

## 社交媒体文本数据的抑郁症检测研究综述

徐东东, 蔡肖红, 刘 静, 曹 慧

山东中医药大学 智能与信息工程学院, 济南 250355

**摘 要:**近年来,机器学习被逐渐运用到基于社交媒体文本数据的抑郁症检测中并凸显重要应用价值。为梳理其应用现状和发展方向,对用于抑郁症检测的社交媒体文本数据集、数据预处理和机器学习方法进行整理分类。在数据特征表示方面,对比分析了基础特征表示、静态词嵌入和语境词嵌入。全面分析了利用不同基础特征和不同算法类型的传统机器学习以及深度学习进行抑郁症检测的性能和特点。总结并建议未来在中文数据集的创建、模型的可解释性、基于隐喻的检测和轻量级预训练模型方面做进一步的探索。

**关键词:**社交媒体;文本数据;抑郁症;机器学习

**文献标志码:**A **中图分类号:**TP391 **doi:**10.3778/j.issn.1002-8331.2209-0042

### Review of Depression Detection Using Social Media Text Data

XU Dongdong, CAI Xiaohong, LIU Jing, CAO Hui

College of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China

**Abstract:** Machine learning has been gradually applied to depression detection using social media text data, and has prominently shown important application value in recent years. Firstly, this paper organizes and classifies social media text datasets, data preprocessing and machine learning methods used for depression detection. In addition, in terms of data feature representation, the basic feature representation, static and contextual word embedding are compared and analyzed. Secondly, this paper analyzes comprehensively the performance and characteristics of traditional machine learning with different basic features and different algorithm types as well as deep learning for depression detection. Finally, this paper summarizes and suggests further explorations in Chinese dataset creation, model interpretability, metaphor-based detection and lightweight pre-training model.

**Key words:** social media; text data; depression; machine learning

抑郁症是全世界主要致残原因,也是造成全球疾病总负担的主要因素之一<sup>[1]</sup>,其主要特点有持续的悲伤、失去兴趣或快乐等。抑郁症会给患者身心带来极大的影响且容易复发,当抑郁症发展到中度或重度时,将成为一个严重的健康疾患甚至导致患者自杀<sup>[2]</sup>,从而对家庭和社会造成严重的损失。

目前,抑郁症的诊断主要基于国际疾病分类标准(international classification of diseases, ICD)和精神疾病诊断统计手册(diagnostic and statistical manual of mental disorders, DSM),这种诊断方式易受患者的主观信念和医师的诊断经验影响<sup>[3]</sup>。此外,现代社会对于抑郁症等精神类疾病的接受程度普遍较低,抑郁症患者自身也常由于存在病耻感和对精神疾病缺乏了解等原因

而没有选择求医,结果造成众多患者因未得到及时、准确的诊断而错失最佳治疗机会<sup>[1]</sup>。

随着互联网的兴起,人们越来越倾向于在社交媒体上分享自己的生活状态,且患有精神疾病的人也以倾诉自己的精神状态作为一种解脱<sup>[4]</sup>。同时,越来越多的证据表明,社交媒体平台上发布的特定语言和情绪可能提供了关于抑郁症的线索<sup>[5-8]</sup>。在此背景下,机器学习也逐渐被运用到基于社交媒体文本数据的抑郁症检测中<sup>[9-11]</sup>。基于传统机器学习的方法可以执行自动、客观和有效的评估<sup>[12]</sup>,但是其性能在很大程度上依赖于特征的构建和选择,并且泛化性受到所使用特征和算法的限制。而深度学习以理解复杂自然语言句子的上下文为目标,彻底改变了潜在特征提取过程。现有的基于深度学习的抑

**基金项目:**国家自然科学基金(82074579, 81973981, 82174528);山东省自然科学基金(ZR2020MH360)。

**作者简介:**徐东东(1996—),男,硕士研究生,研究方向为医疗健康大数据;蔡肖红(1989—),女,硕士,讲师,研究方向为自然语言处理、医疗数据分析;刘静(1980—),女,硕士,副教授,CCF会员,研究方向为医学图像处理;曹慧(1963—),通信作者,女,硕士,教授,研究方向为中医药智能化研究、医疗健康大数据, E-mail: caohui63@163.com。

**收稿日期:**2022-09-05 **修回日期:**2022-11-28 **文章编号:**1002-8331(2023)04-0054-10

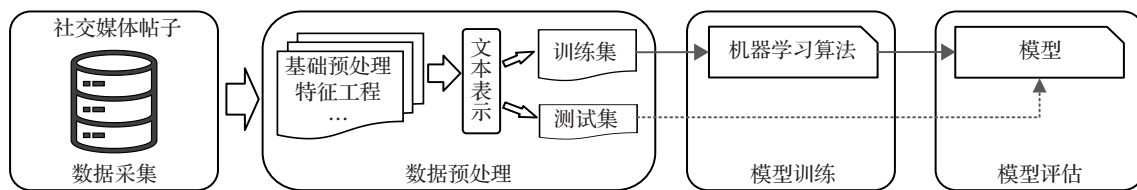


图1 利用机器学习方法检测抑郁症的一般流程

Fig.1 General process of detecting depression using machine learning

郁症检测系统能够执行预处理、特征提取和抑郁症检测等连续过程,实现了端到端的全自动化抑郁症检测<sup>[13]</sup>,在抑郁症的预防和治疗方面具有重大意义。

目前,国外关于利用机器学习在基于社交媒体文本数据的抑郁症检测研究仍在不断发展进步,但国内少有关于此领域的研究和报告。本文对在社交媒体文本中运用机器学习检测抑郁症进行综述,以期国内研究提供借鉴。

## 1 机器学习及流程概述

机器学习(machine learning, ML)是指利用计算机通过对已有数据进行自主学习以改善自身功能,从而能够在下一次执行相同任务时做得更好或者效率更高的一种技术。机器学习可根据用于学习的数据性质分为监督学习、无监督学习、半监督学习,也可根据模型结构的深度分为传统机器学习和深度学习<sup>[14]</sup>。

利用机器学习方法在基于社交媒体文本数据中检测抑郁症的一般流程如图1所示,主要分为以下步骤:数据采集,数据预处理(基础预处理和特征工程),利用机器学习算法对文本表示进行学习,以及使用测试数据评估已学习好的模型。

目前,广泛使用在社交媒体文本中检测抑郁症的传统机器学习算法有逻辑回归(logistic regression, LR)、决策树(decision tree, DT)、支持向量机(support vector machine, SVM)、朴素贝叶斯(naive Bayes, NB)和随机森林(random forest, RF)等。而随着深度学习的发展,卷积神经网络(convolutional neural network, CNN)、循环神经网络(recurrent neural network, RNN)和基于Transformers的双向编码器表示(bidirectional encoder representation from transformers, BERT)等算法得以推广和使用。

