.issn.2095-915x.2026.02.006

层判断患者评论情感倾向。双重注意力机制的协同配合，在精准定位主题语义场景的同时动态捕捉情感线索，提升情感分类的准确性。通过进一步对负面评论文本展开分析，深入洞察医患关系中存在的矛盾点，为改善医患关系、提供高质量的医疗服务给予充分的数据支撑与决策依据。

1 相关研究综述

1.1 主题识别

在线健康社区主题识别任务是指借助新型主题挖掘和文本分析技术，提炼和解析社区文本数据中的深层信息，对知识主题及其关联特征展开深入研究[6]。Wang 等[7]基于词义共现分析和 LDA 模型探讨在线健康社区对 COVID-19 的信息需求，归纳出症状、预防、检查和治疗四个主要主题。于本海等[6]借助 LDA 模型分析出患者在“百度痛风病吧”平台信息交流中聚焦的主题类型，并根据研究结论提出相关建议。成全等[8]为从用户生成内容中深度挖掘用户需求，提出融合病情特征的多任务用户需求识别模型（MUNI-DC），构建涵盖提问意图和提问实体的用户需求主题体系。

1.2 情感分析

情感分析本质上是一种剖析文本中隐含的情感语义信息，挖掘出文本表达的情感极性和强度特征的分析手段[9]。评估在线健康社区中患者情感是维护稳定医患关系的关键环节，能够深入了解患者的情绪状态和需求，促进医患间的有效沟通和信任。Rodrigues 等[10]提出 SentiHealth-Cancer 的情绪分析工具，专门用于检测癌症患者在线社区中患者的积极、消极或中立情感。熊回香等[11]设计用户评价服务模型，选择 BiLSTM 模型对目标医生服务特征的情感极性进行量化评估。张伟等[12]综合运用 WoBERT Plus、CNN 和 BiLSTM 模型获取语义特征，提出情感增强、知识融合策略和剪枝依存句法树对“甜蜜家园”文本数据进行情感分析研究。

1.3 主题情感协同分析

主题与情感协同分析日益成为在线健康社区患者评论研究的重要方向，部分学者[13-15]先利用 LDA 等主题识别技术从大规模患者评论数据中获取高频讨论主题，继而运用情感词典或 BiLSTM 等方法分析不同主题下评论的情感倾向；而王辉等[16]调整研究流程，先基于情感词典筛选负面患者评论，再利用 BERTopic 模型进行主题分析。虽然两者目标均为明晰患者对主题的情感态度，并深入解析负面情绪的根源，但未能充分利用主题和情感之间相互影响的协同作用。为此，余佳琪等[17]构建主题情感混合模型 dUTSU，旨在挖掘主题与情感间的相互影响关系，并引入时序特征揭示主题情感演化过程，但存在患者类型单一且未整合深度学习方法的局限。

针对现有研究存在的问题，本文充分考虑情感关键词的语义权重及其在不同主题语境下的情感差异，将主题识别过程中挖掘的主题信息用于辅助情感分类，并设计主题情感双重注意力机制增强主题语义与情感表达之间的关联，实现在线健康社区患者评论主题与情感的深度协同分析。

2 研究方法

本文构建了一种针对在线健康社区患者评论的主题识别和情感分析研究框架，如图1所示，主要分为两个部分：（1）构建基于ITFLDA-PPMI的主题识别方法；（2）设计ALBERT-BiGRU融合主题情感双重注意力机制的情感分析方法。

