网络社群中照护者对早产儿养育的需求和关注点：基于网络爬虫的文本挖掘

江蓉，金颖，许莉莉，肖云霞
浙江大学医学院附属妇产科医院，浙江杭州　310006

摘　要：目的　采用文本挖掘技术探讨网络社群中照护者对早产儿养育的关注焦点和情感态度。**方法**　使用 Python 编程语言构建网络爬虫系统抓取微博和小红书平台中“早产儿养育”相关内容，采用隐含狄利克雷分布主题模型挖掘潜在主题，并利用 ROSTCM 6.0 软件进行高频词统计和情感分析，以揭示网络社群中照护者对早产儿养育的需求、关注热点及情感倾向。**结果**　共抓取到 8 180 条评论，主题模型识别出 7 个主题，分别为“日常护理”“喂养与营养”“发育与成长”“医疗与健康”“家庭环境与支持”“父母心身调适”“教育与启蒙”；照护者对“喂养”“护理”等词提及度较高，在一定程度上展现了用户感受和关注重点；网络社群中照护者积极情绪占 51.25%，中性情绪占 29.16%，消极情绪占 19.59%，网络社群中照护者消极情绪主要集中在担心孩子的生长发育及远期发展等方面。**结论**　网络社群中照护者对早产儿养育的关注集中在喂养管理、发育监测等细节，同时存在焦虑情绪。研究结果为构建基于证据的早产儿家庭照护知识体系提供了数据支持，建议通过精准健康宣教和心理支持提升照护质量与早产儿生存质量。

关键词：早产儿；养育；需求；文本挖掘；网络；照护者　　DOI：10.3969/j.issn.1671-9875.2026.06.009

中图分类号：R473.72　　**文献标识码：**A　　**文章编号：**1671-9875(2026)06-0050-05

随着国家政策支持、医学技术进步及社会经济文化发展，早产儿生存率显著提升[1-2]。但因其体重低、胎龄小、各系统发育不完善，部分早产儿需要反复进行有创治疗，易并发呼吸暂停、生长迟缓、神经认知障碍等问题[3]，死亡风险与长期健康隐患依然严峻，需精细化护理干预。在互联网高速发展背景下，微博、小红书等社交媒体已成为公众分享育儿经验的重要平台，用户生成内容（user-generated content，UGC），真实反映了网络社群中照护者的养育需求与情感表达。文本挖掘技术作为高效的数据处理手段，利用自然语言处理和机器学习等技术将无序的非结构化文本信息转化成一系列有序的结构化信息[4-6]。网络爬虫技术通过模拟用户行为自动抓取公开网络数据，突破了传统数据收集在时空上的限制，在护理研究领域展现出独特的应用价值[7-9]。本研究采用网络爬虫技术结合隐含狄利克雷分布（Latent Dirichlet Allocation，LDA）主题模型、高频词共现分析及情感极性分析等文本挖掘方法，对微博、小红书两大社交媒体平台中“早产儿养育”相关 UGC 进行深度挖掘，旨在揭示网络社群中照护者照护知识需求特征与情感倾向，为构建早产儿家庭照护知识体系提供数据支持。

1　资料与方法

1.1　文本数据抓取

本研究基于微博、小红书两大社交媒体平台，以“早产儿养育”为核心关键词采集数据，并用 Python 编程语言构建网络爬虫系统，通过模拟用户行为实现目标页面的自动化遍历与数据提取[10]。数据采集时间范围为 2020 年 3 月至 2025 年 3 月，爬取范围涵盖话题页、用户发帖及评论区。为尽可能确保数据来源于真实照护者，采用多阶段筛选策略。首先，保留包含具体养育经历、问题或经验的文本；其次，结合用户简介、历史发帖主题（如持续聚焦育儿、早产儿相关内容）及互动语境进行辅助判断；最后，人工复核剔除明显的广告、转载及无关内容。经去重和清洗后，最终获得有效数据为 8 180 条评论。

