·定性系统综述·

机器学习在网络社交平台自杀预测领域的研究进展

王焓骁[1,2],康艾嘉[1,3],赵玉宝[4],赵福容[1],蒋晓江[5],郝凤仪[1,6*],唐向东[6]

(1. 重庆两江新区第一人民医院,重庆 401120;
2. 中国人民大学,北京 100872;
3. 多伦多大学,加拿大多伦多 M5S 2E8;
4. 杭州安肯医疗科技有限公司,浙江 杭州 311121;
5. 陆军特色医学中心,重庆 400042;
6. 四川大学华西医院睡眠医学中心,四川 成都 610041
*通信作者:郝凤仪,E-mail:fengyihao@cqljrmyy.com)

【摘要】 本文针对机器学习在网络社交平台自杀预测领域的相关成果进行系统综述,为群体及个体自杀预测提供参考。本文将从机器学习在多个平台自杀预测的现状(博客与轻博客、熟人社交平台、论坛、图片与视频社交平台、临床数据库)和局限性(算法准确性和效率、隐私泄露、污名化问题)等方面进行阐述。

【关键词】 机器学习;自杀;预测;综述

开放科学(资源服务)标识码(OSID):

微信扫描二维码
听独家语音释文
与作者在线交流

中图分类号:R749 文献标识码:A doi:10.11886/scjsws20210521001

Advances in machine learning in suicide prediction on online social platforms

Wang Hanxiao[1,2], *Kang Aijia*[1,3], *Zhao Yubao*[4], *Zhao Furong*[1], *Jiang Xiaojiang*[5], *Hao Fengyi*[1,6*], *Tang Xiangdong*[6]

(1. *The First People's Hospital of Chongqing Liangjiang New Area*, *Chongqing* 401120, *China*;
2. *Renmin University of China*, *Beijing* 100872, *China*;
3. *University of Toronto*, *Toronto* M5S 2E8, *Canada*;
4. *Hangzhou Anken Medical Technology Co., Ltd.*, *Hangzhou* 311121, *China*;
5. *Army Medical Center of PLA*, *Chongqing* 400042, *China*;
6. *Sleep Medicine Center*, *West China Hospital*, *Sichuan University*, *Chengdu* 610041, *China*
**Corresponding author*: *Hao Fengyi*, *E-mail*: *fengyihao@cqljrmyy.com*)

【**Abstract**】 This article systematically reviews the research results related to the machine learning based suicide ideation prediction on social networking platforms, so as to provide references for group and individual suicide prediction. This article will address the current states (issues of algorithm accuracy and efficiency, privacy leakage and stigma) and limitations of machine learning based suicide prediction on different platforms (light blogging, acquaintance social platforms, forums, picture and video sharing applications and clinical databases).

【**Keywords**】 Machine learning; Suicide; Prediction; Review

自杀是个体蓄意或自愿采取各种手段结束自己生命的行为,自杀死亡占所有死亡人数的1.4%[1]。既往研究[2]阐述了潜在的自杀风险因素,包括心理健康状况、经济社会地位、文化和道德因素等,为自杀风险评估提供了理论模型,但传统统计分析方法处理复杂数据的能力有限,且研究同质性强,导致模型仅在较狭窄的限定范围内有意义[3]。且由于自杀意念的隐蔽性,既往自杀预测方法难以对高危人群做出准确的、主动的识别[4-5]。

机器学习是人工智能(Artificial Intelligence,AI)学科的重要分支,它使用计算机模拟人类学习过程,并通过不断适应新数据以优化算法,从而提高模型的预测准确性[6],是一类能从数据中自动分析并掌握规律,再利用规律对未知数据进行预测的方

基金项目:重庆市科技传播与普及专项(项目名称:慢性失眠症网络自助式整合心身干预技术的应用,项目编号:cstc2019kpzx-kphdA0014);重庆市科卫联合医学科研项目(项目名称:矛盾性失眠简易诊断指标筛选及远程心理治疗研究,项目编号:2021MSXM228)