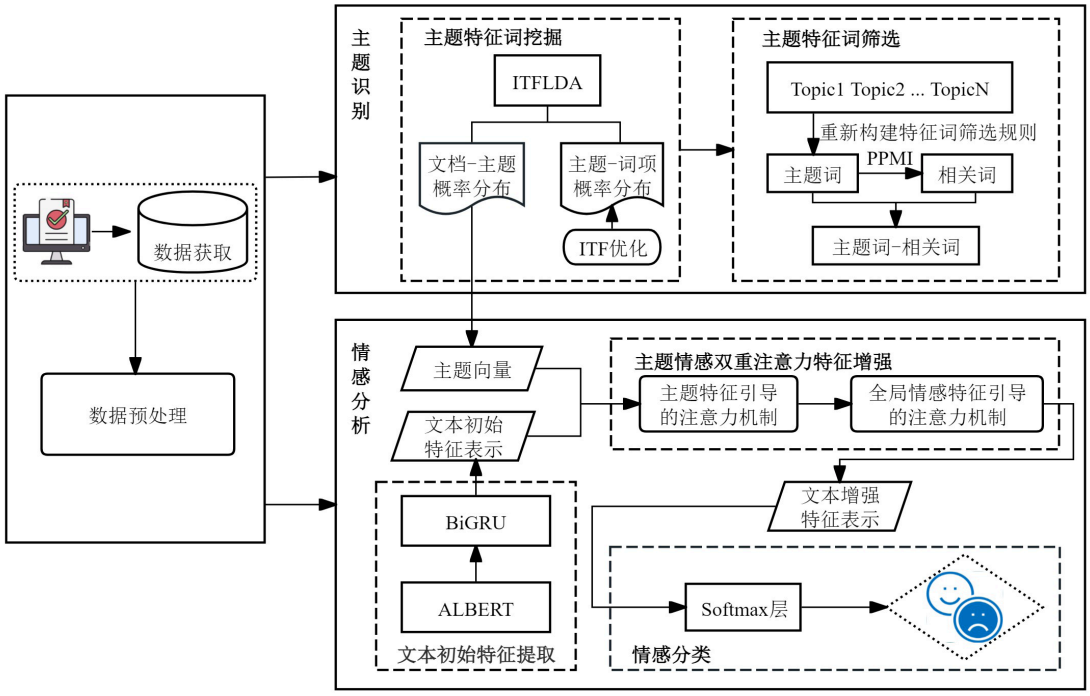

图1 在线健康社区患者评论主题识别和情感分析研究框架

2.1 基于ITFLDA-PPMI的主题识别方法

2.1.1 使用逆主题频率优化LDA模型

LDA[18]是典型的词袋模型，广泛应用于话题提取领域，通常采用Gibbs采样求解“文档-主题”分布和“主题-词项”分布。为改善LDA模型提取的特征词容易向高频噪声词偏移的缺陷，增强特征词对主题的区分程度，本文提出ITFLDA模型，其核心思想是引入逆主题频率（Inverse Topic Frequency，ITF）[19]优化LDA模型的主题-词项概率分布，ITF用于量化特征词与主题的相关性，消除跨主题高频词对主题边界的干扰。具体而言，在主题 k 的候选词集中选取前 n 个高概率词项计算 ITF 值，将词项 t 在主题 k 上的分布概率 $\varphi_{k,t}$ 与对应的逆主题频率 ITF_t 相乘，得到优化后的分布概率 $\varphi^*_{k,t}$，具体公式为

$$ITF_t = \log(1+\frac{C}{C_t}) \quad (1)$$

$$\phi^*_{k,t} = \phi_{k,t} \times ITF_t \quad (2)$$

式中，C 表示主题总数，C_t 代表包含词项 t 的主题数量。

2.1.2 引入PPMI值筛选主题特征词

LDA模型基于词共现挖掘主题的方式难以捕捉语义关联，且依赖主题词概率最大原则筛选特征词，易遗漏低概率强关联词项，导致主题表征不完整。因此，本文使用正点互信息（PPMI）[20]进行语义关联挖掘，重新构建主题特征词的筛选规则。PPMI是将点互信息

doi:10.3772/j.issn.2095-915x.2026.02.006

（Pointwise Mutual Information，PMI）[21] 中的负值置为 0，仅保留正相关性衡量词语间语义关联程度的指标，公式为

$$PMI(w_i,w_j)=\log\frac{p(w_i,w_j)}{p(w_i)p(w_j)},i\neq j \quad (3)$$

$$PPMI(w_i,w_j)=\max(PMI(w_i,w_j),0),i\neq j \quad (4)$$

式中，$p(w_i)$ 和 $p(w_j)$ 分别表示词 w_i 和 w_j 的边缘概率，$p(w_i, w_j)$ 表示联合概率，$PPMI(w_i, w_j)$ 越大，两个词的语义联系越紧密。

重新构建的主题特征词筛选规则以概率最大原则为基础，获得初始主题词项，然后引入主题词项间的 PPMI 值，识别与初始主题词项具有较强语义关联的相关词项，并将其作为主题描述的补充内容。通过该规则能够筛选出概率表现不突出但与主题语义相关的词项，使主题描述更加简洁清晰，具体过程如下。

（1）按照概率最大原则，基于 ITFLDA 模型结果选择前 m 个词作为初始主题词项；

（2）计算初始主题词项间的 PPMI 值，得到对应的初始相关词项；

（3）选取各主题概率位于前 $m/3$ 的词，得到最终主题词项和相关词项，共同描述主题。

2.2 基于ALBERT-BiGRU融合主题情感双重注意力机制的情感分析方法

本文构建基于 ALBERT-BiGRU 融合主题情感双重注意力机制的情感分析方法，实现患者评论文本的情感分类任务。该方法由文本初始特征提取模块、主题情感双重注意力特征增强模块和情感分类模块构成，具体结构如图 2 所示。

