

在线心理健康社区用户评论亲密度水平测度研究 ——基于文本挖掘的方法*

孙佳坤¹ 顾东晓^{1,2} 黄正行³ 苏凯翔¹ 朱凯旋¹ 李敏¹

(1.合肥工业大学 管理学院, 安徽 合肥 230009)

(2.数据科学与智慧社会治理教育部哲学社会科学实验室, 安徽 合肥 230009)

(3.浙江大学人工智能研究所, 浙江 杭州 310008)

摘 要 亲密度是在线心理健康社区成员之间语言和情感交流的关键要素。现有研究主要使用语言查询和词频统计 (linguistic inquiry and word count, LIWC) 词典计算亲密度, 但该方法容易丢失语义信息, 并且存在词汇遗漏和准确性低的局限性。本文基于文本挖掘的方法, 构建了一个具有可解释性的用户评论亲密度水平测度模型。该模型利用中文 BERT 提取文本语义特征, 并引入可解释性模块和文本卷积神经网络 (text convolutional neural network, TextCNN) 模块, 赋予模型高度的可解释性和分类能力。通过对知乎“心理健康”话题下的评论进行实验评估, 验证了模型的有效性。

关键词 在线心理健康社区, 用户评论, 文本挖掘, 亲密度, 可解释性

中图分类号 C931.6

1 引言

中国精神卫生调查显示, 我国精神障碍的终身患病率为 16.6%, 意味着有 2.3 亿人一生中会罹患心理健康疾病, 心理健康问题已成为中国面临的重要挑战^[1]。由于在线预约心理医生较困难且费用较高, 越来越多的人选择通过在线心理健康社区 (online mental health communities, OMHC) 发布问题帖子寻求帮助^[2]。在这些社区中, 拥有相同或相似经历的成员往往会积极发表评论, 为发帖人提供情感或信息支持^[3, 4]。在这种互动过程中, 评论的亲密度水平不仅影响发帖人接受支持的体验, 还可能对其心理状态产生直接影响。

低亲密度的评论往往难以有效缓解发帖者的心理健康问题^[5, 6], 甚至可能降低他们再次寻求心理健康帮助的意愿。在在线心理健康社区中, 这类评论通常表现得冷淡或机械化, 难以建立情感纽带。研究表明, 低亲密度评论可能削弱社区成员参与互动的积极性^[7], 甚至促使社区成员转为潜伏状态, 从而对社区的活跃度和整体氛围造成不利影响。相反, 高亲密度的评论表明社区成员更愿意分享个人的心理健康问题和情感经历^[8], 这使发帖者更倾向于与社区成员建立深厚的情感联系, 并体验到更为愉悦的交流过程^[9]。这种亲密感不仅有助于缓解发帖者的孤独感, 还能增强他们对他人的信任和依赖, 从而进一步促进心理健康^[10, 11]。因此, 实现评论亲密度水平的自动化测量, 不仅能够帮助社区管理员快速筛选出高

* 基金项目: 国家自然科学基金项目 (72271082、72071063、71771077)。

通信作者: 朱凯旋, 合肥工业大学管理学院, 博士, E-mail: kaixuan_zhu@mail.hfut.edu.cn。

亲密度的评论,还可以及时识别出可能存在心理健康问题的社区成员,为其提供适当的支持和建议。此外,这一测量还可以为社区成员量身打造推荐工具,根据亲密度水平向社区成员推荐匹配的帖子,进一步优化社区互动体验和服务质量。

在这一背景下,如何准确识别和分析评论的亲密度水平,成为研究在线心理健康社区互动模式的重要课题。以往研究主要采用 LIWC 词典来测量文本的亲密度^[12-14]。该方法通过将文本中的词与词典中的亲密度词进行匹配,统计其出现频率来计算亲密度。然而,这种方法存在以下局限性:①缺乏对文本语义特征的全面考虑,仅依赖词典匹配,忽视了词在不同语境中可能呈现出不同的含义或词性;②随着在线心理健康社区的快速发展,许多新兴词汇频繁出现并被广泛使用,然而 LIWC 词典往往未能及时更新,导致对亲密度词汇的遗漏,影响了亲密度测量的准确性。

