

基于情感分析的抑郁倾向人群在线健康社区参与行为特征研究

王 哲 向 菲

(华中科技大学同济医学院医药卫生管理学院 武汉 430030)

〔摘要〕 基于主题挖掘和情感分析方法,分析不同情感倾向用户在线健康社区参与行为特征,提出在线健康社区可以为抑郁症患者提供在线社会支持,为平台开展抑郁情绪管理和干预提供理论依据,为抑郁症防治特色服务提供思路。

〔关键词〕 抑郁倾向;情感分析;隐含狄利克雷分布模型;在线健康社区

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.01.003

Study on Behavioral Characteristics of Participation in Online Health Community among People Prone to Depression Based on Sentiment Analysis WANG Zhe, XIANG Fei, School of Medicine and Health Management, Tongji Medical College of Huazhong University of Science and Technology, Wuhan 430030, China

〔Abstract〕 Based on the topic mining and sentiment analysis methods, the behavioral characteristics of online health community (OHC) participation of users with different emotional tendencies are analyzed, and it is proposed that OHC could provide online social support for patients with depression, so as to provide a theoretical basis for the platform to carry out depressive mood management and intervention, and provide ideas for the characteristic services of depression prevention and treatment.

〔Keywords〕 depressive tendency; sentiment analysis; latent Dirichlet allocation (LDA); online health community (OHC)

1 引言

为贯彻落实《健康中国行动(2019—2030年)》中关于心理健康促进行动有关要求,加大抑郁症防治工作力度,遏制患病率上升趋势,鼓励社会心理服务试点地区探索开展抑郁症防治特色服务,国家卫生健康委员会办公厅于2020年印发《探索抑郁症防治特色服务工作方案》^[1]。该方案明

确提出,要求各相关机构通过线上线下多种形式开展抑郁症筛查。由于高发病率和低龄化等趋势,抑郁症日益受到关注,抑郁症患者在线健康社区(online health community, OHC)蓬勃兴起,吸引了大量用户参与。OHC的价值实现正是以大量参与者为基础^[2],积极的用户参与是增强社区黏性、保持社区可持续发展的关键性要素^[3]。目前,国内学界对OHC参与的研究主要存在两种视角:信息行为视角和社会行为视角。前者包括在线健康信息搜寻、在线健康信息素养教育等行为^[4];而社会行为视角将信息行为视为参与行为的一部分,同时也包含诸如表达情感、建立关系甚至经济利益实现等社会行为^[5]。同时,抑郁倾向用户在线健康参与行为与正

〔修回日期〕 2022-10-31

〔作者简介〕 王哲,硕士研究生;通信作者:向菲,博士,副教授,发表论文45篇。

常用户存在一定差别^[6]，这种差别正是进行早期识别和筛查工作的基础。因此，本文基于社会行为视角，分析抑郁倾向用户在线健康社区参与行为特征，以期为平台开展抑郁情绪管理和干预提供理论依据，为抑郁症防治特色服务提供思路。

2 数据与方法

2.1 数据获取与预处理

“百度贴吧－抑郁症吧”是具有代表性的抑郁症患者 OHC。本研究以该贴吧内有发帖行为的用户为研究对象，抓取其在所有贴吧全部发帖及“楼层”回复。数据清洗之后得到 687 名用户发表在全部贴吧内的 6 179 条主题帖、102 472 条回复贴，其中发表在“抑郁症吧”的主题帖和回复帖 6 879 条。

2.2 研究方法

2.2.1 情感分析方法 情感分析是自然语言处理中最热门的研究领域之一^[7]。本文使用开源情感分析系统 Senta_ BiLSTM^[8]，对于输入的每条文本，系统可输出情感类别标签、积极情感打分、消极情感打分。将第 i 条文本积极情感打分记为 P_i 、消极情感打分记为 N_i ，两者分别赋权值 1、-1，将用户发文数量记为 n_1 。根据模型输出结果计算用户情感得分 S 。该得分可以在整体程度上反映用户 OHC 参与与情感倾向。

$$S = \frac{\sum [(1 * P_i) + (-1) * N_i]}{n_1}$$

(1)

2.2.2 文本主题识别 隐含狄利克雷分布模型 (latent Dirichlet allocation, LDA)，由 Blei D M 等^[9]于 2001 年提出。作为一种无监督学习算法，该模型简单有效，在多个领域的主题识别中得到广泛应用^[10]，已被证明具有较强的热点话题发现能