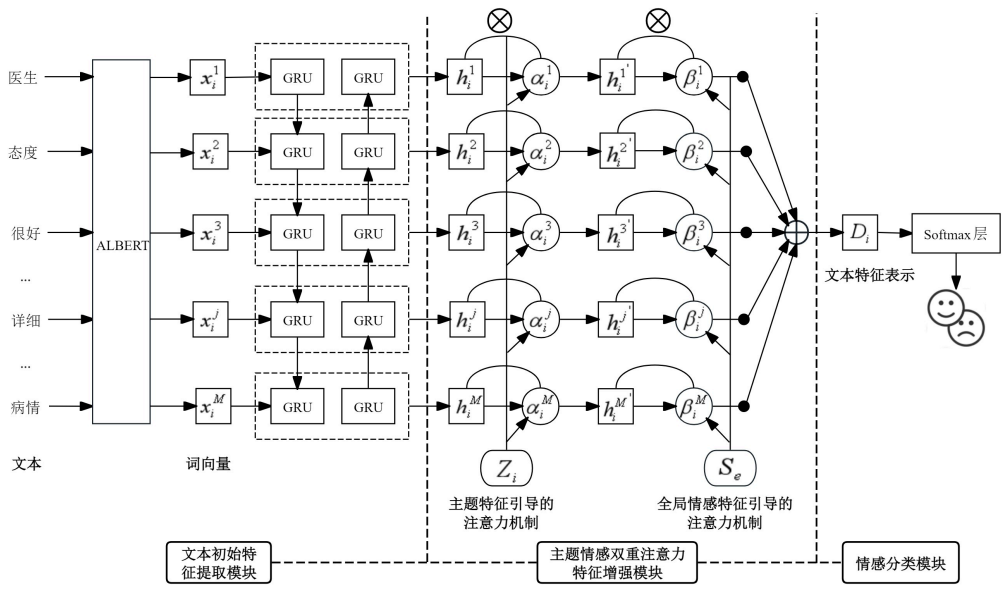

图 2 基于 ALBERT-BiGRU 融合主题情感双重注意力机制的情感分析方法

2.2.1 文本初始特征提取模块

本文结合 ALBERT 和 BiGRU 模型提取评论文本的初始特征。ALBERT[22] 是在 BERT 模型基础上提出的轻量化改进模型，通过采用 Embedding 层参数因式分解和跨层参数共享机制，大幅减少模型的参数量。双向门控循环单元（BiGRU）[23] 由正向和反向两个 GRU 单元构成，每个 GRU 通过重置门和更新门控制信息

流动。双向循环结构使模型同时捕捉序列数据的双向依赖关系，从而获得更完整的上下文信息。将在线健康社区评论集合中第 i 条评论文本中的第 j 个单词 t_i^j 输入到预训练模型 ALBERT 中获得对应单词的特征向量表示，表达式为

$$x_i^j = \mathrm{ALBERT}(t_i^j), j \in [1, M], x_i^j \in R^d \quad (5)$$

式中，M 是单个评论文本序列长度，d 是词向量的维度。为了进一步捕捉文本上下文信息，接着将 x_i^j 输入 BiGRU 模型中，具体计算过程为

$$\overrightarrow{h_i^j} = \mathrm{GRU}(x_i^j, \overrightarrow{h_i^{j-1}}), j \in [1, M] \quad (6)$$

$$\overleftarrow{h_i^j} = \mathrm{GRU}(x_i^j, \overleftarrow{h_i^{j-1}}), j \in [1, M] \quad (7)$$

$$h_i^j = \overrightarrow{h_i^j} \oplus \overleftarrow{h_i^j}, j \in [1, M] \quad (8)$$

式中，时序位置 j 的隐藏状态 h_i^j 由前向传播特征向量 $\overrightarrow{h_i^j}$ 和后向传播特征向量 $\overleftarrow{h_i^j}$ 共同构成；⊕表示向量的拼接操作；h_i^j 为第 i 条评论文本中第 j 个单词的初始特征向量。