为克服以往研究的局限性,从而更准确地测量评论的亲密度水平。本文基于文本挖掘的方法,构建了一个具有可解释性的用户评论亲密度水平测度模型。相比传统依赖 LIWC 词典的测量方法,该模型不仅充分考虑了文本的深层语义特征,还具备较强的可解释性,有效解决了传统方法中亲密度测量存在的词汇遗漏和准确率低的问题。本文使用一个人工标注的中文评论数据集,共包含 2962 条用户评论,由三名健康信息学专业的研究生完成标注工作。评论的亲密度水平测度被定义为三分类任务,分别是低、中、高三类。具体流程如下:首先,利用中文 BERT 模型生成文本的词向量表示;其次,通过可解释性模块中的标签感知注意力层识别评论中对预测贡献较大的词汇,赋予模型可解释能力。由于亲密度词汇通常由 2 至 4 个字组成(如“抱抱你”“抑郁”“兄弟姐妹”等),利用 TextCNN 在文本卷积上的优势,提取语义关系并进行分类^[15]。这样整合可解释性模块与 TextCNN 模块,可以实现对评论亲密度水平的有效测度。

综上所述,本文通过引入文本的深层语义特征和可解释性机制,克服了传统词典方法对语境动态变化和新兴词汇适应性不足的局限,同时提高了亲密度测量的准确性和可解释性,为在线心理健康社区的互动分析提供了更加全面可靠的工具支持。本文在方法和理论方面的主要贡献总结如下。

(1) 本文基于文本挖掘的方法,构建了一个具有可解释性的用户评论亲密度水平测度模型。该模型可以自动识别出高、中、低亲密度的评论,并且具有可解释性。本文使用中文 BERT 模型获取文本的词向量表示。然后输入可解释性模块和 TextCNN 模块进行分类器的训练。实验结果表明,该模型的性能优于其他基线分类模型,优于 LIWC 词典的方法。

(2) 本文设计了一个可解释性模块。该模块中的标签感知注意力层能够识别出对标签预测贡献更为重要的输入序列的特定部分,从而更直观地展现模型的预测过程。这不仅提高了模型的分类性能,还增强了模型的可解释性,使研究人员和从业者能够更深入地了解模型的决策过程。

(3) 本文采用词向量训练和相似度计算的方法,对 LIWC 词典中的亲密度词进行了扩充,最终得到 5890 个不同等级的亲密度词汇,这些亲密度词汇不仅可供其他研究者参考使用,而且拓展了语言表达的广度,能够帮助人们在不同的情境中更好地表达爱情、亲情、友情等各种情感,使情感表达更为丰富和贴切。

本文内容安排如下:第 2 部分总结分析相关研究;第 3 部分描述所提出的模型;第 4 部分通过实验评估所提模型的性能;第 5 部分是对实验结果进行展示和分析;第 6 部分是研究的理论与实践意义;第 7 部分总结全文。