力^[11]。本研究综合《哈工大停用词表》《四川大学机器智能实验室停用词库》《百度停用词表》，经去重后构建停用词表，并在试验中对词典进行改进调整。

3 结果与分析

3.1 “抑郁症吧” 用户参与总体主题特征

对发表在“抑郁症吧”的文本进行主题识别，从整体上探究抑郁症吧用户参与的主题特征。困惑度是评价模型有效性的通用标准^[12]。根据困惑度曲线手肘法，选取拐点处 $K = 8$ 为最佳主题个数。根据主题分类结果，“抑郁症吧”用户关注主题除了疾病症状与治疗相关话题外，还包括用户情感表达与交流等相关话题，见表 1。

表 1 “抑郁症吧” 用户参与主题特征

主题描述	关键词
情感表达	喜欢 开心 情绪 希望 痛苦 讨厌 难过 自私
治疗经验分享	治疗 吃药 医生 心理 药物 大脑 效果 方法
在线情感支持	加油 谢谢 朋友 老师 小姐姐 理解 抱抱 倾诉
生活碎片分享	锻炼 可爱 冬天 机会 小姐姐 勇敢 可惜
现实陪伴需求	妈妈 病房 姐姐 病友 叔叔 浣熊 我爸 朋友
人生哲学思考	世界 孤独 害怕 学会 悲伤 生命 生活 善良
症状讨论	痛苦 身体 焦虑 难受 思维 改变 绝望 噩梦
住院治疗	病区 手机 护士 吃饭 出院 聊天 监控 病友

3.2 不同情感倾向用户参与行为特征

3.2.1 情感得分计算与情绪倾向分类 使用 Bi-LSTM 文本情感分类模型进行情感分析，计算用户单条发帖或回复的情感得分，见表 2。对用户情感得分设置阈值，将用户分为 3 类，见表 3。

表 2 部分情感得分示例

文本示例	积极情感得分	消极情感得分	文本情感得分
我丑 我有病 我爱哭 我矫情 我嘴硬 我难缠 我一无是处 我脑子有病	0.000 1	0.999 9	-0.999 8
不知道怎么和人相处，就想一直在家里，只有在熟悉的环境下我才能安心	0.477 6	0.522 4	-0.044 4
整个清晨都散发着一一种新的朝气	0.943 6	0.056 4	0.887 2
新年快乐。烟火向星辰，所愿皆成真，岁岁年年，万喜万般宜	0.999 2	0.000 8	0.998 4

表 3 用户情感倾向分类

研究对象	情感得分 <i>S</i>	用户数量 (个)	发帖及回复 数量 (条)
积极用户	$1 \geq S > 0.2$	73	3 132
中性用户	$0.2 \geq S > -0.2$	213	18 538
消极用户	$-0.2 \geq S > -1$	360	24 394

3.2.2 用户主要关注社群类型 分别计算 3 类用户在不同贴吧发文占比，见图 1。对 3 类用户参与的前 100 种贴吧进行分类并绘制雷达图，见图 2。除了抑郁类，消极用户还较多参与其他心理障碍和游戏类社群，中性用户较多参与文学类社群，积极用户则更多参与运动类、宗教类等。

