

社交网络中抑郁风险特征的识别:深度学习的方法*

穆 易¹, 李 强¹, 黄 鑫¹, 孟祥寒²

(1. 南开大学社会学院社会心理学系, 天津 300071; 2. 天津行政学院, 天津 300191)

摘 要:随着社交网络的发展,越来越多的人倾向于在社交网络社区中表达自己情感和需求,其中包括抑郁症状的相关信息,从而为抑郁风险特征的识别创造了新路径。本文考察了多种深度学习模型对新浪微博“抑郁症社区”文本中抑郁风险特征的识别效果。为验证算法模型的有效性,结合词频分析结果编制了抑郁风险特征注释方案,并比较了九种前沿算法模型的预测效果,结果表明基于文本的卷积神经网络模型更适用于在多分类任务中捕捉关键语言特征。研究发现为基于社交网络和深度学习的抑郁风险预警提供了借鉴。

关键词:社交网络;抑郁风险特征;注释方案;深度学习

中图分类号:B848

文献标志码:A

文章编号:1003–5184(2026)01–0050–08

1 引言

抑郁症是临床常见的精神障碍,预计到2030年将跃升为全球最大的疾病负担,成为公共健康领域的关键问题(张小崔等,2024)。早期识别和干预是有效防控的前提,然而我国当前抑郁症患者的治疗率仍然偏低,成人患抑郁症的比例约为6.8%,但仅有约0.5%的患者能够获得充分规范的治疗(张伟霞等,2023)。除此之外,心理障碍污名化等问题导致许多患者“谈病色变”(李强等,2020),不能及时察觉病情。近年来,人们开始从互联网中寻求非正式帮助(袁钦涓等,2020),社交网络填补了心理健康资源方面的空缺(Cavazos–Rehg et al., 2016; Yao et al., 2021),匿名分享的方式保护了抑郁症患者的自尊并赋予其归属感,提高其对专业帮助可获得性的认识,从而为抑郁症干预策略提供了新思路(Li et al., 2018; Yu et al., 2021)。

近年来,自然语言处理技术(Natural Language Processing, NLP; Karmen et al., 2015)在社交媒体数据分析中得到了广泛应用。NLP能够将自然语言转换为计算机可解析的数值表示,进而通过构建预测模型,实现对抑郁症的识别与检测。常见的文本数值化方法包括“语言探索与字词计数”软件(Linguistic Inquiry and Word Count, LIWC)、潜在狄利克雷分配等(Blei et al., 2003; Pennebaker et al., 2015)。在此基础上,结合机器学习算法(De Choudhury et al., 2013),可以识别数据中的复杂关

系和模式,预测个体的情感变化和心理状态(Tausczik & Pennebaker, 2009)。这类计算方法在抑郁症等精神疾病特征捕捉中展现了重要价值(Squires et al., 2023)。

以往研究中,有学者基于Twitter行为数据,采用支持向量机、随机森林等方法预测抑郁发作(De Choudhury et al., 2013; Reece et al., 2017);也有学者采用K-近邻方法,比较Facebook用户中抑郁与非抑郁个体在语言使用上的差异(Islam et al., 2018)。相关成果表明,通过社交媒体生成的文本语料库(田玮,朱廷劭,2018),结合特征分析方法(Wang et al., 2024)与机器学习技术(Burkart & Huber, 2021),可以有效支持抑郁症状的识别与早期预防(董健宇等,2020;苏悦等,2021)。随着网络语言的复杂性增加,网络用语、情感波动和隐喻、俚语等语言特征变得难以捕捉。为了解决这一问题,深度学习模型被广泛应用于抑郁症状检测中,其在捕捉文本中复杂、隐晦的语义信息和语言特征上,比传统的机器学习方法更有效(Devlin et al., 2018; Mikolov et al., 2013; Pennington et al., 2014)。已有研究结合极度梯度提升树(eXtreme Gradient Boosting, XGBoost)与双向长短期记忆网络(Bidirectional Long Short–Term Memory, BiLSTM)对早期抑郁症进行检测,评估结果表明其识别性能优于传统支持向量机方法(Cong et al., 2018);也有研究发现,采用卷积神经网络与循环神经网络相融合的混合模型,