法。与传统分析方法相比，机器学习能为给定数据集确定最有效的模型，并且更适合处理复杂数据[7]，但需要更大的数据集来构建预测模型。目前在自杀预测领域常用的算法有随机森林、支持向量机、神经网络、自然语言、深度学习等，均表现出良好潜力。近年来，自杀意念的表达不再局限于口头形式，通过电子手段（包括论坛、博客、轻博客、即时消息、电子邮件、私信等）表达痛苦和自杀意念的情况逐渐增多。青年人是网络平台的主要用户，也是自杀的高风险人群，网络平台的数据公开化为自杀预测的机器学习提供了数据来源。

1 资料与方法

1.1 资料来源与检索策略

1.1.1 资料来源

于2021年4月-5月对PubMed、中国知网、万方医学网的相关文献进行检索。检索时限为2016年1月1日-2020年12月31日。

1.1.2 检索策略

中文检索词："机器学习""人工智能""决策树""分类树""支持向量机""随机森林""神经网络""深度学习""自然语言"和"自杀"；中文检索式：（机器学习+人工智能+决策树+分类树+支持向量机+随机森林+神经网络+深度学习+自然语言）*（自杀）；英文检索词："Machine Learning""Artificial Intelligence""Decision Trees""Classification Trees""Support Vector Machines""Random Forests""Neural Network""Deep Learning""Natural Language""Suicide""Social media""Social Network""Facebook""Twitter""Reddit""Instagram""Snapchat""YouTube""Weibo""Forums"；英文检索式：((Machine Learning OR Artificial Intelligence OR Decision Trees OR Classification Trees OR Support Vector Machines OR Random Forests OR Neural Networks OR Deep Learning OR Natural Language) AND (Suicide) AND (Social media OR Social Network OR Facebook OR Twitter OR Reddit OR Instagram OR Snapchat OR YouTube OR Weibo OR Forums)) AND ((Machine Learning OR Artificial Intelligence OR Decision Trees OR Classification Trees OR Support Vector Machines OR Random Forests OR Neural Networks OR Deep Learning OR Natural Language) AND (Suicide) AND (Social media OR Social Network OR Facebook OR Twitter OR Reddit OR Instagram OR Snapchat OR YouTube OR Weibo OR Forums))。

1.2 文献纳入与排除标准

由三位作者共同制定文献的纳入与排除标准。纳入标准：①采用各类机器学习方法，从网络社交平台采集数据并预测自杀的研究；②具有代表性的关于基于机器学习的网络社交平台用户自杀预测的重要综述和原创研究性文献。排除标准：①重复的文献；②非中英文文献；③无法获取全文的文献。

1.3 文献筛选与质量评估

由两名研究者独立进行文献检索，在剔除重复文献后，由两名研究者阅读文献标题、摘要和全文，进行人工交叉复审；严格按照纳入和排除标准筛选文献。

2 结 果

2.1 纳入文献基本情况

初步检索共获取文献114篇，其中中文文献44篇，英文文献70篇。排除重复文献18篇，剩余96篇。再通过阅读文献标题、摘要及全文，排除60篇，最终纳入文献36篇。见图1。