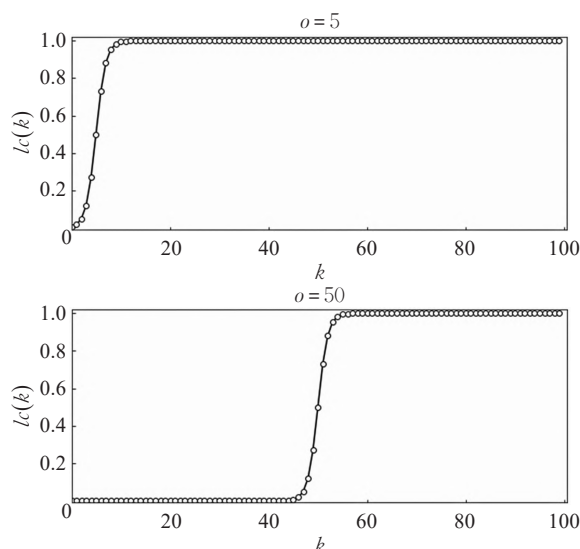
衡量抑郁症检测算法性能的常用评价指标有准确率(accuracy, Acc)、精确率(precision, P)、召回率(recall, R)和F1值。而这些评价指标没有考虑到时间因素,对此Losada等人<sup>[15]</sup>提出了早期风险检测误差(early risk detection error, ERDE)指标。该指标同时考虑二元决策的正确性和模型做出决策所用的延迟,而延迟通过在模型给出预测之前所输入文章(帖子或评论)的数量( $k$ )来衡量。ERDE指标的计算如式(1)所示:

$$ERDE_o(d, k) =$$

$$\begin{cases} c_{fp}, d = positive \text{ AND } gt = negative \text{ (FP)} \\ c_{fn}, d = negative \text{ AND } gt = positive \text{ (FN)} \\ lc_o(k) \cdot c_{tp}, d = positive \text{ AND } gt = positive \text{ (TP)} \\ 0, d = negative \text{ AND } gt = negative \text{ (TN)} \end{cases} \quad (1)$$

$$lc_o(k) = 1 - \frac{1}{1 + e^{(k-o)}} \quad (2)$$

其中,  $d$  为模型所做出的决策,  $gt$  为黄金真理(golden truth),  $c_{fp}$  和  $c_{fn}$  分别为假阳性和假阴性的代价。函数  $lc_o(k)$  ( $\in [0, 1]$ ) 代表检测真阳性的延迟的代价,其计算公式如式(2)所示。 $o$  为延迟成本函数中代价增长更快的  $k$  轴的位置,也是  $lc_o$  和  $ERDE_o$  的下标,决定着延迟做出决定的代价的高昂程度。图2为  $lc_5(k)$  和  $lc_{50}(k)$  的函数图像。 $c_{tp}$  通常被设置为与  $c_{fn}$  相同的值。在抑郁症数据集中,假设共有  $p$  个不同的个体,因此模型将做出  $p$  个决定,总体ERDE值将是这  $p$  个ERDE值的平均值。

图2 延迟成本函数  $lc_5(k)$  和  $lc_{50}(k)$ Fig.2 Latency cost functions  $lc_5(k)$  and  $lc_{50}(k)$ 

## 2 数据采集与预处理

### 2.1 社交媒体文本数据采集

社交媒体文本数据主要来源于各社交媒体中用户发布的帖子和评论。研究者们用于抑郁症检测的数据一般是从Reddit、Twitter和新浪微博等平台上爬取或使用API获取。目前,常用的公开数据集较少,主要有RSDD

表1 常用公开数据集统计信息

Table 1 Statistics of common public datasets

数据集	抑郁 用户/个	对照 用户/个	抑郁用户 帖子数/条	对照用户 帖子数/条	申请地址
RSDD <sup>[16]</sup>	9 210	107 274	8 924 490	103 948 506	<a href="https://georgetown-ir-lab.github.io/emnlp17-depression/">https://georgetown-ir-lab.github.io/emnlp17-depression/</a>
ERiskD 2017 <sup>[17]</sup>	训练集	83	403	30 851	—
	测试集	52	349	18 706	
ERiskD 2018 <sup>[18]</sup>	训练集	135	752	49 557	<a href="https://tec.citius.usc.es/ir/code/eRisk.html">https://tec.citius.usc.es/ir/code/eRisk.html</a>
	测试集	79	741	40 665	
CLPD <sup>[19]</sup>	训练集	327	573	—	<a href="https://www.cs.jhu.edu/~mdredze/clpsych-2015-shared-task-evaluation/">https://www.cs.jhu.edu/~mdredze/clpsych-2015-shared-task-evaluation/</a>
	测试集	150	300	—	
MDDL <sup>[1]</sup>	1 402	>3 亿	292 564	100 亿	<a href="http://depressiondetection.droppages.com/">http://depressiondetection.droppages.com/</a>

(Reddit self-reported depression diagnosis)数据集<sup>[16]</sup>、ERisk(early risk prediction on the Internet)任务中的抑郁症早期检测数据集ERiskD 2017<sup>[17]</sup>和ERiskD 2018<sup>[18]</sup>、CLPsych 2015(computational linguistics and clinical psychology)共享任务中用于抑郁症检测任务的数据集CLPD<sup>[19]</sup>和由Shen等人利用Twitter API创建的抑郁症检测数据集MDDL<sup>[1]</sup>。上述数据集由用户发布的帖子集合构成,一般根据用户自我陈述的诊断(诸如“我已经被诊断为抑郁症”等)和人工审查进行标注。各数据集的统计信息见表1。

## 2.2 数据预处理

原始数据经过基础预处理和特征工程生成文本表示,然后输入机器学习模型进行分类检测。基础预处理一般包含数据清洗、分词和标准化等步骤,其意义在于减少词汇量和非重要信息所带来的干扰。特征工程旨在从原始语料或经过基础预处理的文本数据中生成计算机能够理解的数值化数据。

自然语言处理中的文本表示可分为基础特征表示、静态词嵌入和语境词嵌入,具体如图3所示。基础特征表示需人工构建特征以表示文本,通常与传统机器学习方法搭配使用,也可以作为深度学习的输入;静态词嵌入和语境词嵌入则一般与深度学习结合使用。

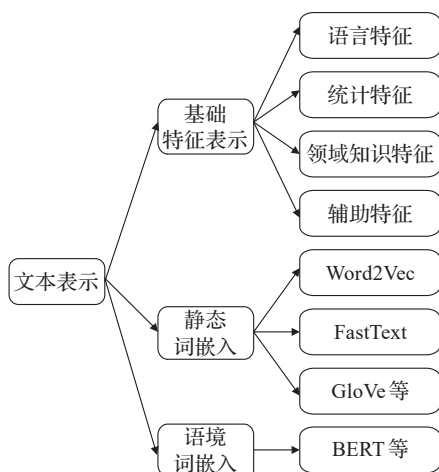


Fig.3 Classification of text representation

基础特征表示能够提取文本中的关键信息,甚至能够考虑到单词出现的次序,但是不能够结合上下文语义信息,而上下文语义信息在自然语言理解中至关重要。静态词嵌入的方法表达了单词原本的含义、单词相似度甚至是上下文关系,通常与深度神经网络配合使用,在自然语言处理中具有不错的效果。基于语境的词嵌入尽可能地学习单词的上下文语义,其极大规模的数据量、训练强度和模型容量,以及利用无监督模型的训练方式,使得它拥有强大的语言表征能力和特征提取能力,在多项自然语言处理任务中表现优异。