2 文献综述

2.1 在线心理健康社区

随着互联网技术的普及,在线心理健康社区的快速发展不仅提供了大量的心理医疗健康数据,还改变了人们获取心理健康信息的方式^[16,17]。尤其是远程心理医疗和互联网心理医疗等领域的发展,使更多人通过在线途径获取心理健康方面的信息和服务^[18]。现有研究主要围绕以下三个维度展开:首先,远程医疗与文本挖掘技术的融合,为心理健康数据的结构化分析提供了技术基础。例如,Win Myint 等基于 Transformer 架构开发的多任务学习模型,实现了情感分类与情绪识别的联合优化,为社区评论的情感分析提供了技术支持^[19]。顾东晓等进一步指出,大模型的“涌现”效应为医疗领域的知识管理带来了新的机遇,推动了智能知识管理体系的创新与优化^[20]。其次,在线社区通过降低信息获取门槛显著提升了健康服务的普惠性。沈彦琪等指出在线健康信息的替代性搜寻对于老年人健康素养的提升具有积极作用,尤其在弥合数字鸿沟方面取得了显著成效^[21]。邓朝华等基于价值共创视角,研究发现在线健康平台能够促进价值共创,有助于缓解健康不公平问题,为不同社会群体提供更加公平的健康信息获取渠道^[22]。值得注意的是,Ke 等揭示的健康焦虑群体信息检索注意力偏差,暗示了用户心理状态与在线互动行为的复杂关联^[23]。最后,医生在在线医疗社区中的表现成为新的研究热点。孙士伟等基于信任传递理论的分析发现,医生职称和在线声誉对其在线医疗绩效有显著的正向影响^[24]。邓朝华等进一步指出,医生团队的结构与声誉不仅影响患者的择医决策,还进一步促进了在线医疗服务的发展^[25]。

综合来看,已有研究主要探讨了在线心理健康社区中的数据价值挖掘、用户行为及医患关系等问题,但鲜有专门讨评论亲密度水平对用户关系建立及心理健康影响的研究。本文将亲密度测度作为研究重点,构建具有可解释性的测度模型,旨在弥补现有研究在评论语义深层次分析和亲密度量化测度方面的不足,为在线心理健康社区的互动模式研究和心理健康干预提供创新方法和理论支持。

2.2 亲密度分析

亲密度作为一个复杂而多层次的研究领域,近年来受到了广泛的关注。通过回顾相关文献,本文将亲密度细分为两个关键领域:人际交往的亲密度和语言的亲密度。在人际交往的亲密度方面,研究主要关注人际交往中的信任、情感支持以及自我披露对亲密度的影响。Fan 和 Lederman 指出在基于文本的关系导向在线健康社区中,情感信任有助于促进亲密关系的形成^[26]。Ma 等做了一项关于内容亲密度与社交媒体上自我披露意愿之间关系的在线实验。研究结果显示,随着内容亲密度的增加,人们的自我披露意愿会减少^[27]。Ogba 等设计了在线自我披露问卷,采用 Hayes 提出的基于回归的 Process 模块进行适度分析,得出自我披露与亲密度呈正相关^[28]。Sheng 和 Kairam 通过研究 Twitch 的聊天环境发现,减少身体线索和强调共同点,能够促进自我披露和交流,从而在用户间建立紧密的关系^[29]。

在语言的亲密度方面,研究聚焦于分析文本内容所表达的亲密度。Gilbert 和 Karahalios 使用 LIWC 词典测量了亲密度,具体方法是将词典中的亲密度词与文本词进行匹配,并统计匹配词汇出现的频次^[12]。Seol 等提出了一种基于共同邻居相似度的社交网络服务亲密度测量方法,利用用户之间的链接关系进行亲密度测量^[30]。Reuel 等基于 LIWC 词典 LDA 主题分布、情绪类别等特征预测了在线健康社区交互文本中的亲密度^[13]。Zeng 等使用改进与拓展的中文 LIWC 词典来计算支持帖中的亲密度^[14]。Spiliotopoulos 和 Oakley 基于先前研究使用的情感分析词典,检测帖子中亲密性词汇的数量来测量亲密度^[31]。传统的

文本亲密度分析方法通常依赖于特征工程或词典匹配等技术。然而这些方法无法充分捕捉文本中更为深层的语义信息,导致亲密度测量的准确率低。相较之下,Pei 和 Jurgens 则采用了大规模文本挖掘的方法,并利用了 RoBERTa 预训练模型,成功实现了对英文社交媒体文本亲密度的计算,并取得了显著成果^[32]。