1.2　数据预处理

数据预处理流程包括以下步骤：文本清洗，运用正则表达式匹配方法，即按照特定语法规则找出并大批量删除文本中存在的一些表情、超链接、标点符号和空白符，统一格式，去除无关“噪声”；分词处理，使用 jieba 分词工具（开源的中文自然语言处理库），将语句切分成单独的词，构成词袋模型，实现对中文文本的颗粒度细粒化处理；停用词

作者简介：江蓉（1994－），女，本科，硕士在读，主管护师
通信作者：金颖，jinyin@zju.edu.cn
收稿日期：2025－06－09
基金项目：浙江省医药卫生科技计划项目，编号 2023KY127

过滤，基于自然语言工具包自带停用词表及领域专用词库，去除“的”“了”“在”等无实际语义的虚词，保留有效词项。

1.3 数据分析方法

1.3.1 主题分析

本研究采用 LDA 主题模型[11]对预处理后的数据进行语义挖掘。LDA 主题模型基于“文档由多个潜在主题混合而成”的假设，通过学习大量文本数据，自动挖掘文本隐含的主题结构，揭示文本集合的语义内涵。具体实施步骤如下：特征工程，以 Python 的 scikit-learn 为基础搭建 LDA 主题模型分析框架。Scikit-learn 是机器学习领域的开源工具包，包含丰富、易用的特征提取以及模型构建相关的代码实现。词频-逆文档频率（term frequency-inverse document frequency，TF-IDF）是从海量文本数据中抽取高频率词项来描述文本特征的方法，本研究基于 TF-IDF 算法提取高频特征词，生成词项-I 文档矩阵，为模型训练提供数据基础。参数优化，采用困惑度与主题一致性进行参数优化，其中，困惑度评估模型对未见数据（训练时未使用过的测试集数据）的预测能力（值越小，泛化能力越强），主题一致性通过计算主题内关键词的语义相关性评估主题质量（值越高，词语聚集度越好）。针对 LDA 主题模型的核心参数主题数 K（$2 \leqslant K \leqslant 15$），采用网格搜索法遍历取值区间，并通过绘图库绘制困惑度-主题数曲线。主题数 K 直接影响主题划分的颗粒度，过小易导致语义重叠，过大则会产生语义碎片化。最终确定主题数 $K=7$ 为最优值，该取值满足困惑度极小且主题一致性得分高的条件。主题命名：结合高频关键词与临床实践经验，对 7 个主题进行命名。

1.3.2 基于 ROSTCM 6.0 软件的高频词分析和情感分析

为揭示网络社群中照护者的 UGC 语义特征及情感倾向，本研究采用 ROSTCM 6.0 软件进行数据可视化分析。高频词提取：将预处理后的文本数据导入 ROSTCM 6.0，采用 TF-IDF 算法提取核心关键词。该算法利用词项频率与逆文档频率相乘的办法去除通用词汇，选取有意义的高频词，得到包含 129 个关键词的词集。词云可视化：使用微词云在线工具生成动态词云图。该工具作为专业的文本可视化软件，能以字体大小和颜色深浅直观呈现高频词。通过设置字体、配色方案、布局算法等参数，清晰展现高频词的分布特征与语义关联。情感极性分析：基于 ROSTCM 6.0 内置的情感词典（含条情感词及程度副词），经过一般性处理后，得出情感分布结果，并进一步分离其中的积极情绪（含褒义词汇或正面评价）、中性情绪（陈述事实，无情感倾向）和消极情绪（含贬义词汇或负面评价）。

2 结果

2.1 主题分析结果

LDA 主题模型将文本数据分为 7 个主题，各主题及主题词见表 1。

表 1 “早产儿养育”相关文本的 7 个主题及主题词

主题	主题词
日常护理	护理，睡眠，洗澡，衣服，清洁，皮肤，温度，体温，消毒
喂养与营养	喂养，母乳，奶粉，辅食，喂奶，营养，消化，蛋白，食物，维生素，强化剂
发育与成长	发育，成长，身高，体重，运动，智力，语言，翻身，精细动作
医疗与健康	医院，医生，疫苗，住院，感染，检查，治疗，随访，干预，体检，疾病
家庭环境与支持	家庭，家里，回家，环境，沟通，交流
父母心身调适	父母，家长，妈妈，爸爸，焦虑，担心，情绪，缓解，着急，安抚，快乐
教育与启蒙	早教，学习，游戏，育儿，科学