* 基金项目:国家社科基金项目(23BSH131)。

通信作者:李强, E-mail: liqiangp@126.com。

能够更精准地预测 Facebook 用户的抑郁与压力状态(Rosa et al., 2019)。

尽管深度学习在处理网络语言的复杂性方面相较传统机器学习更具优势,其在抑郁症识别中的应用仍面临诸多挑战,尤其是对大量标注数据的依赖和模型可解释性的缺失。以往研究更关注模型的预测性能指标,却忽视了特征背后心理学语义的深度解析,尤其缺乏对社会心理背景与精神病理学的综合考量(Rosa et al., 2019)。例如,有研究通过 DSM-5 框架标注 Twitter 中涉及重度抑郁症的文本症状(Mowery et al., 2017),或借助流调中心抑郁量表(CES-D)对社交媒体文本进行情绪状态标注(De Choudhury et al., 2013; Reece et al., 2017),但这类标准化工具在面对非结构化社交语境、非典型抑郁症状或潜在求助信号时,存在语义覆盖不足、上下文感知能力低等局限性。基于此,我们结合对目标社群文本语料的主题分布与心理特征分析,编制了新的抑郁风险特征注释方案,覆盖情绪表现、认知表现、动力性表现、植物性和躯体性表现、其他精神表现、求助及救助、人际问题与其他(即无关文本)这八个维度。该方案在延续抑郁症状分类理论(Beck & Alford, 2009)的基础上,进一步结合词频分析结果,纳入社交互动信息与非结构化语境下的潜在抑郁风险特征,实现了对社交媒体语言中非典型症状、求助行为及人际障碍的扩展,从而增强了模型在多

维心理特征识别中的适应能力与语义覆盖广度。相比于传统标签方式,多维度注释系统能够更有效地识别语义复杂、情绪模糊的文本(Cong et al., 2018)。

总而言之,依托具有强大的特征提取和上下文建模能力的模型,并使用符合社交网络环境的注释方案是提升抑郁风险特征文本的识别准确性和鲁棒性的关键(Squires et al., 2023)。基于以上考虑,通过广泛地采集新浪微博的“抑郁症社区”中 2018 至 2024 年间的文本数据(Yao et al., 2020),结合抑郁症状分类标准(Beck & Alford, 2009)形成的新型注释方案,对多种前沿的深度学习模型在多类别特征上的识别效果进行比较,以期为社交媒体数据在抑郁风险特征识别中的应用提供方法参考。

2 方法

2.1 数据收集与处理

采用的数据来自 2018 年 1 月 1 日至 2024 年 4 月 30 日期间,通过“抑郁症超话”关键词公开获得的 451071 条新浪微博推文。经剔除来源不属于抑郁症社区的推文,最终实际采用数据 344817 条,其中包含 55618 名社区成员,每个成员平均发布推文数量约为 6.20 条(见表 1)。研究充分尊重参与者的合法权益,所使用的公开数据已通过所在单位的伦理审核,可能涉及个人身份的信息均已删除或隐匿,处理后的数据公开发布在心理科学数据银行中。

表 1 2018 - 2024 微博推文详情情况

总数量	实际数量	成员数量	人均推文数量	最小值	最大值	最高点赞	最高评论	最高转载
451071	344817	55618	6.20	1	1424	7635	16103	1441

2.2 推文内容的主题分析

基于 Python 3.8,结合 Jieba 中文分词工具包对数据进行词频统计分析。由于文本中可能包含与研究无关的词汇、网址和广告等,故手动构建了停用词表,剔除了特殊符号、生僻字和语气词。最后,对筛选后的 344817 条数据进行了关键词分析,直观地呈现了关键词的分布情况,旨在挖掘文本的核心主题和成员的主要需求,为抑郁风险特征的分类提供更全面的支持。

2.3 抑郁风险特征分类的注释

数据标记工作由两位来自临床与心理咨询方向的研究人员完成,分别对随机抽取的 3000 条文本进行了标记。采用 Cohen's Kappa ($K = 0.9568$) 法验证了注释者之间的一致性(Cohen, 1960)。值得注意的是,社交媒体语料的标签呈现不均衡分布,例如“情绪表现”和“其他”类别的文本数量明显多于“求