综上所述,现有研究在人际交往亲密度领域强调了信任、情感支持和自我披露对亲密度的影响,但对在线心理健康社区中互动评论亲密度的动态演变及其影响机制关注不足。在语言亲密度方面,传统词典匹配方法虽然提供了初步的分析路径,但难以充分全面挖掘深层语义信息。最新的大规模文本挖掘与预训练模型为亲密度分析带来新思路,但其在中文文本和特定场景中的应用仍有待完善。基于此,本文结合中文在线心理健康社区的特点,探索互动评论的亲密度测量方法,引入深度学习技术以弥补传统方法的不足,提升亲密度分析精准性,为用户行为分析与个性化服务优化提供支持。

2.3 文本挖掘方法

自然语言处理技术的进步推动了面向在线心理健康社区的文本挖掘技术的发展,主要包括词嵌入、主题提取、文本分类等核心领域。在词嵌入方面,一些研究人员使用 Word2Vec^[33-35]、ELMo^[36, 37]、BERT^[38, 39]等进行文本挖掘。Mikolov 等提出了改进后的 Skip-gram 模型,能够捕捉大量精确的句法和语义关系,并能够更快地学习到更高质量的词向量表示^[40]。Peters 等提出了 ELMo 词向量表示方法^[41]。相比 Word2Vec,ELMo 在训练中考虑了词语在不同语境下的多义性,但推断时只能单向处理上下文,可能导致某些情况下无法捕捉到完整的语境信息。Devlin 等提出了 BERT 预训练模型,该模型通过使用双向上下文信息来预训练词汇表中的词语,从而更好地捕捉语言的复杂语境和语义关系^[42]。因此,可采用 BERT 模型来获取文本的深层语义信息,将其转化为词向量矩阵,从而更全面地捕捉复杂的语境和语义关系。

在文本分类方面,一些研究使用深度学习模型对在线心理健康社区文本内容进行分类,取得了很好的效果,主要包括 BiLSTM^[43, 44]、Transformer^[45, 46]、TextCNN^[47, 48]。BiLSTM 通过引入双向结构和“门”机制来控制信息流动,能够同时处理文本的前向和后向依赖,捕捉上下文中的长期依赖关系,因此更适合处理包含复杂语境的长文本输入。Transformer 是一种基于自注意力机制的深度学习模型,能够同时处理整个序列,适用于处理长文本输入。TextCNN 可以通过在单词向量的维度上对文本进行卷积来提取语义关系,在处理短文本时通常表现良好。

综上所述,自然语言处理技术的进步为在线心理健康社区文本挖掘提供了丰富的工具,词嵌入模型从静态向量(如 Word2Vec)到动态语境表示(如 BERT),显著提升了语义理解能力。同时,一些深度学习分类模型如 BiLSTM、Transformer 和 TextCNN 等展现出了独特的优势。基于此,本文利用文本挖掘方法来提取文本的深层语义特征,并结合深度学习分类器,实现对亲密度的高精度分类,进一步提升了在线心理健康社区评论亲密度测度的准确性。

3 方法

本部分首先介绍了可解释性的用户评论亲密度水平测度模型的基本框架及其三个关键组成部分:词嵌入层、可解释性模块、TextCNN 模块。其次介绍了本文扩充 LIWC 词典中亲密度词汇的过程。表 1 为模型训练过程中的主要参数介绍。

表 1 模型训练过程中的主要参数

主要参数	解释	值
lr	分类模型的学习率	0.0001
n	评论的最大长度	128
V	词向量的维度	768
num_filters	TextCNN 模型的卷积核数量	128
filter_sizes	TextCNN 模型的卷积核大小	2, 3, 4
l	分类标签数量	3
num_folds	K 折验证	10
epoch	训练次数	50
seed	随机种子	10