2.2.2 主题情感双重注意力特征增强模块

情感表达具有语境依赖性，词语情感色彩会随着情境或主题动态转变，因此本文对在线健康社区患者评论进行情感分析时融入主题信息的影响。首先通过 ITFLDA 模型获得每条评论文本的主题分布，并将其编码为 K 维的概率向量（K 为主题个数），向量值表示评论对应主题的概率；然后设计主题特征引导的注意力机制，将每条评论文本表示为句子特征矩阵 $h_i = \{h_i^j \mid h_i^j \in R^d, j = 1, 2, ..., M\}$，每列向量与对应的主题特征向量 Z_i 相拼接，结果输入单层全连接层，经非线性激活变换得到中间特征 r^i；再利用 Softmax 函数计算主题特征引导的注意力权重 α_i，具体计算过程为

$$r^i = \tanh(W(h_i \oplus Z_i) + b) \quad (9)$$

$$\alpha_i = soft\max(Wr^i + b), \alpha_i \in R^M \quad (10)$$

式中，W 表示权重矩阵，b 表示偏置项，⊕表示拼接操作。α_i 包含句子中每个单词的注意力权重 α_i^j，基于该注意力权重分布，得到与主题特征相关的句子特征矩阵 $h_i^{'}$，其表达式为

$$h_i^{j'} = \alpha_i^j h_i^j \quad (11)$$

$$h_i^{'} = \{h_i^{j'} \mid h_i^{j'} \in R^d, j = 1, 2, ..., M\} \quad (12)$$

文本中的情感信息往往与特定单词紧密相关，应着重强化这些词在情感分类中的作用，为此引入全局情感特征辅助向量 S_e。S_e 经过训练为文本中的关键情感词赋予较大的注意力权重，得到全局情感特征引导的注意力权重分布 β_i^j，具体计算过程为

$$r_i^{'} = \tanh(Wh_i^{j'} + b) \quad (13)$$

$$\beta_i^j = soft\max(r_i^{'\mathrm{T}} S_e) \quad (14)$$

最后基于该权重分布，生成经主题情感双重注意力机制增强后的文本特征表示 D_i，如公式（15）所示，从而为后续的文本情感分类提供更具判别性且富含情感语义的特征输入。

$$D_i = \sum_j \beta_i^j h_i^{j'} \quad (15)$$

2.2.3 情感分类模块

为实现情感分类任务，将主题情感双重注意力机制增强后的特征表示 D_i 通过 softmax 层执行分类操作。Softmax 函数将输入向量映射至概率空间，生成数值在 [0,1] 的概率分布，将最大概率值对应的情感标签判定为患者评论文本的情感倾向，计算公式为

$$\hat{y}_i = soft\max(WD_i + b), \hat{y}_i \in R^c \quad (16)$$

式中，W 和 b 表示可训练的权重矩阵，$\hat{y}_i$ 表示第 i 条评论文本的情感概率分布，c 表示情感类别的数量。

doi:10.3772/j.issn.2095-915x.2026.02.006

3 实验与结果分析

3.1 数据采集与预处理

实验共使用两个数据集：

第一个数据集采用 SMP2020-EWECT 中的“疫情”主题数据集，与在线健康社区患者评论同属于公共卫生事件情境下，因此选择在该数据集下开展情感分类模型的训练与测试工作。将积极情绪赋值为 1，合并愤怒、悲伤、恐惧和惊奇为消极情绪并赋值为 –1，无情绪类不参与模型训练。

第二个数据集是从“39 健康网”爬取的患者评论数据。从全国各省份中随机选取多所三甲医院，因北京、上海、广东、湖南、四川和陕西等省份医疗综合实力较强，增加医院选取数目，最终确定医院总数为 53 所，时间跨度为 2014 年 1 月 1 日—2024 年 12 月 1 日，共爬取 11207 条患者评论。

在数据预处理阶段，需剔除重复评论文本，进行中文分词和去停用词。为提高医疗文本处理的适配性，本文基于四川大学机器智能实验室停用词库、哈工大停用词表、中文停用词表和百度停用词表等多方权威词库构建高覆盖度停用词表，并在去除重复词项的基础上补充医患交互场景中的噪声词（如医患称谓、医疗机构名称等）。

3.2 患者评论主题识别

3.2.1 确定最佳主题数

本文使用困惑度（Perplexity）和一致性分数（Coherence）双指标协同优化主题数选择问题，二者的最小比值对应的主题数即为最优主题数。计算 LDA 模型在主题数为 1~25 时的比值，如图 3 所示。当主题个数为 3 时比值最小，因此确定最佳主题数为 3。