2.2 基于 ROSTCM 6.0 软件的高频词分析

通过 ROSTCM 6.0 软件进行高频词分析，词云图见图 1。图 1 显示，由词云图形式对词频排列在前 100 的词语进行展示，可以直观看到网络社群中照护者在话题及帖子中提及的高频词汇。其中“喂养”（9 007 次）、“护理”（9 544 次）、“体重”（5 387 次）等词高频出现，提示照护者对喂养方式、营养摄入及生长发育监测的重视；“妈妈”（7 910 次）、“焦虑”（1 932 次）等词体现照护者面临的心理压力与情感困扰；“消毒”（1 030 次）、“环境”（1 246 次）、“睡眠”（1 594 次）、“按摩”（1 374 次）等词提示照护者关注居家护理细节。词云中字体大小代表词频高低，字体越大，表示词频越高，词频的高低能直接反映用户感知的高低，在一定程度上展现了网络社群中照护者对早产儿养育关注的重点，直观展现了“喂养困难”“追赶生长”“神经发育”等核心议题的突出地位。

图 1　高频词词云图

2.3　基于 ROSTCM 6.0 软件的情感分析结果

本研究通过 ROSTCM 6.0 软件内置的情感词典对 4 482 条有效评论进行情感极性分析，见表 2。结果显示：积极情绪占比 51.25%（2 297/4 482），主要集中在“喂养成功”“体重增长”“发育进步”等正向反馈，如“宝宝现在纠正胎龄 3 个月，体重已达 12 斤”“母乳强化剂效果显著，宝宝追赶生长明显”等表述体现了照护者对喂养方案的认可；中性情绪占比 29.16%（1 307/4 482），多为客观描述性内容，如“每天按时喂养 1 次/3 h”“每周记录生长曲线”等，反映出网络社群中照护者对科学育儿流程的遵循；消极情绪占比 19.59%（878/4 482），主要涉及“生长发育迟缓”“喂养困难”“健康焦虑”等议题，高频负面词包括“肌张力低”“不会翻身”“追赶失败”等。进一步分析显示：一般消极情绪占比 13.88%（622/4 482），表现为对喂养细节的担忧，如“每次喂奶耗时 1 h，宝宝体重增长缓慢”；中度消极情绪占比 5.02%（225/4 482），涉及对医疗干预的焦虑，如“眼底筛查未通过，需二次手术”“肺炎反复发作，担心影响肺功能”；高度消极情绪占比 0.69%（31/4 482），集中于对长期预后的恐惧，如“医生怀疑脑瘫倾向，康复训练效果未知”“早产儿免疫力低下，频繁感染住院”。情感分布与高频词分析结果存在关联，“发育”“护理”等代表照护实践与发育监测的高频词在积极情绪文本中出现频率较高，而“焦虑”“担心”等表达负面情绪与担忧的词汇在消极情绪文本中出现频率也较高。研究结果提示，网络社群中照护者对早产儿养育的情感态度呈现“关注-焦虑”特征，需通过科学且有针对性的健康教育来缓解网络社群中照护者的心理压力。

3　讨论

3.1　早产儿养育需求的多维特征分析

社交网络的普及革新了信息传播模式，UGC 因其真实性、即时性和海量性特点，成为公众健康认知研究的重要数据来源，其在健康传播领域的优势已获验证[12]。本研究通过 LDA 主题模型识别出 7 个主题，包括日常护理、喂养与营养、发育与成长、医疗与健康、家庭环境与支持、父母心身调适、教育与启蒙等领域，与国内外早产儿护理指南的核心内容高度契合。WHO 建议：早产儿或低出生体重婴儿的喂养策略、发育监测和感染预防是护理核心要素[13]。本研究发现，“喂养与营养”主题词频次较高（如喂养、母乳、奶粉等），这与国内一项最佳证据总结中“母乳喂养为早产儿首选方式，需根据胎龄、体重及并发症等个体差异选择捐赠人乳、早产儿配方奶或母乳强化剂”[14]的结论一致。网络社群中照护者对“辅食添加”的关注早于传统认知的早产儿纠正胎龄 4～6 个月，提示网络社群中照护者存在过早引入固体食物的潜在风险，需进行标准化营养教育干预。在“发育与成长”主题中，“身高”“体重”“运动”等主题词表明追赶生长是早产儿家庭的核心关注点。有研究[15]表