助及救助”或“植物性和躯体性表现”等类别。此现象并非人工筛选导致,而是生态语料的特性使然。在多标签分布差异显著的情境下,提升模型识别能力亦是研究的核心目标之一。

2.4 神经网络模型的建模

使用 Python 3.8 和 PyTorch 深度学习库实现了九种算法的建模,比较和分析它们在预测任务中的性能。这九种模型主要包括八种前沿的深度学习架构(含预训练模型)与一种浅层模型(FastText)。在数据集划分方面,从标注后的语料中进一步梳理并选取表述清晰、标注一致性高的语料($n = 2887$),并以约 6:2:2 的比例分别划分为训练集($n = 1732$)、测试集($n = 578$)和验证集($n = 577$),用于对模型的训练和评估。这种划分方式有助于充分利用数据并有效提升模型的泛化能力。模型详细情况见表 2。

表 2 深度学习模型列表

模型	描述	参考文献
TextRCNN	结合 BiLSTM 和池化,提高文本特征提取能力。	Lai et al. ,2015
TextCNN	利用 CNN 提取文本局部特征,擅长段文本分类。	Kim,2014
TextRNN	基于 BiLSTM 网络,捕捉序列中的前后依赖关系。	Liu et al. ,2016
DPCNN	通过堆叠卷积与残差连接构建金字塔结构,提取长文本中的全局语义特征。	Johnson & Zhang,2017
TextRNN_Att	结合注意力机制(Attention Mechanism)的 TextRNN,增强对重要文本片段的关注。	Zhou et al. ,2016
FastText	通过词袋模型和 N-gram 技术,实现快速文本分类。	Joulin et al. ,2016
Transformer	基于注意力机制的无序列依赖模型,适合并行计算。	Vaswani et al. ,2017
BERT	预训练语言模型,通过微调适应不同下游任务。	Devlin et al. ,2018
ERNIE	增强版的 BERT 模型,进一步提升了模型的表现。	Sun et al. ,2019

2.5 模型评估指标

为了评估模型的综合性能,采用预测准确率(Accuracy)、宏平均值(Macro Average)和加权平均值(Weighted Average)评估模型的综合预测性能。其中,预测准确率衡量所有样本中正确预测的比率,宏平均值计算了每个类别的平均评估指标,加权平均值则根据每个类别的样本数量加权计算平均评估指标。在对抑郁风险特征的预测方面,选择了多特征分类任务中常用评价指标,包括精确率(Precision)、召回率(Recall)和 F1 值。其中,精确率表示模型正确预测为正类别的样本占有所有预测为正类别样本的比例,召回率表示模型正确预测为正类别的样本占有所有实际正类别样本的比例,F1 值是精确率和召回率的调和平均值,这些指标能够分别评估模型在不同类别上的表现。

3 结果

3.1 社区成员关注的主题情况

基于 Python 3.8 和中文分词工具包 Jieba,结合停用词表,提取了总体数据中词频前 100 的关键词。分析结果揭示了抑郁症社区成员在文本表达中的核心主题和常用词汇。通过对这些关键词的深入解读,进一步揭示了社区成员的关注焦点。

研究结果显示,抑郁症社区成员普遍关注情感状态、社交关系以及生活和健康问题。高频词汇如“为什么”、“不想”、“哭”、“难受”等,反映了成员在面对心理困扰时的困惑、无助和痛苦,表现出明显的负面情绪倾向(Sabshin, 1968)。与此同时,“大

家”、“朋友”、“妈妈”等词汇则表明,尽管表达负面情绪,社区成员仍在积极寻求社会支持和情感慰藉。生理与心理症状交织的特征也十分显著,诸如“失眠”、“焦虑”、“崩溃”等关键词揭示了心理困扰对个体身心状态的双重影响(Cohen et al. , 1985)。值得注意的是,尽管负面情绪词汇频繁出现,社区成员仍表现出对积极情感和康复的渴望。词汇如“医生”、“吃药”、“医院”反映了成员对专业医疗帮助的寻求,而“希望”、“开心”、“坚持”、“加油”等则表达了对生活的期待与自我恢复的努力,或强烈的求助和自我修复意愿。总的来说,词频分析结果反映了成员对信息、消极情绪的表达、社交关系的关切和积极情感的寻求。此外,频繁出现的与死亡、自杀相关的词语表明了极端情绪的存在,需要特别关注和支持。

3.2 注释方案的编制

基于贝克的《抑郁症》(DEPRESSION: Causes and Treatment, 2nd Edition)中对抑郁症状的分类(Beck & Alford, 2009),结合微博文本关键词统计结果,提出了一套新的抑郁风险特征注释方案。注释方案的编制工作由一名注册督导师全程监督、指导,三名临床与心理咨询方向的博士研究生通过多轮注释和评估共同完成。根据抑郁症状表现的轻、中、重度进行文本解读,最终注释分为八类,分别是:情绪表现、认知表现、动力性表现、植物性和躯体性表现、其他精神表现、求助及救助、人际问题以及其他(见表 3)。