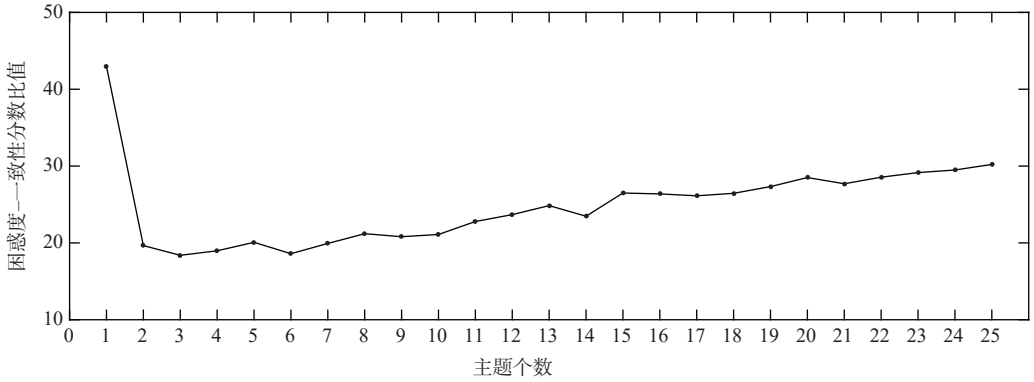

图 3 困惑度 – 一致性分数比值

3.2.2 主题识别结果

实验首先使用 ITFLDA 模型对“39 健康网”患者评论数据进行主题特征词挖掘，初步选择概率最大的前 60 个单词作为主题的特征词，部分结果如表 1 所示。

然后根据重构的特征词筛选规则，计算初始主题词项间的 PPMI 值，得到各词项对应的最相关主题词项，如表 2— 表 4 所示。表 2— 表 4 按照主题 – 词项概率排序，对比发现仅按照概率大小选出 20 个主题词项会忽略排序靠后但能够丰富主题语义内涵的词，结合相关词列表可进一步挖掘出这类词项。

表 1 初始主题词项分布

序号	Topic1		Topic2		Topic3	
	词项	概率	词项	概率	词项	概率
1	治疗	0.034203	手术	0.044493	病人	0.036434
2	好转	0.009991	医术	0.026586	态度	0.028433
3	孩子	0.009754	病人	0.015598	耐心	0.027104
4	效果	0.009319	医德	0.013476	医术	0.014604
5	朋友	0.008673	高超	0.013457	患者	0.013581
6	病情	0.008193	高尚	0.012535	详细	0.013445
...	...	...	...	...	...	...
20	发现	0.004102	和蔼	0.006375	满意	0.006437
...	...	...	...	...	...	...
40	好多	0.002700	治疗	0.003294	谢谢	0.003600
...	...	...	...	...	...	...
60	负责	0.001307	非常感谢	0.001586	希望	0.001997

表 2 主题 1 对应的主题词 – 相关词

Topic1							
T- 排序	主题词	相关词	R- 排序	T- 排序	主题词	相关词	R- 排序
1	治疗	方案	52	11	诊治	用心	54
2	好转	慢慢	31	12	感谢	真的	22
3	孩子	情况	59	13	找到	介绍	7
4	效果	不错	29	14	网上	找到	13
5	朋友	介绍	7	15	来到	仁爱	32
6	病情	控制	10	16	检查	发现	20
7	介绍	朋友	5	17	希望	康复	25
8	专业	负责	60	18	推荐	朋友	5
9	恢复	慢慢	31	19	很快	康复	25
10	控制	很快	19	20	发现	扩散	47

表 3 主题 2 对应的主题词 – 相关词

Topic2							
T- 排序	主题词	相关词	R- 排序	T- 排序	主题词	相关词	R- 排序
1	手术	成功	11	11	成功	手术	1
2	医术	高明	24	12	谢谢	天伦	31
3	病人	认真负责	41	13	术后	成功	11
4	医德	高尚	6	14	耐心	细致	37
5	高超	医术	2	15	诊治	小孩	17
6	高尚	医德	4	16	报答	真心	47
7	态度	和蔼	20	17	小孩	诊治	15
8	精湛	医术	2	18	感谢	真心	47
9	病患	责任心	45	19	检查	仔细	51
10	医治	诊断	44	20	和蔼	态度	7

doi:10.3772/j.issn.2095-915x.2026.02.006

表 4 主题 3 对应的主题词 – 相关词

Topic3							
T- 排序	主题词	相关词	R- 排序	T- 排序	主题词	相关词	R- 排序
1	病人	着想	25	11	询问	详细	6
2	态度	和蔼	22	12	预约	下次	31
3	耐心	解答	10	13	治疗	方案	38
4	医术	高明	39	14	仔细	询问	11
5	患者	着想	25	15	看病	仔细	14
6	详细	讲解	9	16	解释	详细	6
7	病情	询问	11	17	热情	服务	32
8	专业	负责	28	18	关心	病人	1
9	讲解	详细	6	19	热心	服务	32
10	解答	耐心	3	20	满意	下次	31

最后绘制 3 个主题下的词云图，如图 4 所示。综合词云图特点和主题词 – 相关词列表描述主题，从多角度更精准把握主题涵盖的内容和意义。通过分析将 3 个主题分别定义为：治疗效果、医术医德和诊疗态度，表 5 展示了各主题的详细内容及评论数据的分布情况。