表 2　“早产儿养育”相关文本情感分析（n=4 482）

情绪分类	分析结果		一　般		中　度		高　度	
	数量/条	百分比	数量/条	百分比	数量/条	百分比	数量/条	百分比
积极情绪	2 297	51.25	1 313	29.29	573	12.78	411	9.17
中性情绪	1 307	29.16	—	—	—	—	—	—
消极情绪	878	19.59	622	13.88	225	5.02	31	0.69

注：中性情绪由于文本为客观事实陈述，缺乏情感强度梯度，故不作程度分析。

明，早产儿发育迟缓和神经发育异常的发生率显著高于足月儿。本研究中“翻身”“精细动作”等主题词的出现，反映出网络社群中照护者对早产儿发育与成长的焦虑，与早产儿发育监测的指南[16]一致，即通过出院随访、定期评估、早期干预及多学科协作优化长期健康结局。

3.2 早产儿养育需求关注的热点分析

高频词“消毒”“环境”“睡眠”“按摩”等说明网络社群中照护者重视居家感染防控和护理操作细节，这与 WHO 的建议[13]一致。“维生素”“强化剂”等高频词及主题词的出现，可能反映部分家庭因母乳不足或喂养困难而依赖营养补充剂，需警惕营养失衡风险。“家庭环境与支持”主题中“沟通”“交流”等词的出现，揭示家庭系统内部协作的重要性。“回家”“家里”等主题词的出现，提示早产儿出院后适应期是家庭支持的关键阶段，这与国内多学科团队联合家庭提供支持护理的研究[17]结果一致。

3.3 情感特征与心理干预需求

情感分析显示，消极情绪占比 19.59%，集中于生长发育迟缓、喂养困难、健康焦虑等，这一结果与国外研究[18]一致，说明早产儿家庭面临更高的心理压力。本研究中“肌张力低”“追赶失败”等专业术语的出现，反映网络社群中照护者通过网络自主学习获取信息，但可能因医学知识不足导致判断偏差。因此，需建立权威信息平台，提供基于循证医学的发育评估工具以帮助早产儿家庭科学识别异常信号。本研究还发现高度消极情绪(0.69%)主要与脑瘫倾向、反复感染等严重并发症相关，与国内早产儿脑保护相关研究[19]的发现一致。针对这一群体，应建立完整的“医疗-心理-社会”三位一体支持体系，如利用正念减压疗法[20]等措施缓解照护者焦虑。

4 结论

本研究发现，喂养管理、发育监测及神经认知发展是网络社群中照护者关注的焦点，同时焦虑情绪存在表明网络社群中照护者需要社会支持。本研究将网络爬虫技术应用于早产儿养育领域，突破传统问卷调查的时空限制，深入剖析网络社群中照护者对早产儿养育的中心需求及情感特征。这些结果可为构建模块化早产儿健康教育课程、开发早产儿发育追踪小程序和建立网络社群中照护者情感识别系统等提供依据。本研究存在一定局限性：首先，数据来源于微博、小红书两大社交平台，其用户群体可能存在年龄、性别、代际及社会经济地位等方面的偏差，且用户的表达意愿、表演性呈现及内容选择性可能影响数据的真实性；其次，平台的内容分发受商业推广和推荐算法影响，可能使某些议题被放大或遮蔽，导致数据存在平台特性引发的选择偏倚；此外，情感分析依赖预设词典，难以完全捕捉中文语境下的语义细微差别；主题命名受研究者主观经验影响，后续可采用德尔菲法进行专家论证以提升命名的客观性。未来研究可结合质性访谈，深入探究照护者决策过程，同时扩大样本覆盖范围，纳入更多元化的平台和渠道，关注不同社会经济与文化背景家庭的需求差异，以增强研究结果的普适性。

参考文献：

[1] 中华医学会妇产科学分会产科学组，胡娅莉，杨慧霞. 早产临床防治指南(2024 版)[J]. 中华妇产科杂志，2024，59(4)：257-269.