表 3 注释方案

一类	二类
情绪表现	沮丧的情绪;对自我的消极情感;欲望满足感缺失;情感依恋缺失;哭泣;欢乐感缺失
认知表现	自我评价低;消极期望;自责和自我批评;优柔寡断;自我形象扭曲
动力性表现	意志瘫痪;回避、逃避和戒断;自杀观念;依赖性增加
植物性和躯体性表现	缺乏食欲;睡眠紊乱;性欲减退;疲劳
其他精神表现	精神型(妄想和幻觉);焦虑型;迟缓型;自我伤害和虐待
求助及救助	服药;咨询;就医;建议;互助;自救与自我鼓励

续表 3

一类		二类
人际问题	人际相关的敌意、孤独感	
其他	与抑郁风险主题无关的文本	

3.3 模型预测效果

为验证所提出模型的有效性,本文对不同模型的预测准确率、宏平均值及加权平均值进行了系统比较。图 1 至图 3 展示了各模型的预测表现。结果

表明,TextRCNN 模型在各项指标中均优于其他模型,表现最为优异,其预测准确率为 42.39%,宏平均精确率为 44.88%,加权平均精确率为 44.63%。

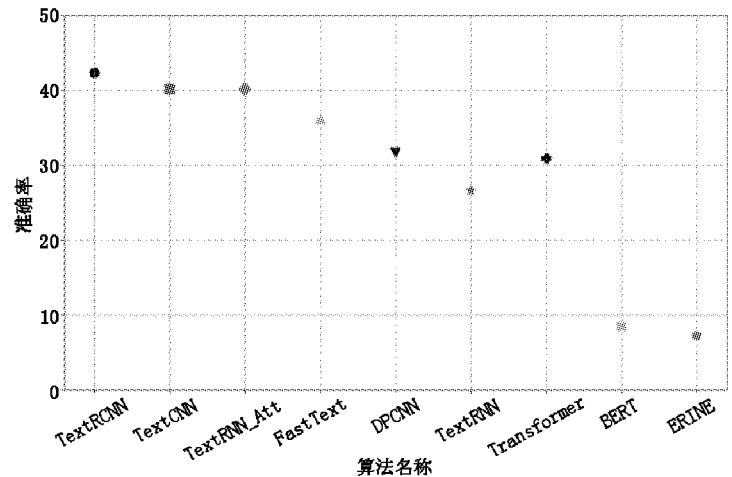


图 1 模型的准确率对比图

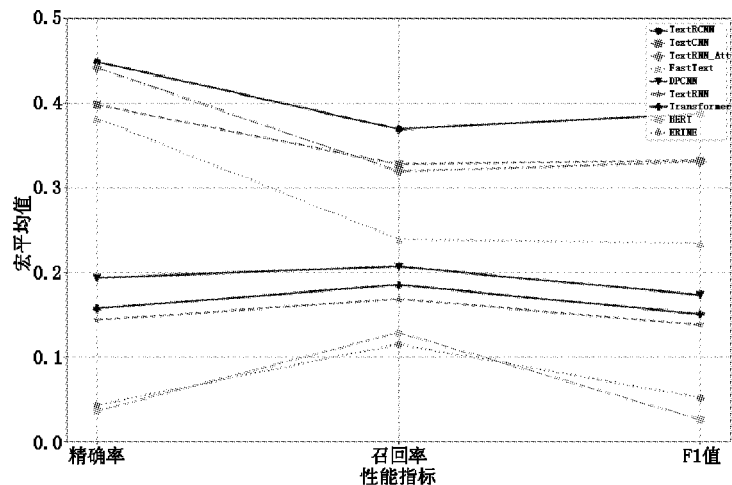


图 2 模型的宏平均值对比图

在特征识别方面,图 4 至图 6 呈现了各模型在特征识别任务中的表现,具体评估指标包括精确率、召回率以及 F1 值。结果显示,相对于其他模型,TextRCNN 在所有评估指标上预测效果更好且波动更小,表现出更高的稳定性。尤其是在“情绪表现”、“植物性和躯体性表现”与“求助及救助”类文本的识别中,该模型体现出显著的优势。同时,该模型在识别并排除无关文本时也表现出较高的准确性。