## 3 基于传统机器学习的抑郁症检测

传统机器学习利用社交媒体文本数据对用户进行抑郁症检测主要分为两个研究方向:基于不同特征的研究和基于不同机器学习算法的研究。基于不同特征检测抑郁症专注于发掘多样、可靠的特征,其使用的算法通常是诸如支持向量机等经典的单一算法;基于不同机器学习算法的研究则侧重于构建更复杂、综合的算法。

### 3.1 基于不同基础特征的检测

运用传统机器学习进行抑郁症检测前,需要从用户帖子中人为构建特征。不同基础特征及其特点如表2所示。其中语言特征能够显示抑郁症患者与心理健康者不同的语言风格,进而揭示两者不同的心理过程。常用的语言特征是语言探索与字词计数(linguistic inquiry

表2 不同基础特征及其特点

Table 2 Various basic features and their characteristics

基础特征	代表用例	特点
语言特征	LIWC等	拥有解释抑郁的能力,使用简便,但不适合社交媒体等口语文档
统计特征	BOW、TF-IDF、N-Gram等	能够充分利用关键字的原始含义且通用性较强,但是仅能从频率和词序方面表示文本,无法很好地利用上下文信息
领域知识特征	主题和情绪等	与抑郁症领域强相关,能够有效解释患者和健康用户的差异
辅助特征	用户行为和生活方式等	通常作为补充信息与其他特征结合使用



and word count, LIWC)。LIWC将文本中的单词与特定词典进行比对,从而输出单词的类别和词频。Nguyen等人<sup>[20]</sup>证明了LIWC在帖子级别预测抑郁症时显示出强大的指示力。Fatima等人<sup>[21]</sup>利用LIWC对抑郁症帖子和非抑郁症帖子进行了较好的区分。

语言特征提供了解释抑郁的能力,且仅通过分析单词语义就能够使用,但是它更适合于新闻或文章等正式文档,而非社交媒体帖子等非正式或口语文档。与基于语言模式的方法相比,基于词袋(bag of words, BOW)和词频逆向文件词频(term frequency inverse document frequency, TF-IDF)等统计特征通过统计单词频率,从而充分利用关键字的原始含义<sup>[22]</sup>且通用性更强。Prieto等人<sup>[23]</sup>使用简单的词袋模型,提取  $N$ -Gram特征并应用基于相关性的特征选择后进行抑郁症的检测,实现了较好的分类精度和速度提升。Dos Santos等人<sup>[12]</sup>则发现TF-IDF可以从非常小的数据集中做出潜在有用的预测。

对于精神疾病的检测而言,诸如主题和情绪等领域知识特征显示出良好的有效性。通常,抑郁症患者与心理健康用户所感兴趣的主题有所不同,因此可根据谈论主题的区别将两类人群进行有效区分。例如,Nguyen等人<sup>[24]</sup>发现主题和语言心理特征是高度有效的预测因子,联合两种特征在帖子级别检测抑郁症,达到了很好的效果。基于情绪的特征则能从更抽象的情绪方面提供信息且更具相关性,同样可以有效地揭示抑郁症患者和心理健康用户之间的差异。例如,Chen等人<sup>[25]</sup>在LIWC的基础上加入一组细粒度情感特征,证明了情感特征的有效性。Leiva等人<sup>[26]</sup>引入TF-IDF的同时,还引入了三情感极性特征(积极、中性、消极情绪),证明包含情感分析的方法比仅利用TF-IDF的方法更准确。

除了利用语言、统计和领域知识特征外,不少学者对辅助特征进行了探索。辅助特征例如用户的行为特征和生活模式特征等,通常作为上述特征的补充,能够从更为现实和细致的角度将抑郁症用户和健康用户进行对比,并且可利用的信息也更加全面。Hu等人<sup>[27]</sup>在语言特征的基础上加入行为特征,并比较不同时间观察窗口下模型的分类精度,发现语言和行为特征可以准确识别用户是否抑郁,而在观察时间为2个月时,效果最好。Chen等人<sup>[25]</sup>组合LIWC和生活模式特征,证明了组合特征的有效性。

整体看来,在基于社交媒体文本数据的抑郁症检测中,最原始的单一特征往往缺乏足够的信息,因而更多的特征被不断探索和加入。在综合的特征下,用户的各种信息能够得到利用,但是过多的甚至冗余的特征又会使模型运行效率下降。因此在利用传统机器学习方法进行抑郁症检测的领域中,构建何种特征以及如何选择具有代表性的特征仍然是一个重要问题;此外,如何构

建合适的学习算法以和选择的特征相匹配,从而使模型发挥更好的性能,也是值得考虑的问题。

### 3.2 基于不同算法类型的检测

在机器学习中,特征的构建和选择至关重要,而学习算法的选择和改进同样举足轻重,二者相辅相成。在基于社交媒体文本数据的抑郁症检测中,研究者们对于算法的研究旨在匹配多种特征以提高检测性能,解决标记数据量少和不支持增量学习等现实问题,以及进行抑郁症的早期检测等。

综合的特征能够较为全面地包含抑郁症用户的信息,但是并不是所有的学习算法都能够与之进行匹配而发挥出良好的效果。为此,许多学者进行了探索。例如, Peng等人<sup>[28]</sup>基于用户档案特征、用户行为特征和帖子文本特征,提出使用多核支持向量机进行抑郁文本分类。多核支持向量机能够针对不同特征自适应选择最优核,因而相比于单一核的支持向量机性能更好。尽管多核支持向量机性能表现良好,但仍存在一些限制,比如不适合更大的数据集,对缺失数据更敏感等。而集成学习能够克服单一分类器的局限,从而在检测性能和泛化性上得到提升。例如, Liu等人<sup>[29]</sup>使用特征选择方法,将多个单一分类器作为基学习器,并将逻辑回归作为组合策略来构建堆叠模型。提出的模型既能够降低数据维度,提高模型效率,又克服了单一模型自身的局限性,提升了模型的泛化性,在抑郁症患者识别中的准确率高达90.27%。

经典机器学习在社交媒体上识别抑郁症要么需要足够的历史数据,要么不支持增量学习。为解决这些问题, Tariq等人<sup>[30]</sup>采用联合了随机森林、支持向量机和朴素贝叶斯的半监督联合训练模型。提出的模型只需要少量的标记数据便可将大量未标记的数据进行标记,从而节省了大量的人力成本。Burdizzo等人<sup>[31]</sup>提出支持在文本流上进行增量训练的SS3模型,在抑郁症早期检测方面取得了先进的表现。SS3模型虽然表现突出,但是存在的一个缺陷是模型的输入部分使用词袋进行处理,因而无法考虑文本词序等问题。