[2] 蒋思远，杨传忠，田秀英，等. 中国新生儿协作网出生胎龄 22～25 周超早产儿出院预后及治疗现状[J]. 中华儿科杂志，2024，62(1)：22-28.

[3] WILLIAMSON M，POORUN R，HARTLEY C. Apnoea of prematurity and neurodevelopmental outcomes：current understanding and future prospects for research[J]. Frontiers in Pediatrics，2021，9：755677.

[4] ALEQABIE H J，SFOQ M S，ALBEER R A，et al. A review of textmining techniques：trends，and applications in various domains[J]. Iraqi Journal for Computer Science and Mathematics，2024，5(1)：125-141.

[5] 李颖，周亮，陈娴，等. 基于文本挖掘的爆炸伤延时现场救护的技术需求分析[J]. 军事护理，2024，41(4)：66-70.

[6] AVASTHI S，CHAUHAN R，ACHARJA D P. Techniques，applications，and issues in mining large-scale text databases[J]. Journal of Data Mining and Knowledge Discovery，2021，35(4)：567-582.

[7] FENG G. Implementation of web data mining technology based on python[J]. Journal of Physics：Conference Series，2021，2066(1)：012033.

[8] WU H，HOU D. A focused event crawler with temporal intent[J]. Applied Sciences，2023，13(7)：4149.

[9] 苑秋辰，麻盛淼，沈晨，等. Python 语言在护理研究领域的应用现状与展望[J]. 中国护理管理，2023，23(5)：796-800.

[10] 郭婺，郭建，张劲松，等. 基于 Python 的网络爬虫的设计与实现[J]. 信息记录材料，2023，24(4)：159-162.

[11] 彭灿华，韦晓敏. 基于 LDA 主题模型的多数据库主题词挖掘算法[J]. 计算机仿真，2023，40(8)：483-487.

(下转第 57 页)

3 讨论

儿童因特殊的解剖、生理特点，不能正确表达疼痛等因素，其压力性损伤的发生特征、危险因素等与成人相比有显著不同[7]。骨科患儿在使用石膏、支具等外固定装置时，常因泡沫敷料不贴合皮肤，时常卷边、脱落，增加了患儿皮肤损伤的风险，患儿一旦发生压力性损伤，很容易导致外观受损、功能丧失，甚至发生骨骼暴露及感染，增加患儿痛苦[8]。在儿童医院骨科，各个年龄段患儿之间的体型差距大，临床可供选择的泡沫敷贴型号有限，在临床使用中，常因为泡沫敷贴型号不合适、生理解剖特点等原因，使得泡沫敷贴预防压力性损伤作用有限。本研究基于人体解剖学设计了翼形贴、沙漏贴、雪花贴、四叶草贴 4 种异形泡沫敷贴，分别用于患儿躯干骨隆突处、足跟、内外踝、肘部等，其中翼形贴、沙漏贴等通过增大接触面积、优化压力传递路径，有效减少施加在皮肤上的外力，从而完成压力的重新分布，雪花贴、四叶草贴通过多触点拆分压力、动态结构适配活动实现压力分散。因此异形泡沫敷贴起到良好的减压作用，能有效预防压力性损伤的发生，同时缓解压力性损伤严重程度。本研究结果显示，虽然试验组使用异形泡沫敷贴后压力性损伤发生率低于对照组，差异无统计学意义（$P>0.999$），但试验组的压力性损伤分期低于对照组，且试验组发生部位少于对照组。两组患儿压力性损伤发生率比较无统计学意义，可能原因在于观察时间仅为 2 周，时间不够长；本研究纳入的对象为无压力性损伤的骨科患儿，患儿基础条件相似，预防压力性损伤的效果相似；同时，评价指标为压力性损伤发生率，对患儿主观感受未作评价，可能存在一定的偏倚性。