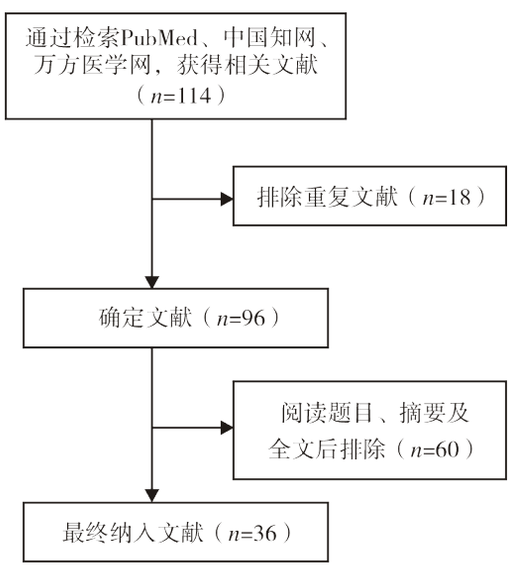

图1 文献筛选流程图

2.2 机器学习预测网络用户自杀行为

2.2.1 微博与轻博客是目前机器学习预测自杀的主战场

微博与轻博客因用户可自由匿名发言且信息公开，容易实现数据采集，为机器学习提供了海量

训练素材。在基于Twitter的研究中[8]，自杀预测的准确率为68%~92%，使用神经网络可探索与自杀相关的心理因素，包括负担、压力、孤独、绝望、失眠、抑郁和焦虑，并预测自杀行为发生风险较高时间。

在中国，基于微博的“树洞行动”以已故用户的微博账号下的留言为数据库，筛查具有情绪低落甚至包含自杀意念的信息。杨芳等[9]研究显示，留言用户主要集中在16~26岁年龄段，跳楼、割腕、烧炭等是高风险人群表达的主要自杀方式。留言用户在各时间段中负性情绪的表达均多于正性情绪，留言文本内容可概括为情绪倾诉、人际关系和社会支持、睡眠、死亡等方面[10]。章宣等[11]提出混合架构的神经网络模型，进一步提升了自杀风险的预测精度。庄婷婷等[12]研究表明，微博用户自杀敏感信息的发布具有周期规律，约50%的信息发布于23:00至次日05:00。许立鹏等[13]提供了较为完备的中国互联网用户“自杀词典”，以提高自杀意念模型的分类准确率。Cheng等[14]研究显示，高自杀风险者代词、前置词、多功能词的使用频率高，而动词使用频率较低，总字数较多。然而，由于网络信息的真实性问题，发言文本中信息的准确性仍需人工进一步甄别。

2.2.2 熟人社交平台Facebook已启动自杀审查与监测系统

个体在熟人社交平台暴露自杀意念可能意味着更迫切的求助与发泄需求。2017年，Facebook开始自动化监测自杀相关内容，利用网站即时消息界面与用户交流情绪和认知，洞察用户行为模式，此外，平台还包括情绪追踪、每日签到和心理教育等功能。如果监测到用户存在自杀风险，则会启动危机应对方案，包括向用户提供心理支持资源和危机干预热线，或提醒当地应急人员。Facebook正在扩大自动监测范围，以监视和删除包含敏感视频的帖子，防止自杀直播[15]。

在Facebook的自杀审查监测系统中，使用随机森林加上Deep Text（由Facebook发布，能够准确识别聊天内容）和线性回归是最有效的，机器学习在自杀表达上得到了更加精确的训练，使工作人员能够更好地区分自杀意念的讽刺表达和严肃表达，从而使模型更加健全和准确[16]。

2.2.3 机器学习可识别讨论论坛中的自杀内容

自杀意念的表达有时兼具抒情、澄清、告别和遗嘱的功能，这些内容被用户发表在相应的“社区”以引起共鸣。国外学者在讨论论坛Reddit进行了调查[17-18]，结果表明，使用自然语言处理，可识别用户的情绪困扰和自杀风险。Logistic回归和支持向量机分类器算法显示，在线帖子中的自杀内容监测准确率为80%~92%[19]。一些担心被污名化者，例如阿片类药物使用者也倾向于在论坛求助。过量使用阿片类药物是其常见的自杀手段，然而机器学习对该类人群的自杀风险识别具有较多假阳性结果[20]。在线心理健康论坛可以为心理痛苦者提供支持性网络环境，同时生成大量数据，可利用机器学习挖掘这些数据以预测其心理健康状态[21]。在COVID-19流行期间，机器学习也被用来识别自杀相关的论坛发言，并发现其数量增加了1倍多，且边缘型人格障碍患者和创伤后应激障碍患者存在较高的自杀倾向[22]。