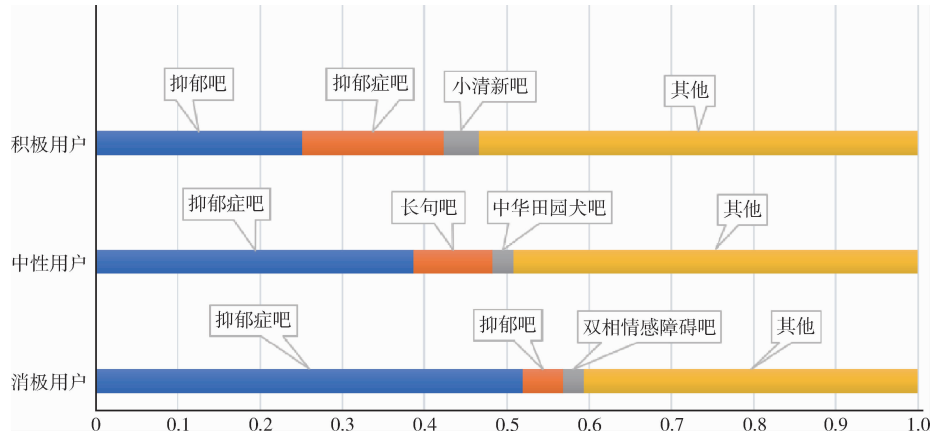


图 1 3 类用户关注社群构成比

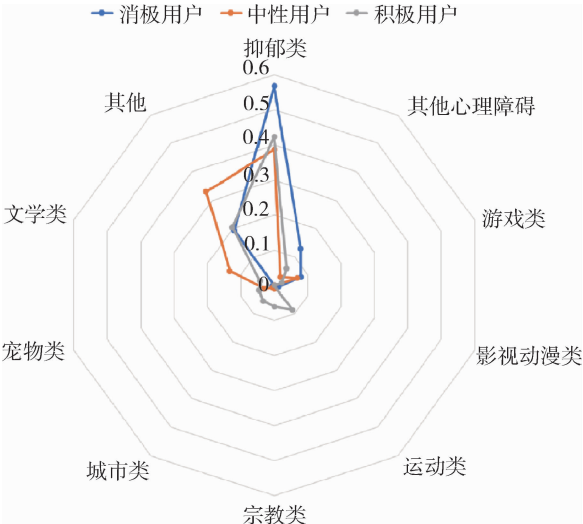


图 2 社群分类

表 4 用户关注主题总体特征

用户类别	主题描述
积极用户	期盼与热爱、陪伴与鼓励、倾诉与祝福、宗教、积极应对、治愈书籍分享、抑郁症治疗
中性用户	消极情绪倾诉、求助、抑郁症治疗、日常生活分享、倾诉、心理问题、祝福与期盼
消极用户	消极情绪倾诉、求助、抑郁症状、抑郁症治疗、求助、日常生活分享

积极用户在线上社区关注的主题多与“鼓励”“祝福”有关，呈现出积极性心态和“正能量”；中性用户既包含积极倾向的“期盼”“祝福”，还包括消极情绪发泄、求助和抑郁症疾病知识分享，主题特征较为复杂，但相较于消极用户来看，此类用户并无严重的消极情绪；消极用户关注主题均与抑郁症有关，见图 3。

3.2.3 用户关注主题特征 对 3 类用户全部文本内容分别使用 LDA 模型进行主题识别。通过困惑度曲线确定 3 类的最佳主题数量，见表 4。



图 3 不同情感倾向用户词云

3.2.4 用户活跃时间特征 将用户在 1 天内不同时段下的发文与回复数量 (x) 视为用户活跃度。首先将数据归一化处理:

$$x' = \frac{x - \text{Min}}{\text{Max} - \text{Min}} \tag{2}$$

数据处理后绘制 3 类用户活跃度随时间变化曲线,横坐标为时间,纵坐标为用户活跃度,见图 4。3 类用户活跃度随时间呈现明显变化。夜间(21~24 时)用户普遍较为活跃,白天活跃用户主要为消极用户和中性用户。凌晨时段(1~6 时)3 类用户都有活跃度明显下降至最低的过程,但也不难发现其中的差别:积极用户活跃度最早降至最低,中性用户其次,而消极用户在此期间均有相对较高的活跃程度,至凌晨 6 时降至最低,并在 7 时 3 类用户活跃度达到同一水平。

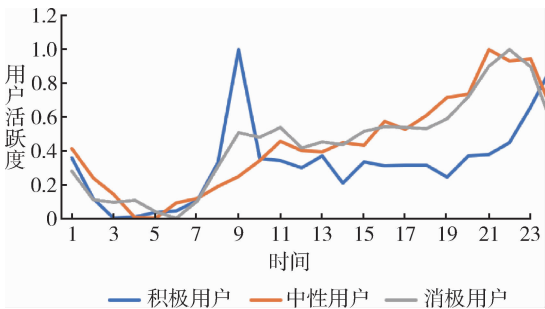


图 4 3 类用户活跃时间特征

3.2.5 情感得分与回复数量相关性分析 根据用户发帖数量 (n_2) 与每条发帖下的“楼层”回复数量 (m),计算用户平均“楼层”数量 L 如下,将其视为用户受关注程度。

$$L = \frac{\sum_{i=1}^{n_2} m_i}{n} \tag{3}$$

以用户情感得分 S 和平均“楼层”数量 L 为变

量进行相关性分析,结果显示,用户情感得分与回复数量相关性假设检验具有显著性,且 $r > 0$,说明两者之间存在正相关关系,见表 5。

表 5 相关性分析结果

变量	$L(P)$	$S(P)$
L	1.000 (0.000 ***)	0.107 (0.009 ***)
S	0.107 (0.009 ***)	1.000 (0.000 ***)