Topic1 Topic2 Topic3

图 4 不同主题下的词云图

表 5 患者评论主题内容及分布情况

主题分类	详细内容	文档数	占比
治疗效果	患者经过治疗后病情的好转和康复情况	3126	27.89%
医术医德	医生的专业能力和道德素养	3728	33.26%
诊疗态度	在患者诊疗过程中，医护人员对患者的态度是否热情、和蔼，与患者讲述病情是否耐心、细致	4353	38.84%

由表 5 可知，诊疗效果、医术医德和诊疗态度是患者就医时的关注重点。其中治疗效果相关评论 3126 份，占比 27.89%，聚焦治疗后病情的好转和康复情况，尽管作为患者就医的核心诉求但并非关注首位。医术医德相关评论 3728 份，占比 33.26%，涉及对医生诊断、制定诊疗方案的专业能力及医德责任感的评判。诊疗态度相关评论 4353 份，占比最多，高达 38.84%，主要反映医护人员在接待患者、沟通病情和解释治疗方案过程中的服务表现。

结合临床实际场景和数据综合分析诊疗态度占比最大的原因：（1）从挂号、问诊到沟通

方案等就诊全环节中，医护态度直接影响患者即时体验；（2）部分危急重症患者紧急送医时，家属已对病情发展程度及可能出现的不良后果有心理预期，因此更需要从医护态度中获取安慰和支持。

3.3 患者评论情感分析

3.3.1 不同模型实验结果对比分析

在重新归类后的“疫情”数据集上，结合五折交叉验证方法，将本文模型与 BERT 模型、ALBERT 模型、BiLSTM 模型、BiGRU 模型和 ALBERT-BiGRU 模型进行对比实验，具体结果如表 6 所示。

表 6 模型对比结果

Model	精准率	召回率	F1
BERT	88.51	87.36	87.93
ALBERT	87.33	86.62	86.97
BiLSTM	86.24	86.01	86.12
BiGRU	86.78	86.43	86.60
ALBERT-BiGRU	88.82	87.68	88.24
ALBERT-BiGRU + 主题情感双重注意力机制	89.57	89.25	89.41

由表 6 实验结果对比可知，本文所提模型在精准率、召回率和 F1 值上均显著优于基线模型。BERT 作为经典预训练语言模型，能够有效捕获词语的上下文语义关联及长距离依赖特征；ALBERT 模型通过参数共享等轻量化设计降低计算成本，但语义建模能力略低于 BERT；基于传统循环神经网络架构的 BiLSTM 和 BiGRU 模型，在捕捉长距离语义依赖上存在局限，整体性能弱于基于 Transformer 的预训练模型；ALBERT-BiGRU 模型融合了预训练语言模型的全局语义表征能力和循环神经网络的序列建模优势使性能提升；进一步引入主题情感双重注意力机制，结合主题信息和全局情感特征强化情感极性判别，最终使模型的综合指标提升至 89.41%，充分表明本文模型在情感分析任务中的有效性。

3.3.2 患者评论情感分析结果

按照主题识别结果将“39 健康网”平台患者评论语料划分为 3 个主题子集，运用训练好的模型对每条患者评论进行情感二元分类（积极 / 消极），不同主题下患者消极情感占比情况如图 5 所示。为深入分析负面评价的生成原因，进一步生成各主题下负面评论的高频词云图，如图 6 所示。

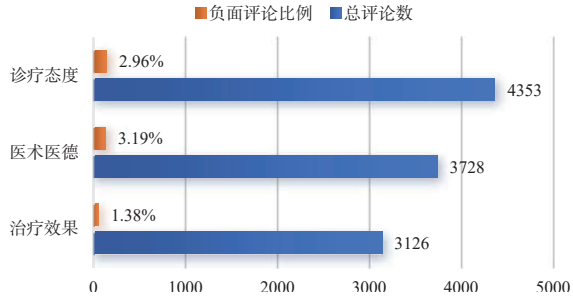

图 5 各主题下的患者负面情感评论占比

尽管各主题中的负面评论在整体评论中占比均低于 5%，但这部分信息能够反映医患关系紧张的底层原因，对改善医患关系具有重要参考价值。结合词云图进行分析，梳理归纳出各主题下医患之间存在的核心矛盾，具体内容如表 7 所示。

在治疗效果方面，“病情”“疼痛”“加重”等具象化症状词汇构成词云图的主体区域，“没什么”“好转”与“治疗”“效果”形成语义冲突，均显示出患者治疗后的恢复状态与预期之间的落差，这一结果揭示出医疗服务中的核心矛盾：

doi:10.3772/j.issn.2095-915x.2026.02.006

图 6 各主题负面评论高频词云图

表 7 各主题下医患间存在的核心矛盾

主题	医患核心矛盾
治疗效果	当患者经过治疗后，病情并未得到有效控制甚至恶化，引发对诊疗方案的质疑。
医术医德	专业能力有待提升，破坏公平秩序安排熟人插队、过度医疗小病大治、推荐天价药
诊疗态度	在诊疗过程中缺乏耐心且态度恶劣。