2.2.4 图片与视频社交平台数据具有潜力，但需更精准的图像识别技术

Brown等[23]研究表明，Instagram上活跃程度和语言使用的差异与急性自杀无关。机器学习的其他机制（如识别图片内容）可能更有价值。Dagar等[24]分析了YouTube上有关青少年自杀预防和相关健康教育视频的用户留言，约7.5%的用户坦率表达了自杀意念或留言寻求帮助。机器学习可监视各类照片和视频共享网站，例如Instagram、Snapchat和YouTube，以减少涉及自伤和自杀图像的传播[25-26]。随着计算机视觉研究和深度学习技术的发展，AI图像分类技术也许会从血腥、暴力或悲伤的图片或视频信息中识别出潜在的自杀风险。

2.2.5 机器学习结合临床数据库，用于群体筛查

机器学习可适用于各种临床环境和人群，且可以胜任对疾病高危人群的初级筛查工作。2018年初，加拿大公共卫生局与AI公司Advanced Symbolics合作，启动了对区域自杀模式的研究。该公司从加拿大社交媒体帐户中公开获取匿名数据，以监测自杀高危人群并预测自杀高峰[15]。Zheng等[27]通过开发基于人群的风险分层监测系统，使用机器学习算法和深度神经网络建立具有电子健康记录的模型，结合社会经济因素及人口学数据，预测未来12个月的自杀未遂概率。Walsh等[28]将机器学习算法应用于纵向临床数据，以预测青少年的自杀未遂风险，将预测准确性提高了9倍。

目前机器学习已从大型数据库中识别出的自

杀相关危险指标包括临床风险(精神疾病或躯体疾病史)与认知风险(生活满意度、目标、绝望、自尊和自我感知能力等)[29-30]。群体纵向临床数据的使用不仅提供了结合健康数据库进行纵向预测的可能性,更有利于对高危者进行长期管理。

2.3 机器学习应用现状的局限性

2.3.1 准确性和效率需进一步提升

机器学习的算法需不断完善,以兼顾预测准确性和处理速度,这主要是由于:①关键信息难以识别。目前基于互联网平台的算法更多关注与自杀相关的关键词,而忽略包含压力、痛苦、抱怨等可能含有自杀风险的部分。②由于自杀死亡是低概率事件,机器学习算法需在精度和召回率之间寻找平衡。过多地将用户判读为高风险人群会增加非必要的人工筛选和救援工作量,反之,则可能遗漏需要被救援的用户。由于存在自杀意念的人群比例相对较高,而自杀死亡率相对更低,大多数自杀预测模型会存在极低的阳性预测值[5],且即使在自杀高风险人群中,该现象仍存在,这限制了机器学习的实际应用。自 1990 年以来,全球自杀死亡率大幅降低,其中中国下降幅度最大,达到 64.1%[31]。大量人群在临床和生活中表达过负性想法或存在自杀意念,但最终不会付诸行动。潜在的解决方法是,首先保证较高的召回率和偏低的精度,然后引入“触发事件”机制,在识别到有自杀倾向的情况下,获取可靠的触发事件(例如用户在线询问如何购买自杀工具)有助于提高预测的精度。③不同人群有其特殊性,单一算法难以适配所有人群,例如,患有抑郁症、双相障碍、焦虑症、物质滥用、冲动控制障碍以及社会经济地位较低,都被认为是与自杀未遂事件相关的重要特征,至少患有一种精神障碍的个体自杀未遂风险是无精神障碍者的 10 倍以上[27]。因此,应建立针对特殊人群的自杀风险预测模型。

2.3.2 隐私泄露、污名化问题

首先,基于社交媒体的网络平台尚未受到隐私法规的管制。用户自杀相关信息的收集可能侵犯隐私权,从而引发不信任感,并降低用户寻求支持的可能性。同时,个体自杀意念与行为被泄露可能对其工作和生活造成困扰。例如,在军事系统和校园中,单位对个人健康状况有一定的知情权,这将影响其职业和学业生涯,导致当事人利益受损,尤其是自杀识别失误,不仅未能提供帮助,还会给当事人带来污名化[32]。