经典的抑郁症检测方法时效性差,原因在于抑郁症检测需要患者首先能够意识到自身的心理问题,其次需要患者克服病耻感去求医,这一过程往往需要很长时间。通常患者被确诊为抑郁症时,已经到达严重的程度甚至存在自杀的倾向。考虑到这些问题,许多学者对抑郁症的早期检测进行了研究。Briand等人<sup>[32]</sup>认为来自新用户的帖子若在语义上接近风险用户的帖子,则新用户也可能处于患抑郁症的风险中。为此,构建了信息检索子系统和监督学习子系统,每个子系统的预测输出根据一种决策算法进行合并。提出的模型不仅能够检测现有用户的患病情况,而且能够尽早地对新增用户进行抑郁症的检测。Cacheda等人<sup>[10]</sup>提出双例方法进行抑郁

表3 抑郁症检测中的传统机器学习算法总结

Table 3 Summary of traditional machine learning algorithms for depression detection

主要算法	文献年份	数据来源	优点	局限性	评价指标
多核支持向量机 <sup>[28]</sup>	2019	新浪微博	融合多种特征;能够针对不同特征自动寻找最优核	没有考虑抑郁水平和动态状态随时间的变化	$Acc=0.834\ 6, P=0.755\ 6, R=0.766\ 9, F1=0.761\ 2$
特征选择和堆叠集成策略 <sup>[29]</sup>	2021	微博	消除冗余特征;能够弥补多个单一分类器的缺点	数据集小;抑郁用户年龄小,无法代表一般人群	$Acc=0.902\ 7, P=0.879\ 1, R=1.000\ 0, F1=0.935\ 7$
半监督联合训练 <sup>[30]</sup>	2019	Reddit	减少对标注数据的依赖,模型更具鲁棒性	未进行参数优化	$P=0.70, R=0.69, F1=0.70$
SS3 <sup>[31]</sup>	2019	ERiskD 2017	支持增量学习;具有可解释性;计算成本低	未考虑情绪变化等对于抑郁症症状的影响;未考虑文本的词序问题	$P=0.63, R=0.60, F1=0.61$
信息检索和监督学习系统 <sup>[32]</sup>	2018	ERiskD 2017	能够发掘新用户和风险用户帖子语义上的联系	搜索引擎中未添加与抑郁症相关特征的索引;分类器较为局限	$P=0.48, R=0.60, F1=0.53, ERDE_5=14.03\%, ERDE_{50}=12.29\%$
双例学习 <sup>[10]</sup>	2019	ERiskD 2017	奖惩检测效率,较单例方法性能更好	未考虑情感语义特征	$ERDE_5=11.88\%, ERDE_{50}=8.68\%$

症的早期检测。双例方法使用两个独立的随机森林分类器,一个用于检测抑郁个体,另一个用于识别非抑郁个体,两个选项(抑郁和非抑郁)独立预测,从而避免了单例方法中两选项相互竞争所造成的延迟。结果表明,双例方法的性能明显优于单例方法,并且能够将当前最先进的模型检测性能提高10%以上。

总体来看,在利用传统机器学习进行抑郁症检测上,特征的构建和选择已经较为全面和成熟,并且匹配多特征的算法也取得了良好的成果。但是当前研究对于标记数据量少等现实问题的探索较少,这在未来应当加强。此外,已有部分研究者对于抑郁症的早期检测进行探究,并且提出了新颖的方法,但是总体上,此类算法所取得的效果仍具有一定的提升空间。

抑郁症检测中的传统机器学习算法总结如表3所示。

#### 4 基于深度学习的抑郁症检测

传统机器学习需要人工构建大量特征,但是构建有效的特征往往会耗费研究者大量的时间和精力,而深度学习能够基于原始文本向量自动进行特征提取,并且拥有对事物进行抽象概括的能力。在许多情况尤其是拥有大量数据时,深度学习表现出优秀的性能。在基于社交媒体文本数据的抑郁症检测中,常用的深度学习算法有CNN、RNN,加入注意力组件的算法和基于Transformers的BERT等。

##### 4.1 基于CNN的抑郁症检测

在基于社交媒体文本数据的抑郁症检测中,CNN由于强大的特征抽取能力而被研究和应用。利用CNN进行抑郁症检测的基本框架如图4所示。文本数据通过词嵌入技术转化为数值化数据,形成词嵌入矩阵;然后利用多个不同大小的卷积核进行卷积操作;最后通过池化层和全连接层后输出为二分类结果。

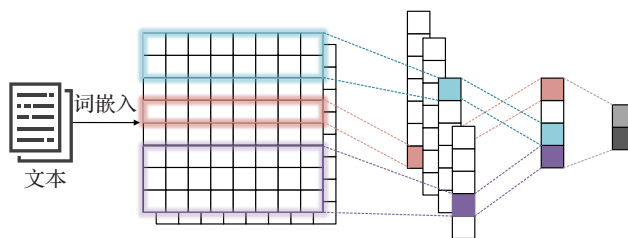


图4 基于CNN的抑郁症检测框架

Fig.4 Depression detection framework based on CNN

在应用中,Trotzek等人<sup>[33]</sup>将基于维基百科的FastText预训练词嵌入输入CNN,同时利用逻辑回归处理用户级语言元数据,最后将两者的输出进行简单融合进而分类。结果显示,构建的模型在抑郁症的早期检测中的综合性能最好。考虑到现实数据大多存在类别不平衡的问题,Kim等人<sup>[34]</sup>在CNN基础上加入SMOTE(synthetic minority oversampling technique),从而克服了数据的类别不平衡带来的性能损失。在利用CNN进行特征提取过程中,门控单元能够突出重要信息和剔除不重要信息,找出问题的关键影响因素和减少网络的参数量,使得模型性能进一步提升。Rao等人<sup>[35]</sup>在CNN中加入门控单元,结合门控单元的CNN拥有强特征抽取能力的同时可以过滤掉不重要的信息,因而模型能够选择性地捕捉用户帖子中的关键情绪信息,具有较强的检测性能和稳定性。

##### 4.2 基于RNN的抑郁症检测

CNN能够提取文本中的局部信息,且具有良好的并行计算能力,但是无法捕捉长距离的文本语义信息。相较于CNN,RNN由于引入了记忆单元而能够存储先前文本的信息,在文本数据的处理上具有一定优势。RNN的基本框架如图5所示。RNN单元按次序读取各个单词的词嵌入信息,其中 $h_i$ 为隐藏层的输出单元,包含了上一时间步 $h_{i-1}$ 的信息。传统的RNN存在梯度消失(gradient vanishing)问题,为此学者们提出RNN的变