4 讨论

研究以微博抑郁症社区用户文本为研究对象,旨在探索深度学习模型在抑郁风险特征预测中的可行性。词频分析结果显示了社区成员对消极情绪的表达、社交关系的关切和积极情感的需求,基于此,本文编制了全新的抑郁风险特征注释方案,进一步对比了多种深度学习模型在该数据集上的综合预测效果及特征识别能力。

在模型综合评估方面,采用识别准确率、宏平均

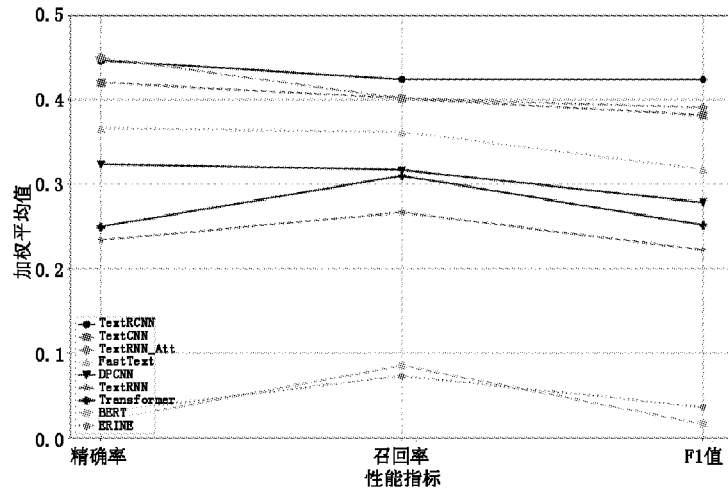


图3 模型的加权平均值对比图

值和加权平均值,共同衡量了模型的综合预测效果。其中,深度学习模型 TextRCNN 的预测准确率表现优于其他深度模型和传统浅层模型 FastText。通常情况下,在多类别分类任务中,随着特征数量的增

加,模型的评估指标往往会有所下降,尤其是在样本分布不均和特征重叠较大的环境下,会大幅度降低模型预测效果 (Spelman & Porkod, 2018)。这些因素不仅增加了模型的复杂性,还降低了模型在不同