3.1 可解释性的用户评论亲密度水平测度模型

在用户评论分析中，亲密度是一个主观且复杂的维度，需要捕捉评论中的细粒度语义信息和语境。因此，本文基于文本挖掘方法，构建了一个具有可解释性的用户评论亲密度水平测度模型框架。如图 1 所示，该框架由词嵌入层、可解释性模块和 TextCNN 模块组成。首先，词嵌入层（如 BERT、Word2Vec、GloVe）旨在将文本转化为高维语义向量表示，能够捕捉上下文信息和词之间的语义关系^[40, 42, 49]。本文使用中文 BERT 模型作为词嵌入工具，借助其强大的预训练能力生成能够表征评论语义的高质量向量，为分类任务奠定语义基础。其次，可解释性模块（如注意力机制）在诸多任务中被用于识别输入特征与输出标签的相关性，从而为模型预测提供解释^[50]。本文在该模块中引入标签感知注意层，能够评估每个词对不同亲密度水平的贡献，增强模型结果的可解释性，使模型预测过程更加透明。最后，TextCNN 是一种经典的文本分类模型，通过卷积层提取文本的局部特征（如 n-grams 模式），具备参数少、训练效率高的优势^[15]。本文将 TextCNN 作为最终的分类模块，接收来自可解释性模块的特征输入，进一步提取关键词局部信息并生成最终的亲密度预测结果。

3.1.1 词嵌入层

给定一个由单词序列 $X = (x_1, x_2, \dots, x_n)$ 组成的用户评论，其中 n 表示评论长度。本文使用中文 BERT 模型对用户评论进行编码，BERT 的输出可以表示为一个向量序列 $H = (h_1, h_2, \dots, h_n)$ ， h_n 表示第 n 个单词。 H 的形状为 $R^{n \times V}$ ， $V = 768$ 表示中文基础版 BERT 的词向量维度。

3.1.2 可解释性模块

将词向量 H 输入到可解释性模块中，为增强模型对预测标签的可解释性，本文构建了基于标签感知注意机制的可解释性模块。该模块使每个标签能够关注输入序列的不同部分，从而帮助模型识别哪些部分对标签预测的贡献更大。可解释性模块分为标签感知注意层和输出层，标签感知注意层的公式如下所示：

$$A_1 = \text{softmax}(HU_1) \tag{1}$$

$$Z_1 = A_1^T H \tag{2}$$

其中, $U_1 \in R^{V \times l}$ 表示标签感知注意层的参数矩阵; l 表示分类标签数量。此外, $A_1 \in R^{n \times l}$ 表示注意力权重矩阵; $Z_1 \in R^{l \times V}$ 表示标签感知注意层的输出矩阵。