3 小结与展望

机器学习可以依据收集到的社交网络文字、图片和视频等资料预测用户的自杀风险。在现有自杀预测手段难以满足大规模筛查需求、海量自杀相关数据真假难辨的情况下,机器学习有望成为突破口。机器学习在轻博客、Facebook、讨论论坛、图片与视频社交平台用户自杀预测中的表现值得期待,然而,也需要进一步提升算法的准确性和效率,平衡精度与召回率之间的矛盾,建立不同人群的自杀预测模型,注重隐私保护与污名化问题,并解决后续自杀干预手段不足的问题。

目前,由于机器学习算法仍不够成熟,由计算机进行海量数据的甄别,再由医师做出临床判断的人机结合的预测方式可能是风险最低、效率最高的选择。这需要制定安全处理高风险病例、假阳性或假阴性的预案以及在专家判断和算法判断有冲突时给出决策。在未来的研究中,应注重自杀预测模型的优化。智能手机收集的用户输入信息及穿戴式设备收集的生理数据可能是自杀预测模型的重要补充[33];结合临床数据,如电子病历、就诊记录[34]及静息态功能磁共振数据[35]等也可能有助于提高预测准确率和效率;机器学习亦有望通过对自杀相关脑区的识别[36],并与神经调控技术相结合[37],实现对自杀的实时监测与干预。

参考文献

[1] Bernert RA, Hilberg AM, Melia R, et al. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations[J]. Int J Environ Res Public Health, 2020, 17(16): 5929.

[2] Aleman A, Denys D. Mental health: a road map for suicide research and prevention[J]. Nature, 2014, 509(7501): 421-423.

[3] Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research[J]. Psychol Bull, 2017, 143(2): 187-232.

[4] Chan MK, Bhatti H, Meader N, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales[J]. Br J Psychiatry, 2016, 209(4): 277-283.

[5] Torous J, Walker R. Leveraging digital health and machine learning toward reducing suicide-from panacea to practical tool[J]. JAMA, Psychiatry, 2019, 76(10): 999-1000.