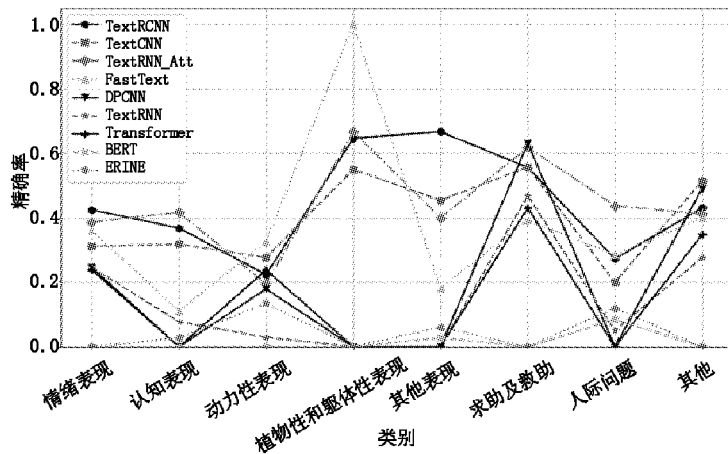


图4 模型的预测精确率对比图

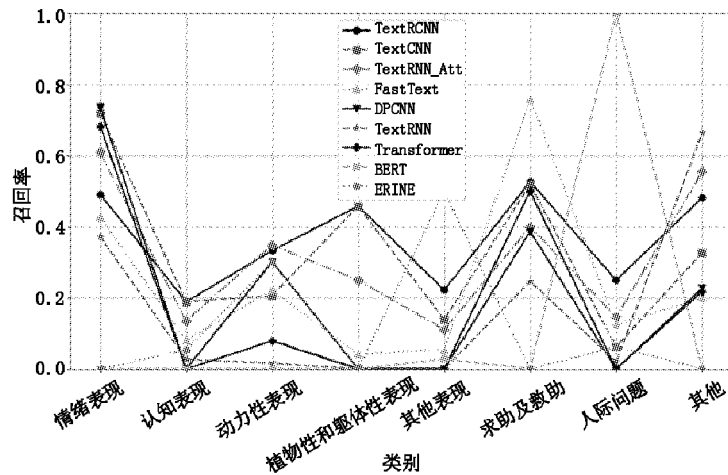


图5 模型的预测召回率对比图

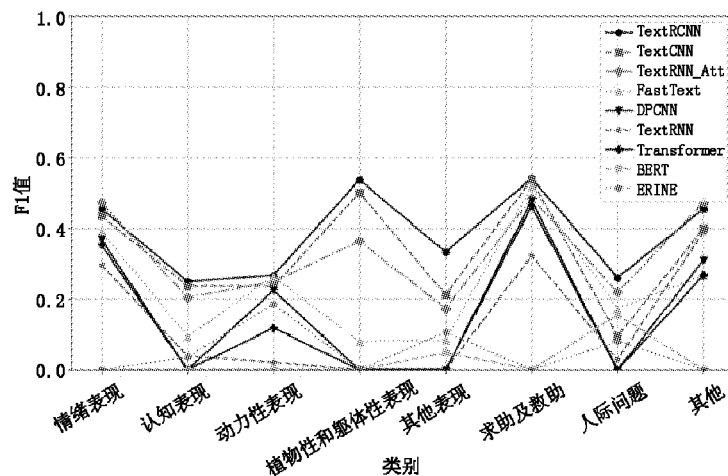


图6 模型的F1值对比图

类别之间的泛化能力。研究结果突显了深度学习模型在这些极端条件下的学习能力,尽管数据的稀缺性和类别的不平衡对模型的预测效果带来挑战,但TextRCNN仍展现出相对较强的特征捕捉和模式识别能力。

在具体的特征识别方面,深度学习模型TextRCNN的预测效果依旧好于其他模型,尤其是在“情绪表现”(F1 = 45.57%),“植物性和躯体性表现”(F1 = 53.66%)与“求助及救助”(F1 = 54.11%)这几类特征中。这表明,该模型在数据分布和文本多样性较为复杂的环境下,仍擅长建模长文本和上下文语义,进一步验证了深度学习模型在非结构化文本处理中的优势(Fu & Liu, 2023; Lu et al., 2023)。类似地,TextRNN_Att也可以利用循环神经网络(RNN)层的时间序列特性,捕捉文本中的情感波动和语义变化,而TextCNN的卷积层则能帮助模型通过多通道特征学习不同维度的信息(Conneau et al., 2017)。这使得此类模型在多类别特征识别任务中展现出稳定且较好的效果,尤其是在处理复杂、非标准语言时,能够保持持续且较高的稳定性(Fu & Liu, 2023)。然而,DPCNN尽管也是深层结构模型,其采样机制在处理微博这种篇幅较短、信息密度高的文本时,容易导致关键信息丢失(Zhang et al., 2023)。而FastText作为浅层模型,采用词平均方式建模,缺乏对上下文和复杂语义的捕捉能力,尤其在情感识别任务中表现不佳(Joulin et al., 2016)。在预训练模型方面,BERT和ERNIE理论上具备强大的语义建模能力,但却在“求助文本”等关键类别上存在识别不足,这可能与微博环境下的网络俚语、隐喻表达等特征与预训练语料差异较大有关(Devlin et al., 2018; Sun et al., 2019)。Transformer尽管具备全局感知能力,但并行计算机

制缺乏对上下文连续性的建模,可能是影响识别效果的原因(Vaswani et al., 2017)。

总而言之,通过人工标注方案将临床抑郁特征转换为可计算的文本指标,验证了症状学理论在社交媒体环境中的部分可迁移性,模型识别的是抑郁相关语言表达模式而非诊断结果,进一步体现了数字表型(digital phenotype)概念的可操作化(Guntuku et al., 2019)。研究基于临床理论,实现了症状学维度与社交媒体语言特征的映射,为数字表型分析提供了启示。相较于以往研究(Reece et al., 2017),TextRCNN模型实现了对中文社交媒体语境下抑郁风险特征的多维细粒度识别(八类特征),从技术层面验证了混合神经网络结构在捕捉中文抑郁表达复杂语义方面的优势。尽管如此,当前研究仅探索了对文本中抑郁风险特征的识别,未能实现对个体抑郁风险的预测,未来研究可尝试引入多模态特征(如图像、音频和用户行为数据),结合特征预测方法,细化特征标签并平衡样本数据,在个案层面提供更加个性化的预测。尤其是在早期识别和干预方面,通过精准捕捉个体的情感变化轨迹,及时进行心理健康干预,进一步提升模型的预测性能和应用价值。