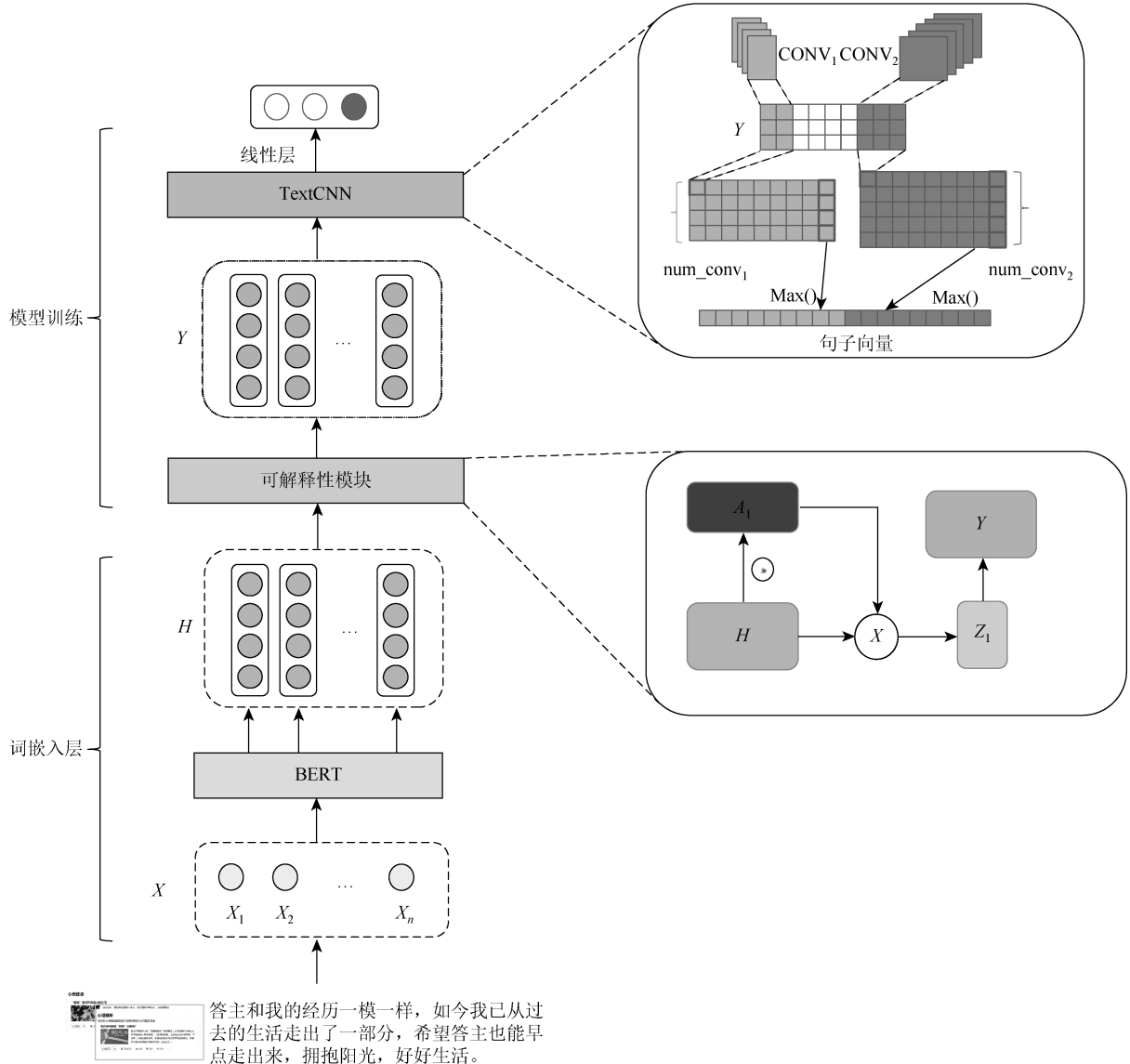


图 1 可解释性的用户评论亲密度水平测度模型框架

如上所述, 在注意力权重矩阵 $A_1 \in R^{n \times l}$ 中, 任意元素 $A_{1_{ij}}$ 都可以作为置信度来表示单词 i 对标签 j 预测的贡献。因此, 通过筛选出置信度最高的 k 个单词, 可以识别出用户评论中哪些单词与特定标签的预测关联性更强。

可解释性模块的输出层公式如下所示:

$$Y = U_2 Z_1 \quad (3)$$

其中, $U_2 \in R^{n \times l}$ 表示可解释性模块输出层的参数矩阵; $Y \in R^{n \times V}$ 表示可解释性模块的输出矩阵, 同时作为 TextCNN 模型的输入矩阵。

3.1.3 TextCNN 模块

可解释性模块得到评论新的词向量 $Y = (y_1, y_2, y_3, \dots, y_n)$ ，接下来将 Y 输入到卷积层。大小为 $V \times c$ 的卷积核 w 将词向量在词维上进行卷积：

$$c_i = f(w \cdot y_{i:i+c-1} + b) \quad (4)$$

其中， $y_{i:i+c-1}$ 表示第 i 个向量和第 $i+c-1$ 个向量之间的连接； b 表示偏置项； f 表示非线性激活函数。对于长度为 n 的评论，使用大小为 $V \times c$ 、数量为 m 的卷积核来处理窗口 $\{y_{1:h}, y_{2:h+1}, \dots, y_{n-h+1:n}\}$ ，以生成特征矩阵：

$$c = [c_1, c_2, \dots, c_{n-c+1}] \quad (5)$$

其中， c_i 表示长度为 m 的向量，将特征矩阵逐列最大化，得到 c_i 维的最大值，最终生成特征向量：

$$\hat{C} = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_{n-c+1}] \quad (6)$$