[6] Miller DD, Brown EW. Artificial intelligence in medical

practice: the question to the answer?[J]. Am J Med, 2018, 131(2): 129-133.
[7] Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning[J]. Clin Psychol Sci, 2017, 5(3): 457-469.
[8] Roy A, Nikolitch K, McGinn R, et al. A machine learning approach predicts future risk to suicidal ideation from social media data[J]. NPJ Digit Med, 2020, 3(1): 1-12.
[9] 杨芳, 黄智生, 杨冰香, 等. 基于人工智能技术的微博"树洞"用户自杀意念分析[J]. 护理学杂志, 2019, 34(24): 42-45.
[10] 陈盼, 钱宇星, 黄智生, 等. 微博"树洞"留言的负性情绪特征分析[J]. 中国心理卫生杂志, 2020, 34(5): 437-444.
[11] 章宣, 赵宝奇, 孙军梅, 等. 面向微博文本的自杀风险识别模型[J]. 计算机系统应用, 2020, 29(11): 121-127.
[12] 庄婷婷, 李冬梅, 檀稳, 等. 基于分层支持向量机的微博用户自杀倾向预测与分析[J]. 哈尔滨工程大学学报, 2019, 40(11): 1890-1895.
[13] 许立鹏, 宋文爱. 基于中文微博语言特征的自杀意念检测[J]. 中北大学学报(自然科学版), 2019, 40(4): 350-357.
[14] Cheng Q, Li TM, Kwok CL, et al. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study[J]. J Med Int Res, 2017, 19(7): e243.
[15] Fonseka TM, Bhat V, Kennedy SH. The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors[J]. Aust N Z J Psychiatry, 2019, 53(10): 954-964.
[16] Gomes de Andrade NN, Pawson D, Muriello D, et al. Ethics and Artificial Intelligence: suicide prevention on Facebook[J]. Philos Technol, 2018, 31(4): 1-16.
[17] Cavazos-Rehg PA, Krauss MJ, Sowles SJ, et al. An analysis of depression, self-harm, and suicidal ideation content on tumblr[J]. Crisis, 2017, 38(1): 44-52.
[18] Kavuluru R, Williams AG, Ramos-Morales M, et al. Classification of helpful comments on online suicide watch forums[J]. ACM BCB, 2016: 32-40.
[19] Mörch CM, Côté LP, Corthésy-Blondin L, et al. The darknet and suicide[J]. J Affect Disord, 2018, 241: 127-132.
[20] Yao H, Rashidian S, Dong X, et al. Detection of suicidality among opioid users on reddit: machine learning - based approach[J]. J Med Internet Res, 2020, 22(11): e15293.
[21] Howard D, Maslej MM, Lee J, et al. Transfer learning for risk classification of social media posts: model evaluation study[J]. J Med Internet Res, 2020, 22(5): e15371.
[22] Low DM, Rumker L, Talkar T, et al. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19: observational study[J]. J Med Internet Res, 2020, 22(10): e22635.
[23] Brown RC, Bendig E, Fischer T, et al. Can acute suicidality be predicted by Instagram data? Results from qualitative and quantitative language analyses[J]. PloS One, 2019, 14(9): e0220623.
[24] Dagar A, Falcone T. High viewership of videos about teenage suicide on YouTube[J]. J Am Acad Child Adolesc Psychiatry, 2020, 59(1): 1-3.
[25] Chhabra N, Bryant SM. Snapchat toxicology: social media and suicide[J]. Ann Emerg Med, 2016, 68(4): 527.
[26] Miguel EM, Chou T, Golik A, et al. Examining the scope and patterns of deliberate self-injurious cutting content in popular social media[J]. Depress Anxiety, 2017, 34(9): 786-793.
[27] Zheng L, Wang O, Hao S, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records[J]. Transl Psychiatry, 2020, 10(1): 72.
[28] Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning[J]. J Child Psychol Psychiatry, 2018, 59(12): 1261-1270.
[29] Choi SB, Lee W, Yoon JH, et al. Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea[J]. J Affect Disord, 2018, 231: 8-14.
[30] Ryu S, Lee H, Lee DK, et al. Use of a machine learning algorithm to predict individuals with suicide ideation in the general population[J]. Psychiatry Investig, 2018, 15(11): 1030-1036.
[31] Naghavi M, Global Burden of Disease Self-Harm Collaborators. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the global burden of disease study 2016[J]. BMJ, 2019, 364: 194.
[32] Mörch CM, Gupta A, Mishara BL. Canada protocol: an ethical checklist for the use of artificial Intelligence in suicide prevention and mental health[J]. Artif Intell Med, 2020, 108: 101934.
[33] Haines-Delmont A, Chahal G, Bruen AJ, et al. Testing suicide risk prediction algorithms using phone measurements with patients in acute mental health settings: feasibility study[J]. JMIR Mhealth Uhealth, 2020, 8(6): e15901.
[34] Sanderson M, Bulloch AG, Wang J, et al. Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning[J]. EClinicalMedicine, 2020, 20: 100281.
[35] Bohaterewicz B, Sobczak AM, Podolak I, et al. Machine learning-based identification of suicidal risk in patients with schizophrenia using multi-level resting-state fMRI features[J]. Front Neurosci, 2020, 14: 605697.
[36] Weng JC, Lin TY, Tsai YH, et al. An autoencoder and machine learning model to predict suicidal ideation with brain structural imaging[J]. J Clin Med, 2020, 9(3): 685.
[37] Barredo J, Bozzay M, Primack J, et al. Translating interventional neuroscience to suicide: it's about time[J]. Biol Psychiatry, 2021, 89(11): 1073-1083.

(收稿日期:2021-05-21)
(本文编辑:陈 霞)