注: *** 表示 $P < 0.001$ 。

4 讨论

4.1 “抑郁症吧”用户总体特征

4.1.1 以消极用户为主 根据统计分析结果,“抑郁症吧”以消极用户为主,积极用户与中性用户占比不足 1/2。通过对发表在“抑郁症吧”的文本进行主题识别,并对聚类结果进一步归纳,发现该社区内用户参与行为以线上疾病治疗经验分享与情感表达为主,证明用户 OHC 参与的主要原因是获得在线社会支持^[13]。用户通过参与疾病治疗经验讨论,获得其所需的信息性帮助支持,有助于开展自我诊断与护理;通过情感表达与交流,可以感知来自社区内其他用户的共鸣、理解和鼓励等精神支持。

4.1.2 OHC 参与行为有助于缓解消极情绪 抑郁症患者往往伴随着消极情绪,由于社会对抑郁症的偏见以及疾病给个体自尊心带来的冲击,患者会具有一定程度的“病耻感”,对于抑郁症患者在正常社会交往中的情绪表达更为不利。而在线社区这一匿名化平台规避了污名化相关的负面后果,对于抑郁症患者而言,心理门槛更低、更易参与^[14-15]。

因此, 抑郁症患者 OHC 参与行为在消除孤独感、打破耻辱感障碍、促进自我赋权以及鼓励个人积极态度等方面是有效的^[16]。

4.2 不同情感倾向用户参与行为特征

4.2.1 积极用户的“利他性”行为 积极用户在社区内多扮演“支持提供者”角色, 参与行为有明显的“利他性”。用户关注话题与社群中出现“宗教”类, 国外有关研究表明, 各群体中宗教信仰与较低程度的抑郁症状呈显著正相关^[17-18]。同时, 积极患者较关注运动、宠物类社群, 该类用户有着较为正向、积极的生活态度。

4.2.2 中性用户的“自我疗愈”行为 中性用户以抑郁症相关话题讨论为主, 但其抑郁类社群参与在 3 类用户中占比最少, 其更多地参与文学、宠物类社群。从在线社区关注范围可以看出该类用户有着更为广泛的兴趣爱好。因此, 中性用户虽然存在一定的抑郁倾向或症状, 但其对于自身症状具有更强的认知, 同时积极尝试各种途径以试图缓解自身症状, 其在 OHC 参与中更倾向于“自我疗愈”行为。

4.2.3 消极用户“抑郁特征”明显 消极用户抑郁类贴吧参与度最高, 在其他心理障碍类社群参与中也具有较高占比。结合关注话题, 推测该类用户可能是程度较为严重的抑郁症或双相情感障碍患者, 或是具有较强抑郁倾向的潜在抑郁症患者。建议在线健康社区平台将此类用户设为重点关注对象, 并有针对性地开展抑郁情绪管理, 促进在线社区良好健康发展。

4.3 游戏类社群与抑郁倾向

本研究发现相比于积极用户, 消极用户和中性用户较多参与游戏类社群, 这也从在线社区参与角度证实了游戏成瘾可能是抑郁症患病的风险因素^[19-20]。因此, 建议相关游戏平台加强用户治理, 避免游戏成瘾, 进而遏制存在抑郁倾向玩家抑郁程度加深; 对 OHC 平台而言, 作为抑郁症“高发区”, 应加强对该类社区关注, 尽早识别, 倡导用户合理利用网络游戏, 采用正确方式进行自我心理

疏导, 培养良好的生活习惯。

4.4 用户活跃度特征

从活跃时间特征看, 积极用户受工作时间影响较大, 作息时间为规律。中性用户和消极用户在夜间活跃度较高, 说明伴有一定程度的睡眠障碍, 这是抑郁症最为常见的躯体化症状之一。3 类用户中, 消极用户最为依赖在社区中进行情感倾诉与情绪发泄, 活跃度全天普遍处于最高水平, 中性用户与消极用户活跃度接近, 这与 SAEB 等^[6]的研究结果基本一致, 手机使用时长和频率与抑郁程度之间呈正向相关。

4.5 用户受关注程度特征

用户情感得分与回复数量间相关性分析表明, OHC 中用户的情感倾向越积极, 所受到的关注度越高。

5 结语

本文对可能存在抑郁倾向人群的 OHC 参与行为进行多维度分析, 进一步明确其 OHC 参与行为特征。研究发现, 抑郁症吧用户关注主题存在共同之处, 说明 3 类用户均具有一定抑郁倾向, 区别在于严重程度不同, 不同程度抑郁倾向导致其关注主题、活跃度和受关注程度等方面存在明显差别, 后续研究与平台管理者可据此特征开展对抑郁症患者的早期在线识别和筛查。