最后，通过线性层对评论亲密度水平进行分类，生成相应的类别概率：

$$\hat{y} = \text{Linear}(\hat{C}) \quad (7)$$

3.2 亲密度词汇扩充

随着在线心理健康社区的快速发展，许多新兴词汇频繁出现并被广泛使用，然而 LIWC 词典往往未能及时更新，导致对亲密度词汇的遗漏，影响了亲密度测量的准确性。为更好地将本文提出的模型与基于 LIWC 词典的亲密度计算方法进行比较，本文采用词向量训练和相似度计算的方法，对 LIWC 词典中与亲密度相关的词汇进行了扩展与完善。

具体流程如下：首先，对爬取的约 57 万条帖子和评论进行词向量训练，获取词向量表示。其次，根据亲密度水平的定义手动构建亲密度种子词，例如，低亲密度种子词包括“兴趣”“爱好”“观点”“态度”等；中亲密度种子词包括“工作”“学校”“金钱”“财产”等；高亲密度种子词包括“家庭”“身体”“亲密关系”“身体与心理疾病”等。通过计算种子词与其他词汇的相似度，选取相似度大于 0.5 的词汇扩充词典。最后，经过人工筛选剔除重复的亲密度词汇，生成最终的核心亲密度词典，其中包含 1920 个低亲密度词汇，1738 个中亲密度词汇，2232 个高亲密度词汇，共计 5890 个词汇。表 2 是不同亲密度水平词汇的部分展示。

表 2 不同亲密度水平词汇的部分展示

水平	部分种子词	部分亲密度词汇
低	兴趣、爱好、观点、态度	喜恶、交响乐、文学、宗教、书籍、艺人、野猫、杰伦、基督教、演员、电影、啤酒、翻唱
中	工作、学校、金钱、财产	上司、院长、实习、学者、学业、收入、财务、宿舍、成绩、高中化学、金钱、房屋产权、院校、大学、贫困
高	家庭、身体、亲密关系、身体与心理疾病	焦虑、抑郁、痴呆、机能、躯体、消瘦、亲人、爸妈、妻儿、前男友、兄弟、女朋友、住院、亲吻

4 实验

4.1 数据收集和标注

本文从 2015 年 2 月至 2023 年 5 月抓取了知乎“心理健康”话题下的 276 528 条帖子和 305 709 条评论。鉴于长度不足 4 个字符的评论通常信息量较低,多为简单应答或符号组合,难以提供有意义的分析支持,因此本文删除了长度不足 4 个字符的评论,以提升数据集质量并优化分析结果^[32]。最终,经过数据清洗后保留了 287 404 条评论。已有文献表明亲密度是语言的重要组成部分,能够通过自我披露的程度来表达^[32]。因此,本文基于用户在在线心理健康社区中的自我披露程度来评估其亲密度水平。亲密度分类标准及举例如表 3 所示。

表 3 亲密度分类标准及举例

水平	定义	举例
低	通常是比较肤浅的信息,如分享自己的兴趣爱好、表达对某事的观点、态度等。	1.现在呢?一切都怎么样了。 2.我觉得可能还没有正确的三观吧。
中	通常向他们感觉亲密的人分享信息,如分享自己的工作、学校、财务规划、消费偏好等信息。	1.我初中就想学心理学了,现在高一,但是身边很多人说心理学不好就业。 2.我只有跑步好一点了,把跑步撤了我岂不是不及格了。
高	通常只会向他们感觉非常亲密的人分享信息,如个人感受、家庭成员信息,疾病/手术痕迹、亲密关系隐私细节等。	1.太难过了,我的爸爸没有这样对我过。 2.答主和我的经历一模一样,如今我已从过去的生活中走出了一部分,希望答主也能早点走出来,拥抱阳光,好好生活。