5 结论

深度学习模型有望在社交网络中预警抑郁风险,选择符合语言环境的注释方案是应用此类方法的关键,识别的主要困难来自样本的不平衡分布和语义复杂性,尤其是抑郁症状的间接或隐含表述。研究系统地分析了抑郁症社区成员在社交平台上的主题,对比了多种深度学习模型在抑郁风险特征上的识别效果,得出以下结论:(1)基于贝克的抑郁症理论修订后的注释方案具有良好的有效性。(2)相对于其他深度学习模型,融合的深度深度学习模型Tex-

tRCNN 能够好地捕捉社交网络抑郁症社区的文本特征,尤其是在“情绪表现”、“植物性和躯体性表现”与“求助及救助”类文本上。

参考文献

- 董健宇, 韦文棋, 吴珂, 妮娜, 王黎霏, 付莹, ... 彭歆. (2020). 机器学习在抑郁症领域的应用. *心理科学进展*, 28(2), 266 – 274.
- 李强, 魏晓薇, 翟宏堃. (2020). 健康中国视角下优化国民心理健康素养研究: 意义与取径. *西南交通大学学报(社会科学版)*, 21(4), 61 – 68.
- 苏悦, 刘明明, 赵楠, 刘晓倩, 朱廷劼. (2021). 基于社交媒体数据的心理指标识别建模: 机器学习的方法. *心理科学进展*, 29(4), 571 – 585.
- 田玮, 朱廷劼. (2018). 基于深度学习的微博用户自杀风险预测. *中国科学院大学学报*, 35(1), 131 – 136.
- 张小崔, 杨颜慈, 雷辉, 董戴凤, 冯志凌. (2024). 青少年睡眠问题与抑郁症状关系的网络分析. *中国临床心理学杂志*, 32(4), 824 – 830.
- 张伟霞, 席敏, 阴甜甜, 王成, 司书宾. (2023). 基于网络分析的抑郁症产生与演变预测. *心理科学进展*, 31(11), 2129 – 2141.
- Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment* (2nd ed.). University of Pennsylvania Press.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993 – 1022.
- Burkard, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70, 245 – 317.
- Cavazos – Reh, P. A., Krauss, M. J., & Bierut, L. J. (2016). A content analysis of depression – related tweets. *Computers in Human Behavior*, 54, 351 – 357.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37 – 46.
- Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, 98(2), 310 – 357.
- Cong, Q., Feng, Z., Li, F., Xiang, Y., Rao, G., & Tao, C. (2018). X – A – BiLSTM: A deep learning approach for depression detection in imbalanced data. In 2018 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1624 – 1627). IEEE.
- Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*.
- De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2021). Predicting depression via social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), 128 – 137.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre – training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Fu, T., & Liu, H. (2023). Research on Chinese Text Classification Based on Improved RNN. In 2023 *IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)* (pp. 1554 – 1558). IEEE.
- Guntuku, S. C., Schneider, R., Pelullo, A., Young, J., Wong, V., Ungar, L., ... Merchant, R. (2019). Studying expressions of loneliness in individuals using twitter: An observational study. *BMJ Open*, 9(11), e030355.
- Islam, M. R., Kamal, A. R. M., Sultana, N., Islam, R., Moni, M. A., & Ulhaq, A. (2018). Detecting depression using K – nearest neighbors (KNN) classification technique. In 2018 *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)* (pp. 1 – 4). IEEE.
- Johnson, R., & Zhang, T. (2017). Deep pyramid convolutional neural networks for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (pp. 562 – 570). Association for Computational Linguistics.
- Joulin, A., Grave, E., & Mikolov, T. (2016). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.
- Karmen, C., Hsiung, R. C., & Wetter, T. (2015). Screening internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. *Computer Methods and Programs in Biomedicine*, 120(1), 27 – 36.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1).
- Li, A., Jiao, D., & Zhu, T. (2018). Detecting depression stigma on social media: A linguistic analysis. *Journal of Affective Disorders*, 232, 358 – 362.
- Liu, P., Qiu, X., & Huang, X. (2016). Recurrent neural network for text classification with multi – task learning. *arXiv preprint arXiv:1605.05101*.
- Lu, S., Jiang, Y., Xie, S., Ye, J., Zhang, Z., & Zhao, Y. (2023). Sensitive word recognition scheme based on Text – RCNN model. In 2023 *International Conference on Data Science and Network Security (ICDSNS)* (pp. 1 – 6). IEEE.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546*.
- Mowery, D., Smith, H., & Conway, M. (2017). Understanding depressive symptoms and psychosocial stressors on Twitter: A corpus – based study. *Journal of Medical Internet Research*, 19(2), e48.