4 结论

异形泡沫敷贴的设计适用于不同年龄段的患儿，在临床应用中有助于缓解压力性损伤的严重程度。但由于本研究仅对一家医院的患儿进行研究，存在单中心问题，同时该研究是历史性对照研究，可能存在一定偏倚，而且评价指标仅为压力性损伤发生率，对患儿主观感受评价不足，未来研究可以进行多中心、多评价指标、随机对照等研究，以进一步探讨异形泡沫敷贴在骨科患儿压力性损伤预防中的应用效果。

参考文献：

[1] 申屠晓艳，邹瞿超，吴琳珊. 压力性损伤客观诊断与量化评估技术的研究进展[J]. 护理与康复，2023，22(9)：94-97.

[2] 王琳，梁培荣，张素梅. 小儿骨科医疗器械相关压力性损伤的原因分析与预防对策[J]. 天津护理，2021，29(2)：189-191.

[3] 邵晓，陈晓春，袁婷，等. 糖尿病酮症酸中毒患儿救治期间皮肤损伤原因分析及护理对策[J]. 护理与康复，2021，20(7)：78-80.

[4] 王欣怡，宋俐，王萍等. 不同类型敷料对手术压疮预防效果的贝叶斯网状 Meta 分析[J]. 护士进修杂志，2018，33(11)：963-968.

[5] 宣炜嘉，谢珺，冯晓艳. 儿童压力性损伤风险评估量表对手术患儿压疮风险评估及价值分析[J]. 山西医药杂志，2023，52(17)：1351-1355.

[6] 杨桃，胡雯，程懿，等. 老年慢性阻塞性肺疾病患者入院营养标志物与压力性损伤的相关性研究[J]. 华西医学，2023，38(1)：28-33.

[7] 王震，张伟英. 小儿术中获得性压力性损伤危险因素及风险评估的研究进展[J]. 中国护理管理，2022，22(11)：1727-1731.

[8] 陈娜，赵华，王照红，等. 儿童医疗器械相关压力性损伤原因分析与预防措施[J]. 齐鲁护理杂志，2023，29(10)：157-159.

（上接第 53 页）

[12] 陈旖旎，周晓英，岳丽欣，等. 移动 UGC 社区用户健康信息采纳行为意愿的影响因素[J]. 图书情报知识，2022，39(5)：82-95.

[13] World Health Organization. WHO recommendations for care of the preterm or low-birth-weight infant[R/OL]. [2025-05-20]. https://www.who.int/publications/i/item/9789240058262.

[14] 董慧敏，张筱岚，胥琰梨，等. 早产儿营养管理的最佳证据总结[J]. 全科护理，2024，22(18)：3372-3377.

[15] CHEN Z，XIONG C，LIU H，et al. Impact of early term and late preterm birth on infants' neurodevelopment：evidence from a cohort study in Wuhan，China[J]. BMC Pediatrics，2022，22(1)：251.

[16] HEALTH W O A，NURSES N A O. Assessment and care of the late preterm infant evidence-based clinical practice guideline，third edition[J]. Journal of Obstetric，Gynecologic，and Neonatal Nursing：Jognn，2024，53(5)：e1-e62.

[17] 刘娅，倪志超，刘葵葵. 多学科团队照护联合家庭支持护理对新生儿科早产儿家属认知、焦虑和社会支持水平的影响[J]. 湖南师范大学学报(医学版)，2024，21(2)：154-158.

[18] KILIÇLI A，SARAÇOĞLU G，BÜYÜKBODUR A Ç. Stress，anxiety，and postpartum depression in parents with premature infants in neonatal intensive care unit[J]. Florence Nightingale Journal of Nursing，2023，31(2)：82-90.

[19] 石晶，母得志. 早产儿脑保护的研究进展[J]. 中国实用儿科杂志，2023，38(5)：359-364.

[20] 宁时安，姜梅，柳学华，等. 正念干预联合信息支持对早产儿父亲心理健康状况的影响[J]. 中华护理杂志，2023，58(12)：1442-1449.