总之, OHC 可以为抑郁症患者提供在线社会支持, 患者在此“抱团取暖”, 在一定程度上缓解抑郁情绪, 同时避免疾病污名化问题。社群种类繁多, 有些社群的参与可能会进一步加强抑郁症患者症状, 如游戏类社群; 同时虽然大部分社群已有成员审核制度, 但审核质量难以保证, 难免鱼龙混杂, 存在进一步加重抑郁症患者参与的隐患。因此, 通过优化制度规范加强对 OHC 的监管, 并依托人工智能等技术手段, 更加精准地提供特色化服务、开展疾病防治工作, 是促进 OHC 良好健康发展、开展抑郁症防治特色服务的必由之路。

参考文献

- 1 国家卫生健康委员会办公厅. 探索抑郁症防治特色服务工作方案 [EB/OL]. [2022-03-30]. http://www.gov.cn/zhengce/zhengceku/2020-09/11/content_5542555.htm.
- 2 SARASOHN - KAHN J. The wisdom of patients: health care meets online social media [EB/OL]. [2022-03-30]. https://www.researchgate.net/publication/242101734_The_Wisdom_of_Patients_Health_Care_Meets_Online_Social_Media.
- 3 张敏, 刘雪瑞, 张艳. 在线健康社区用户诊疗信息求助行为形成机理的概念模型——基于扎根理论的探索性研究 [J]. 情报科学, 2019, 37 (4): 22-28.
- 4 张帅, 王文韬, 李晶, 等. 国外用户在线健康信息行为研究进展 [J]. 图书馆论坛, 2018, 38 (9): 138-147.
- 5 张鑫. 在线健康社区用户参与行为的类型及偏好研究 [J]. 情报资料工作, 2019, 40 (5): 84-91.
- 6 SAEB S, ZHANG M, KARR C J, et al. Mobile phone sensor correlates of depressive symptom severity in daily - life behavior: an exploratory study [J]. Journal of medical internet research, 2015, 17 (7): e175.
- 7 YUAN J H, WU Y, LU X, et al. Recent advances in deep learning based sentiment analysis [J]. Science China - technological sciences, 2020, 63 (10): 1947-1970.
- 8 TIAN H, GAO C, XIAO X Y, et al. SKEP: sentiment knowledge enhanced pre - training for sentiment analysis [C]. Electr Network: The 58th annual meeting of the association - for - computational - linguistics (ACL), 2020.
- 9 BLEI D, NG A, JORDAN M. Latent Dirichlet allocation [C]. Vancouver: Advances in neural information processing systems 14, neural information processing systems: natural and synthetic, NIPS, 2001.
- 10 JENSEN S, LIU X Z, YU Y Y, et al. Generation of topic evolution trees from heterogeneous bibliographic networks [J]. Journal of informetrics, 2016, 10 (2): 606-621.
- 11 刘干, 林杰豪, 翟雯熠. 基于中心词和 LDA 的微博热点话题发现研究 [J]. 情报杂志, 2021, 40 (5): 143-148, 164.
- 12 KLAKEOW D, PETERS J. Testing the correlation of word error rate and perplexity [J]. Speech communication, 2002, 38 (1): 19-28.
- 13 秦敏, 李若男. 在线用户社区用户贡献行为形成机制研究: 在线社会支持和自我决定理论视角 [J]. 管理评论, 2020, 32 (9): 168-181.
- 14 KENDAL S, KIRK S, ELVEY R, et al. How a moderated online discussion forum facilitates support for young people with eating disorders [J]. Health expectations, 2017, 20 (1): 98-111.
- 15 RAINS S A. The implications of stigma and anonymity for self - disclosure in health blogs [J]. Health communication, 2014, 29 (1): 23-31.
- 16 JOHNSTON A C, WORRELL J L, DI GANGI P M, et al. Online health communities an assessment of the influence of participation on patient empowerment outcomes [J]. Information technology & people, 2013, 26 (2): 213-235.
- 17 MANN J R, MCKEOWN R E, BACON J, et al. Religiosity, spirituality, and depressive symptoms in pregnant women [J]. The international journal of psychiatry in medicine, 2007, 37 (3): 301-313.
- 18 GWIN S, BRANSCUM P, TAYLOR L, et al. Associations between depressive symptoms and religiosity in young adults [J]. Journal of religion and health, 2020, 59 (6): 3193-3210.
- 19 KIM D J, KIM K, LEE H - W, et al. Internet game addiction, depression, and escape from negative emotions in adulthood: anationwide community sample of Korea [J]. The journal of nervous and mental disease, 2017, 205 (7): 568-573.
- 20 鲍成臻, 高丽丽, 韩历丽. 中学生网络沉溺现状及其与抑郁水平的关系 [J]. 中国儿童保健杂志, 2021, 29 (1): 33-36.

欢迎订阅

欢迎赐稿