体模型 LSTM(long short-term memory)和 GRU(gated recurrent unit),旨在解决传统 RNN 的梯度消失问题。

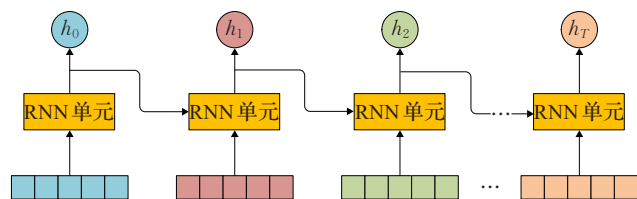


图5 RNN 基本框架

Fig.5 Basic framework of RNN

在利用 RNN 及其变体 LSTM 进行抑郁症检测中, Amanat 等人<sup>[36]</sup>构建了 RNN-LSTM 模型,证明了 RNN 性能优于 CNN。而与 LSTM 相比, BiLSTM 增加了对后文的训练,充分利用了前后文的语义信息,能够使序列分类问题的模型性能得以提升。Ahmad 等人<sup>[37]</sup>提出使用 BiLSTM 进行抑郁症检测,通过对比发现, BiLSTM 在各项指标上都优于 LSTM,但是未考虑数据类别不平衡问题; Cong 等人<sup>[38]</sup>构建了 X-A-BiLSTM 模型,发现在 BiLSTM 上使用 XGBoost 能够缓解数据不平衡问题。

#### 4.3 基于 CNN-RNN 和注意力机制的抑郁症检测

理论上, CNN-RNN 体系结构拥有 CNN 优秀的特征提取能力和 RNN 的序列建模能力,在基于社交媒体文本数据的抑郁症检测中,有学者对其进行了探索。Aragón 等人<sup>[39]</sup>将用户帖子的内容转化为子情绪序列,通过 CNN 提取特征后,使用双向门控循环单元(BiGRU)捕获子情绪序列的上下文,最后利用注意力机制提取句子中的重要子情绪。研究发现,提出的模型相较于单一的 CNN 和 RNN,精度提升了 7% 和 12%。同时,在数据规模较小时,标准 CNN 和 RNN 性能不及传统机器学习方法。Zogan 等人<sup>[40]</sup>构建由堆叠 BiGRU 和 CNN 与结合注意力的 BiGRU 组合的 DepressionNet 框架,其中堆叠 BiGRU 用于处理用户行为特征, CNN 与结合注意力的 BiGRU 用于提取用户帖子的摘要,通过将用户行为和用户发帖史进行融合来自动检测抑郁症。实验表明, CNN+BiGRU 模型已经达到不错的精度,而提出的模型相较于 CNN+BiGRU 在各项指标中至少提升了 2% 的性能。

在抑郁症检测中,注意力机制能够对信息进行权重分配,即赋予与抑郁症相关的重要信息更高的权重,从而使模型学习到用户帖子中包含的关键信息,提升检测性能。在社交媒体中,许多精神疾病患者倾向于通过隐喻等方式间接表达自己的感受和情绪<sup>[41-42]</sup>。鉴于此, Zhang 等人<sup>[43]</sup>提出 MAM(metaphor-based attention model)模型,试图通过发掘隐喻中的关键信息来更好地检测抑郁症。MAM 模型通过 RNN\_MHCA(recurrent neural network multi-head contextual attention)<sup>[44-45]</sup>模块获得句子隐喻和文本隐喻特征,然后基于隐喻特征计算注意

权重。实验表明,带有注意力的 MAM 模型能够学习到用户隐性情绪信息,并且证实了隐喻信息在抑郁症检测中的有效性。同样, Almars<sup>[46]</sup>提出使用注意机制来分析与抑郁症相关的阿拉伯语文本数据,在 BiLSTM 的基础上加入注意力机制,从而使模型学习到抑郁症的重要隐藏特征。相较于 BiLSTM,提出的模型在准确率方面提升了 3%。Ren 等人<sup>[47]</sup>提出包含注意力机制的 EAN(emotion-based attention network)模型。实验中, Ren 等人通过模型对比证明了注意力机制能够有效提升模型性能,并且证实了情感语义信息在抑郁症检测中的有效性。

注意力机制不仅能够提高模型性能,而且能通过可视化其权重分数,分析与抑郁症强相关的单词和句子,从而为发掘抑郁症的重要关联因素提供线索。Song 等人<sup>[48]</sup>提出的特征注意网络(feature attention network, FAN)综合了用户的抑郁症状、情感、反复性思考和写作风格特征,能够模拟专家对抑郁症进行诊断的过程。FAN 模型通过分析注意力权重来产生解释性,并证实了情感信息在抑郁症检测中的重要作用,但是模型的总体性能并不算杰出。Uban 等人<sup>[49]</sup>结合情感等信息,将层次注意网络(hierarchical attention networks, HAN)用于抑郁症的检测,最后通过分析网络层中数据的抽象表示等方法充分解释了模型预测。但是 HAN 模型更多地是对文本语言相关信息的考虑,而忽略了对用户行为、时间等特征的建模。Zogan 等人<sup>[50]</sup>提出基于 HAN 的混合模型 MDHAN(multi-aspect depression detection hierarchical attention network)。该模型结合文本、行为、时间和语义方面的特征,提高了预测性能,并通过分析注意力权重解释了模型预测方法,但是该模型尚缺乏对于情感的分析。

#### 4.4 基于 BERT 的抑郁症检测

Transformer 模型利用了自注意力编码器,能够自主发掘同一句子中各单词之间的相关性,从而获得更深层次的编码信息。此外, Transformer 完全抛弃了类似循环神经网络结构的使用,使得运算速度和对于长句的处理能力大幅提升,而基于两层双向 Transformers 的 BERT 预训练语言模型,更是具有强大的对语义信息建模的能力,其网络结构如图 6 所示。BERT 需将句子前后分别加入标识符作为分隔,然后将单词的位置信息、段落信息和单词嵌入作为两层 Transformer 编码器的输入。BERT 既可以作为一种词嵌入技术,也可在其后直接加上一个简单的分类器作为分类模型。

在抑郁症检测领域, Yadav 等人<sup>[51]</sup>率先提出一种新的基于 BERT 的多任务学习框架 FiLaMTL(figurative language enabled multi-task learning framework)。该框架能够通过检测比喻用法的辅助任务来准确识别抑郁症状。研究结果显示, BERT 具有强特征提取能力,但