- Pennebaker, J. , Boyd, R. , Jordan, K. , & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.
- Pennington, J. , Socher, R. , & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532 – 1543). Association for Computational Linguistics.
- Reece, A. G. , Reagan, A. J. , Lix, K. L. M. , Dodds, P. S. , Danforth, C. M. , & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports*, 7(1), 13006.
- Rosa, R. L. , Schwartz, G. M. , Ruggiero, W. V. , & Rodriguez, D. Z. (2019). A knowledge – based recommendation system that includes sentiment analysis and deep learning. *IEEE Transactions on Industrial Informatics*, 15(4), 2124 – 2135.
- Sabshin, M. (1968). Depression: Clinical, experimental and theoretical aspects. *Archives of General Psychiatry*, 19(6), 766 – 767.
- Spelman, V. S. , & Porkodi, R. (2018). A review on handling imbalanced data. In *2018 International Conference on Current Trends Towards Converging Technologies (ICCTCT)* (pp. 1 – 11). IEEE.
- Squires, M. , Tao, X. , Elangovan, S. , Gururajan, R. , Zhou, X. , Acharya, U. R. , & Li, Y. (2023). Deep learning and machine learning in psychiatry: A survey of current progress in depression detection, diagnosis and treatment. *Brain Informatics*, 10(1), 10.
- Tausczik, Y. R. , & Pennebaker, J. W. (2009). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24 – 54.
- Vaswani, A. , Shazeer, N. , Parmar, N. , Uszkoreit, J. , Jones, L. , Gomez, A. N. , . . . Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Wang, Y. , Tian, J. , Yazar, Y. , Ones, D. S. , & Landers, R. N. (2024). Using natural language processing and machine learning to replace human content coders. *Psychological Methods*, 29(6), 1148 – 1163.
- Yao, X. , Yu, G. , & Tang, J. (2020). Patterns and longitudinal changes in negative emotions of people with depression on Sina Weibo. *Telemedicine and E – Health*, 26(6), 734 – 743.
- Yao, X. , Yu, G. , & Zhang, J. (2021). Extracting depressive symptoms and their associations from an online depression community. *Computers in Human Behavior*, 120, 106734.
- Yu, L. , Jiang, W. , & Hu, X. (2021). Detecting changes in attitudes toward depression on Chinese social media: A text analysis. *Journal of Affective Disorders*, 280, 354 – 363.
- Zhang, M. , Pang, J. , Cai, J. , Huo, Y. , Yang, C. , & Xiong, H. (2023). DPCNN – based models for text classification. In *2023 IEEE 10th International Conference on Cyber Security and Cloud Computing (CSCloud)/2023 IEEE 9th International Conference on Edge Computing and Scalable Cloud (EdgeCom)* (pp. 363 – 368). IEEE.
- Zhou, P. , Shi, W. , Tian, J. , Qi, Z. , Li, B. , Hao, H. , & Xu, B. (2016). Attention – based bidirectional long short – term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 207 – 212). Association for Computational Linguistics.

Identification of Depression Risk Features in Social Networks: Deep Learning – Based Approach

Mu Yi¹, Li Qiang¹, Huang Xin¹, Meng Xianghan²

(1. Department of Social Psychology, School of Social Sciences, Nankai University, Tianjin 300071;

2. Tianjin Academy of Administration, Tianjin 300191)

Abstract: With the development of social networks, an increasing number of individuals are expressing their emotions and needs in online communities, including disclosures of depressive symptoms, thereby opening new avenues for detecting depression risk features. This study evaluated the performance of nine advanced deep learning models in identifying depression risk indicators within posts from the “Depression Community” on Sina Weibo. To validate model effectiveness, we first conducted a term – frequency analysis across the entire corpus to create a standardized annotation scheme for depression risk features, and then compared each model’s predictive accuracy in a multiclass classification task. Experimental results demonstrate that a text – based convolutional recurrent neural network is particularly adept at capturing the critical linguistic cues associated with depression risk. These findings provide a reference for developing depression risk early – warning systems based on social networks and deep learning.

Key words: social networks; depression risk features; annotation scheme; deep learning