患者经过治疗后，病情并未得到有效控制甚至恶化，引发对诊疗方案的质疑。

在医术医德方面，“老百姓”一词说明发表负面评价的患者群体多为普通民众，更具有代表性。“素质”“人品”“差劲”等词体现患者对部分医生道德素质的负面评判，结合“安排”“药店”等词分析可知，部分医生存在安排熟人插队、推荐天价药等不当行为。而“专业”“不负责任”“不顾”“生命”等词反映患者对医生专业水平的质疑，认为部分医生存在小病大治、忽视患者生命健康等问题。这些问题揭示出医疗服务中存在的核心矛盾：部分医生专业能力有待提升，且存在破坏公平秩序、过度医疗和推荐天价药等行为，严重影响患者就医体验和医患信任。

在诊疗态度方面，“态度”“不耐烦”“恶劣”“说话”“难听”“傲慢”等词表明部分医生在诊疗过程中对待患者态度恶劣，语气傲慢，缺乏耐心。“脾气”“暴躁”“询问”“打断”等词说明部分医生会阻止患者表达疑问，使其无法完整阐述病情，难以获取充足的医疗信息，这一系列不当行为严重干扰医患之间的正常互动，对患者的心理状态造成负面影响。这些问题揭示出医疗服务中存在的核心矛盾：部分医生在诊疗过程中缺乏耐心且态度恶劣。

3.3.3 研究建议

综合上述研究，为改善医患关系、提升医疗服务质量提供决策参考，提出以下建议：

（1）医院建立治疗效果动态评估机制，通过阶段性疗效反馈缓解患者焦虑，同时加强个性化治疗方案设计，降低标准化流程与个体差异间的适配误差。

（2）医院建立就诊秩序监督机制，安排专门工作人员在候诊区域巡查，监督就诊秩序。同时制定明确的紧急情况判断标准，对于确实存在紧急病情的患者优先安排就诊，并做好记录与解释工作。

（3）医院建立分级诊疗负面清单制度，明确界定过度医疗行为；同时提升医疗服务透明

度，在门诊服务区域设置药品价格公示屏，实时展示常用药品的价格区间及医保报销比例等信息。

（4）医生积极参与线上线下培训课程，及时了解领域内的新技术新方法。定期回顾诊断困难、病情复杂或治疗效果不佳的病例，潜心研究病情解决方案。同时要具备良好的情绪管理能力，耐心倾听患者阐述病情，关注患者情绪，尤其是对于癌症等重病患者及时给予安慰和支持。

4 结论

本文以在线健康社区患者评论文本为研究对象，聚焦医患关系紧张的潜在成因，构建主题识别与情感分析的协同研究框架。在主题识别层面，创新性地提出 ITFLDA-PPMI 主题建模方法，通过引入逆主题频率 ITF 优化主题词项分布权重，有效抑制高频噪声词对主题表征的干扰；同时结合正点互信息 PPMI 重新构建特征词的筛选规则，精准挖掘出患者在就医过程中关注治疗效果、医术医德和诊疗态度 3 个核心主题。在情感分析层面，设计 ALBERT-Bi-GRU 融合主题情感双重注意力机制的深度学习模型，主题特征引导的注意力机制旨在利用主题信息定位情感表达的具体场景，避免理解歧义；全局情感特征引导的注意力机制旨在捕捉场景内的关键情感词，防止主题常见倾向的干扰；双重注意力机制的协同作用显著增强了情感分类效果。进一步从主题维度对负面评论进行可视化分析，追溯各主题下触发患者负面情绪的医患关系核心矛盾，并据此提出医患关系改善策略与医疗服务质量提升方案，为数字化医疗环境下的医患关系治理提供理论支撑与实践路径。

本文所提方法具有可推广性，能够迁移应用于不同研究领域的数据集。如在电商平台商品评论分析中，本文主题识别方法能挖掘出消费者关注的商品质量、物流服务等核心主题，并结合情感分析模型评估消费者态度，为商家优化服务提供参考；在社交媒体热点事件评论分析中，可识别公众焦点话题及情感态度，助力相关部门把握舆论走向。需要注意当前各类评论不再局限于单一文本模态，图片和表情以直观视觉传递情绪，未来研究需采用多模态视角融合文本语义和图像视觉特征，通过跨模态语义关联深入解码情感表达内涵，为用户行为研究注入多维解析能力。

参 考 文 献

[1] 国务院办公厅 . 关于促进“互联网 + 医疗健康”发展的意见 [EB/OL]. (2018-04-28) [2025-05-21]. https://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm.