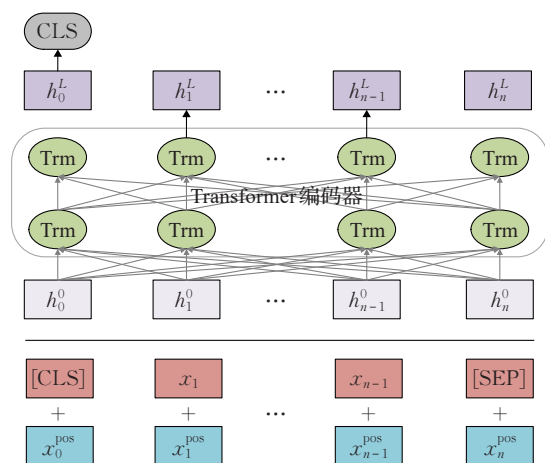


图6 BERT模型结构

Fig.6 Structure of BERT model

是在通用语料上训练的BERT不能够很好地适应特定领域。同时,实验结果也充分证明了引入比喻用法检测对抑郁症状识别的有效性。相比通用的预训练模型,领域内预训练能够学习到特定领域中数据的分布,往往在特定领域中表现更佳。Wang等人<sup>[52]</sup>运用BERT在抑郁症数据集进行领域内预训练(in-domain pretraining, IDP),发现在抑郁症检测和抑郁程度分类任务中,领域内预训练的BERT在所有提出的基于Transformers的模型中取得最佳性能。为解决经典BERT模型因体量巨大而难以在实际应用中部署等问题,Zeberga等人<sup>[53]</sup>提出了一个新的框架,该框架应用将知识从大型预训练网络(BERT)转移到小型网络(Distilled\_BERT)的知识蒸馏技术。相较于BERT,Distilled\_BERT不仅进一步提升了检测性能,而且模型的体量相对较小。在对结构进行了改进的BERT的应用中,Khan等人<sup>[54]</sup>采用DeBERTa(decoding-enhanced BERT with disentangled attention)模型进行抑郁症与其他疾病的区分。DeBERTa的改进之处在于引入了解耦注意力机制和增强型掩码解码器,因而能够同时考虑词汇的内容、相对位置与绝对位置信息,即充分地学习了单词的内容及其依赖关系,在与多个先进模型的对比中,该模型在区分抑郁症与其他疾病方面表现最佳。

综上,在利用深度学习模型进行抑郁症检测的研究中,研究者们从平衡数据类别、特征提取方法和结合多维度特征等角度进行了探索并取得了较好效果。总体看来,相较于传统机器学习,深度学习由于能够自动提取特征而具有更强的稳定性和泛化性,且能够达到更杰出的检测性能。但是深度学习模型的参数量相对较大,且往往需要大规模数据的支撑,在小数据集上深度学习的性能可能不及传统机器学习。在深度学习方法上,需要关注的是注意力机制和BERT预训练模型。注意力机制能够提升模型性能,并且能够为模型预测提供一定的解释性,具有在临床进行应用的潜力。BERT类模型虽然具有强特征提取能力,能够提取文本中表示抑郁的

关键信息,从而达到可观的性能。但是其结构较为复杂,模型参数量巨大,不利于进行重新训练。而使用通用的预训练BERT模型又势必会造成性能上的损失,尤其是在抑郁症这类具有医学特点的领域。

抑郁症检测中的深度学习算法总结如表4所示。

## 5 总结与展望

社交媒体日益成为人们情感表达的平台,抑郁症等心理疾病也逐渐成为人们关注的焦点,从社交媒体用户发布的文本信息中寻找抑郁症的线索,已被诸多学者探索和研究。本文基于上述文献总结当前研究的不足并大胆地对未来研究方向进行展望。

### 5.1 现有研究所面临的问题

(1)缺乏中文数据集。数据是进行科学研究的基础,而当前国内尚缺乏大型公开公认的社交媒体中文抑郁症数据集,这在一定程度上限制了国内抑郁症领域的研究和发展。

(2)模型对于抑郁症的解释不够深入。虽然当前已有许多研究者致力于研究模型的解释性,但其是以分析注意力权重为主。此类分析仅能够展示与抑郁症强相关的单词和句子,而不能揭示抑郁症的发病机理以及模型的推理过程。

(3)缺乏基于隐喻的抑郁症检测研究。患有抑郁症等精神疾病的人群在隐喻的用词上与普通人群有所差异,当前也有少数研究证实了发掘隐喻等表达在区分抑郁症患者和普通人群中的有效性。总体来看,目前基于隐喻的检测是一种较新的思路和方法,拥有较大的研究空间。

(4)BERT模型训练成本高。当前的BERT模型普遍存在着参数量巨大、对数据量要求高和耗费资源大等问题,这些缺陷使得研究者只能在公开的、已训练完成的模型上进行微调,而没有充足的资源进行从头训练,从而难以对模型本身做出改进和提出适用于抑郁症领域的高精度模型。

### 5.2 未来研究展望

(1)中文数据集的创建可以借鉴国外数据集构建的思路,即可以通过在微博等社交媒体平台结合自动筛选用户自我诊断的陈述和人工审查的方式创建中文数据集。此外,在数据集样本标注较少的情况下,构建弱监督学习方法进行抑郁症检测将是重要的研究方向。

(2)当前构建的模型多以数据为驱动,这样的模型难以去深入发掘内部的运行过程,而如果把模型嵌入抑郁症知识,让模型去学习人类进行知识推理的过程,或许能够清晰地解释模型及其运行结果。因此,将抑郁症知识,例如以知识图谱的形式与深度学习方法进行结合,从而构建具有解释性的抑郁症检测模型,是非常具有现实意义的方向。

表4 抑郁症检测中的深度学习算法总结

Table 4 Summary of deep learning algorithms for depression detection

算法类别	算法结构	文献年份	数据来源	优点	局限性	评价指标
CNN	CNN <sup>[33]</sup>	2018	ERiskD 2017	加入用户级语言元数据,提高了模型性能	融合策略陈旧	$P=0.71, R=0.71, F1=0.71$
	CNN+SMOTE <sup>[34]</sup>	2020	Reddit	缓解数据类别不平衡问题	未考虑用户纵向行为模式的时间序列信息	$Acc=0.7513, P=0.8910, R=0.7175, F1=0.7949$
	MGL-CNN <sup>[35]</sup>	2020	RSDD	加入门控单元,有效捕捉帖子中的关键情绪信息	无法捕捉远距离单词之间的联系	$P=0.63, R=0.48, F1=0.54$
RNN	RNN-LSTM <sup>[36]</sup>	2022	Kaggle	较好地处理文本序列信息,性能优于CNN	数据集局限,不能充分反映抑郁症过程	$Acc=0.997, P=0.990, R=0.980, F1=0.990$
	BiLSTM <sup>[37]</sup>	2020	Orabi等人的数据集 <sup>[55]</sup> (含CLPsych数据)	更好地理解前后文语义,性能优于单向LSTM	有限的数据集,未进行数据类别平衡	$Acc=0.93, P=0.89, R=0.91, F1=0.90$
	X-A-BiLSTM <sup>[38]</sup>	2018	RSDD	在BiLSTM中缓解数据集不平衡问题	长文本时信息损失严重	$P=0.69, R=0.53, F1=0.60$
CNN-RNN和注意力机制	CNN+BiGRU+Att <sup>[39]</sup>	2020	ERiskD 2018	能充分捕捉抑郁症患者的细粒度情绪	帖文信息利用不充分	$F1=0.58$
	DepressionNet <sup>[40]</sup>	2021	MDDL	通过提炼摘要压缩信息	模型结构较为复杂,未考虑帖子中的URL内容	$Acc=0.901, P=0.909, R=0.904, F1=0.912$
	MAM <sup>[43]</sup>	2021	ERiskD 2018	创新地提出基于隐喻的方法	模型中的RNN模块不够先进;未能解释隐喻与精神疾病之间的内在联系	$P=0.52, R=0.50, F1=0.51$
	BiLSTM+Att <sup>[46]</sup>	2022	Twitter(阿拉伯语)	有效学习抑郁症的重要隐藏特征	数据集小,模型泛化性弱	$Acc=0.83, P=0.78, R=0.83, F1=0.81$
	EAN <sup>[47]</sup>	2021	Reddit	能够同时捕获上下文语义信息和情感语义信息	情绪分析不够精细	$Acc=0.913, P=0.919, R=0.962, F1=0.940$
	FAN <sup>[48]</sup>	2018	RSDD	模拟专家诊断过程,具有较强可解释性	未纳入用户更多帖子而导致模型性能不高	$P=0.61, R=0.52, F1=0.56$
	HAN <sup>[49]</sup>	2021	ERiskD 2018	具有可解释性	未考虑用户行为、时间等特征	$F1=0.71$
BERT	MDHAN <sup>[50]</sup>	2022	RSDD	考虑了多方面综合特征,具有可解释性	未能结合主题和情感对用户帖文进行分析	$Acc=0.895, P=0.902, R=0.892, F1=0.893$
	FiLaMTL <sup>[51]</sup>	2020	CLPsych 2015	引入比喻用法的检测来提升模型的鲁棒性和可靠性	模型结构较为复杂;未引入模因模态而导致模型对于比喻的理解尚不够深入	$Acc=0.7079$
	BERT_IDP <sup>[52]</sup>	2020	中国微博	具有强特征提取能力,进行了抑郁症程度的划分	未考虑用户级上下文;未加入抑郁症的医学知识	$MicroF1=0.856$
	BERT+知识蒸馏 <sup>[53]</sup>	2022	Reddit+Twitter	应用知识蒸馏技术压缩模型参数量	未考虑用户级上下文	$Acc=0.97$
	DeBERTa <sup>[54]</sup>	2022	Twitter	能够充分地学习单词的内容及其依赖关系	模型参数量大	$P=0.85, R=0.88, F1=0.86$

(3)目前基于隐喻的抑郁症检测尚处于起步阶段,隐喻特征与抑郁症等精神疾病的内在联系还有待进一步论证和研究,并且如何构建模型以发掘隐喻特征也应当予以大量研究。

(4)**BERT模型具有强特征提取能力,但是因其训练成本过高而限制了在领域中的应用。因此,在保证精度的前提下,探索更精简、效率更高的BERT模型或者其他预训练模型,是未来应继续重点关注的话题。**

参考文献:

[1] SHEN G, JIA J, NIE L, et al. Depression detection via har-

vesting social media: a multimodal dictionary learning solution[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017: 3838-3844.

[2] FRIEDRICH M J. Depression is the leading cause of disability around the world[J]. JAMA, 2017, 317(15): 1517.

[3] 区健新, 吴寅, 刘金婷, 等. 计算精神病学: 抑郁症研究和临床应用的新视角[J]. 心理科学进展, 2020, 28(1): 111-127.

OU J X, WU Y, LIU J T, et al. Computational psychiatry: a new perspective on research and clinical applications in depression[J]. Advances in Psychological Science, 2020, 28(1): 111-127.

[4] PARK M, CHA C, CHA M. Depressive moods of users



- portrayed in Twitter[C]//Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining, 2012: 1-8.
- [5] DE CHOUDHURY M, GAMON M, COUNTS S, et al. Predicting depression via social media[C]//Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, 2013.
  - [6] SAXENA S, FUNK M, CHISHOLM D. World health assembly adopts comprehensive mental health action plan 2013—2020[J]. The Lancet, 2013, 381(9882): 1970-1971.
  - [7] ZIEMER K S, KORKMAZ G. Using text to predict psychological and physical health: a comparison of human raters and computerized text analysis[J]. Computers in Human Behavior, 2017, 76: 122-127.
  - [8] WONGKOLAP A, VADILLO M A, CURCIN V. Researching mental health disorders in the era of social media: systematic review[J]. Journal of Medical Internet Research, 2017, 19(6).
  - [9] ISLAM M, KABIR M A, AHMED A, et al. Depression detection from social network data using machine learning techniques[J]. Health Information Science and Systems, 2018, 6.
  - [10] CACHEDA F, FERNANDEZ D, NOVOA F J, et al. Early detection of depression: social network analysis and random forest techniques[J]. Journal of Medical Internet Research, 2019, 21(6): e12554.
  - [11] DE SOUZA FILHO E M, REY H C V, FRAJTAG R M, et al. Can machine learning be useful as a screening tool for depression in primary care?[J]. Journal of Psychiatric Research, 2021, 132: 1-6.
  - [12] DOS SANTOS W R, FUNABASHI A M M, PARABONI I. Searching Brazilian Twitter for signs of mental health issues[C]//Proceedings of the 12th International Conference on Language Resources and Evaluation, 2020: 6111-6117.
  - [13] WU M Y, SHEN C Y, WANG E T, et al. A deep architecture for depression detection using posting, behavior, and living environment data[J]. Journal of Intelligent Information Systems, 2020, 54(2): 225-244.
  - [14] 董健宇, 韦文棋, 吴珂, 等. 机器学习在抑郁症领域的应用[J]. 心理科学进展, 2020, 28(2): 266-274.  
DONG J Y, WEI W Q, WU K, et al. The application of machine learning in depression[J]. Advances in Psychological Science, 2020, 28(2): 266-274.
  - [15] LOSADA D E, CRESTANI F. A test collection for research on depression and language use[C]//International Conference of the Cross-Language Evaluation Forum for European Languages. Cham: Springer, 2016: 28-39.
  - [16] YATES A, COHAN A, GOHARIAN N. Depression and self-harm risk assessment in online forums[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017: 2968-2978.
  - [17] LOSADA D E, CRESTANI F, PARAPAR J. eRISK 2017: CLEF lab on early risk prediction on the Internet: experimental foundations[C]//International Conference of the Cross-Language Evaluation Forum for European Languages. Cham: Springer, 2017: 346-360.
  - [18] LOSADA D E, CRESTANI F, PARAPAR J. Overview of eRisk: early risk prediction on the Internet[C]//Proceedings of the 9th International Conference of the Cross-Language Evaluation Forum for European Languages, 2018: 343-361.
  - [19] COPPERSMITH G, DREDZE M, HARMAN C, et al. CLPsych 2015 shared task: depression and PTSD on Twitter[C]//Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology, 2015: 31-39.
  - [20] NGUYEN T, VENKATESH S, PHUNG D. Textual cues for online depression in community and personal settings[C]//Proceedings of the 12th International Conference on Advanced Data Mining and Applications, 2016: 19-34.
  - [21] FATIMA I, MUKHTAR H, AHMAD H F, et al. Analysis of user-generated content from online social communities to characterise and predict depression degree[J]. Journal of Information Science, 2018, 44(5): 683-695.
  - [22] HAN S, HUANG H, TANG Y. Knowledge of words: an interpretable approach for personality recognition from social media[J]. Knowledge-Based Systems, 2020, 194: 105550.
  - [23] PRIETO V M, MATOS S, ALVAREZ M, et al. Twitter: a good place to detect health conditions[J]. PLoS One, 2014, 9(1): e86191.
  - [24] NGUYEN T, O'DEA B, LARSEN M, et al. Using linguistic and topic analysis to classify sub-groups of online depression communities[J]. Multimedia Tools and Applications, 2017, 76(8): 10653-10676.
  - [25] CHEN X, SYKORA M D, JACKSON T W, et al. Tweeting your mental health: exploration of different classifiers and features with emotional signals in identifying mental health conditions[C]//Proceedings of the 51st Annual Hawaii International Conference on System Sciences, 2018: 3320-3328.
  - [26] LEIVA V, FREIRE A. Towards suicide prevention: early detection of depression on social media[C]//Proceedings of the 4th International Conference on Internet Science, 2017: 428-436.
  - [27] HU Q, LI A, HENG F, et al. Predicting depression of social media user on different observation windows[C]//Proceedings of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2015: 361-364.
  - [28] PENG Z, HU Q, DANG J. Multi-kernel SVM based depression recognition using social media data[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(1): 43-57.
  - [29] LIU J, SHI M. A hybrid feature selection and ensemble approach to identify depressed users in online social media[J]. Frontiers in Psychology, 2021, 12: 802821.
  - [30] TARIQ S, AKHTAR N, AFZAL H, et al. A novel co-training-based approach for the classification of mental illnesses