[2] 陈曦 . 互联网医疗研究现状及未来展望 [J]. 人民论坛学术前沿 , 2017(24): 40-47.

[3] ABAS A R, ELHENAWY I, ZIDAN M, et al. BERT-CNN: a deep learning model for detecting emotions from text[J]. Computers, Materials and Continua, 2021, 71(2): 2943-2961.

[4] CHRIQUI A, YAHAV I. HeBERT and HebEMO: a hebrew BERT model and a tool for polarity analysis and emotion recognition[J]. INFORMS Journal on Data Science, 2022, 1(1): 81-95.

[5] 任楚岚 , 仇全涛 , 劣思敏 . 融合 TCN 和 BiLSTM 的文本情感分析 [J]. 计算机工程与设计 , 2024, 45(7): 2090-2096.

doi:10.3772/j.issn.2095-915x.2026.02.006

[6] 于本海 , 卢畅 . 在线健康社区信息主题特征及其潜在价值研究 —— 基于 LDA 模型对 " 百度痛风病吧 " 案例的分析 [J]. 价格理论与实践 , 2022(3): 195-198, 206.

[7] WANG J, WANG L, XU J, et al. Information needs mining of COVID-19 in Chinese online health communities[J]. Big Data Research, 2021(24): 100193.

[8] 成全 , 林颖茹 . 融合病情特征的在线健康社区用户需求识别模型研究 [J]. 情报理论与实践 , 2025, 48(4): 125-134.

[9] 王克勤 , 高智姣 , 乔亚楠 , 等 . 在线评论中的用户需求识别及其演化趋势挖掘 [J]. 机械科学与技术 , 2023, 42(7): 1070-1080.

[10] RODRIGUES R G, DAS DORES R M, CAMILO-JUNIOR C G, et al. SentiHealth-cancer: a sentiment analysis tool to help detecting mood of patients in online social networks[J]. International Journal of Medical Informatics, 2016, 85(1): 80-95.

[11] 熊回香 , 向瀛泓 , 孟璇 , 等 . 在线健康社区中基于用户评论挖掘的医生服务评价研究 [J]. 情报科学 , 2024, 42(8): 118-125.

[12] 张伟 , 徐宗煌 , 蔡鸿宇 , 等 . 基于情感增强和知识融合的在线健康社区情感分析研究 [J]. 数据分析与知识发现 , 2024, 8(3): 53-62.

[13] 叶艳 , 吴鹏 , 周知 , 等 . 基于 LDA-BiLSTM 模型的在线医疗服务质量识别研究 [J]. 情报理论与实践 , 2022, 45(8): 178-183, 168.

[14] 侯畅 , 李海晨 . 基于主题挖掘与情感分析的在线健康咨询评论研究 [J]. 情报探索 , 2023(6): 48-54.

[15] 郭羽婷 , 姚宣合 . 基于主题挖掘和情感分析的在线健康社区用户评论研究 [J]. 现代情报 , 2025, 45(8): 135-145.

[16] 王辉 , 王晓玉 , 李卫东 , 等 . 在线健康社区重大慢病患者负面评论倾向的关键影响因素分析 [J]. 情报科学 , 2024, 42(6): 12-20, 28.

[17] 余佳琪 , 赵豆豆 , 刘蕤 . 在线健康社区慢性病患者评论主题情感协同挖掘研究 —— 以甜蜜家园为例 [J]. 数据分析与知识发现 , 2023, 7(10): 95-108.

[18] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003(3): 993-1022.

[19] 高烨 . 基于热度高峰趋势的网络舆情事件衍生话题提取方法研究 [D]. 保定 : 河北大学 , 2022.

[20] ALEXANDRE S, ALINE V. Understanding the effects of negative (and positive) pointwise mutual information on word vectors[J]. Journal of Experimental & Theoretical Artificial Intelligence, 2023, 35(8): 1161-1199.

[21] TSCHANNEN M, DJOLONGA J, RUBENSTEIN P K, et al. On mutual information maximization for representation learning[C]// International Conference on Learning Representations. 2020: 1-16.

[22] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: A lite BERT for self-supervised learning of language representations[C]// Proceedings of the 8th International Conference on Learning Representations, 2020: 1-17.

[23] ZHANG T, JIA J, CHEN C. BiGRU-SA: Protein S-sulfenylation sites prediction based on BiGRU and self-attention[J]. Computers in Biology and Medicine, 2023(163): 107145.

（责任编辑：何彦青）