- using social media posts[J].IEEE Access,2019,7:166165-166172.
- [31] BURDISO S G,ERRECALDE M,MONTES-Y-GOMEZ M. A text classification framework for simple and effective early depression detection over social media streams[J]. Expert Systems with Applications,2019,133:182-197.
- [32] BRIAND A,ALMEIDA H,MEURS M J.Analysis of social media posts for early detection of mental health conditions[C]//Proceedings of the 31st Canadian Conference on Artificial Intelligence,2018:133-143.
- [33] TROTZEK M,KOITKA S,FRIEDRICH C M.Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences[J].IEEE Transactions on Knowledge and Data Engineering,2018, 32(3):588-601.
- [34] KIM J,LEE J,PARK E,et al.A deep learning model for detecting mental illness from user content on social media[J].Scientific Reports,2020,10(1):1-6.
- [35] RAO G,ZHANG Y,ZHANG L,et al.MGL-CNN:a hierarchical posts representations model for identifying depressed individuals in online forums[J].IEEE Access,2020,8: 32395-32403.
- [36] AMANAT A,RIZWAN M,JAVED A R,et al.Deep learning for depression detection from textual data[J].Electronics,2022,11(5):676.
- [37] AHMAD H,ASGHAR M Z,ALOTAIBI F M,et al.Applying deep learning technique for depression classification in social media text[J].Journal of Medical Imaging and Health Informatics,2020,10(10):2446-2451.
- [38] CONG Q,FENG Z,LI F,et al.X-A-BiLSTM:a deep learning approach for depression detection in imbalanced data[C]// Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine-Human Genomics, 2018:1624-1627.
- [39] ARAGÓN M E,LÓPEZ-MONROY A P,GONZÁLEZ L C, et al.Attention to emotions:detecting mental disorders in social media[C]//Proceedings the 23rd of International Conference on Text,Speech,and Dialogue,2020:231-239.
- [40] ZOGAN H,RAZZAK I,JAMEEL S,et al.DepressionNet: a novel summarization boosted deep framework for depression detection on social media[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval,2021:133-142.
- [41] MAGAÑA D.Cultural competence and metaphor in mental healthcare interactions:a linguistic perspective[J].Patient Education and Counseling,2019,102(12):2192-2198.
- [42] LLEWELLYN-BEARDSLEY J,RENNICK-EGGLESTONE S,CALLARD F,et al.Characteristics of mental health recovery narratives:systematic review and narrative synthesis[J].PLoS one,2019,14(3):e0214678.
- [43] ZHANG D,SHI N,PENG C,et al.MAM:a metaphor-based approach for mental illness detection[C]//Proceedings of the 21st International Conference on Computational Science,2021:570-583.
- [44] GONG H,GUPTA K,JAIN A,et al.IlliniMet:illinois system for metaphor detection with contextual and linguistic information[C]//Proceedings of the 2nd Workshop on Figurative Language Processing,2020:146-153.
- [45] MAO R,LIN C,GUERIN F.End-to-end sequential metaphor identification inspired by linguistic theories[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,2019:3888-3898.
- [46] ALMARS A M.Attention-based Bi-LSTM model for Arabic depression classification[J].Computers Materials & Continua,2022,71(2):3091-3106.
- [47] REN L,LIN H,XU B,et al.Depression detection on reddit with an emotion-based attention network:algorithm development and validation[J].JMIR Medical Informatics, 2021,9(7):e28754.
- [48] SONG H,YOU J,CHUNG J W,et al.Feature attention network:interpretable depression detection from social media[C]//Proceedings of the 32nd Pacific Asia Conference on Language,Information and Computation,2018.
- [49] UBAN A S,CHULVI B,ROSSO P.On the explainability of automatic predictions of mental disorders from social media data[C]//Proceedings of the 26th International Conference on Applications of Natural Language to Information Systems,2021:301-314.
- [50] ZOGAN H,RAZZAK I,WANG X,et al.Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media[J].World Wide Web, 2022,25(1):281-304.
- [51] YADAV S,CHAUHAN J,SAIN J P,et al.Identifying depressive symptoms from tweets:figurative language enabled multitask learning framework[C]//Proceedings of the 28th International Committee on Computational Linguistics, 2020:696-709.
- [52] WANG X,CHEN S,LI T,et al.Depression risk prediction for Chinese microblogs via deep-learning methods:content analysis[J].JMIR Medical Informatics,2020,8(7): e17958.
- [53] ZEBERGA K,ATTIQUE M,SHAH B,et al.A novel text mining approach for mental health prediction using Bi-LSTM and BERT model[J].Computational Intelligence and Neuroscience,2022:7893775.
- [54] KHAN P I,RAZZAK I,DENGEL A,et al.Performance comparison of transformer-based models on twitter health mention classification[J].IEEE Transactions on Computational Social Systems,2022.DOI:10.1109/TCSS.2022. 3143768.
- [55] ORABI A H,BUDDHITHA P,ORABI M H,et al.Deep learning for depression detection of Twitter users[C]// Proceedings of the 5th Workshop on Computational Linguistics and Clinical Psychology:From Keyboard to Clinic, 2018:88-97.