从 287 404 条评论中随机抽取了 3000 条,由三名健康信息学领域的研究生基于亲密度分类标准对这些评论进行标注。在正式标注之前,三名标注者对随机抽取的 200 条评论的亲密度水平进行评级,并通过计算一致性检验系数(Fleiss Kappa, 弗莱希卡帕系数)评估标注者之间的一致性。结果显示, Fleiss Kappa 值等于 0.884,表明三位标注者之间的意见一致性较高,满足再现性和可靠性的要求。因此,允许标注者对本文所使用的 3000 条评论进行亲密度评级。对于标注过程中存在分歧的部分,通过多轮讨论达成一致,最终保留了 2962 条具有共识的标注数据。随后,将数据集按照 9:1 的比例划分为训练集和测试集,并采用 10 折交叉验证法对模型进行评估,以确保实验结果的稳定性和可靠性。

4.2 评估指标

本文采用宏观 $F1$ 、微观 $F1$ 和曲线下面积(area under the curve, AUC)作为评估模型分类性能的指标。宏观 $F1$ 和微观 $F1$ 的计算公式如下所示,其中 T_p^i 表示模型将第 i 类样本正确地分类为第 i 类的数量, F_p^i 表示模型将其他类别的样本错误地分类为第 i 类的数量, F_N^i 表示模型将第 i 类样本错误地分类为其他类别的数量。其中, i 表示类别的索引 0, 1, 2。

对于宏观 $F1$, 计算公式如下:

$$P^i = \frac{T_p^i}{T_p^i + F_p^i} \times 100\% \quad (8)$$

$$R^i = \frac{T_P^i}{T_P^i + F_N^i} \times 100\% \quad (9)$$

其中, P^i 和 R^i 表示第 i 类的精确度和召回率。则第 i 类的 $F1$ 值的计算公式为

$$F1^i = \frac{2 \times P^i \times R^i}{P^i + R^i} \quad (10)$$

那么可以计算出宏观 $F1$:

$$F1_{ma} = \frac{\sum_{i=1}^l F1^i}{l} \quad (11)$$

对于微观 $F1$, 计算公式如下:

$$F1_{mi} = \frac{2 \times \sum_{i=1}^l T_P^i}{\left(2 \times \sum_{i=1}^l T_P^i + \sum_{i=1}^l F_P^i + \sum_{i=1}^l F_N^i \right)} \times 100\% \quad (12)$$

4.3 模型对比

为了评估模型的有效性, 本文设计了三组对比实验: ①与主流文本分类方法的对比。BiLSTM、Transformer、BERT 等模型是文本分类领域中常用的基线模型, 具有较高的代表性。通过与这些方法的性能对比, 能够全面评估本文模型在不同基线下的表现, 证明其在亲密度分类任务中的优势。②进行消融实验。消融实验是模型性能验证的重要手段, 用于分析不同模块在整体模型中的贡献。本文通过逐步去除或保留可解释性模块和 TextCNN 模块, 能够明确各模块对最终分类效果的影响。③与扩充后的 LIWC 词典进行比较。LIWC 是一种常用的基于词典的情感与语言分析方法, 扩充后的词典能够较好地捕捉情感特征。将本文模型与词典方法进行对比, 可以探讨基于文本挖掘的语义建模方法与传统基于词典的方法在亲密度分类任务中的优劣。最后, 本文从测试集中随机选取高、中、低亲密度的评论, 进一步验证模型的可解释性。

5 结果

5.1 与其他分类模型对比

本文参考了最近关于在线心理健康社区文本分类和亲密度计算的相关文献^[32, 42, 44, 45, 47], 引入了一些基线模型进行对比分析, 包括 BiLSTM、Transformer、TextCNN、BERT 和 RoBERTa (robustly optimized BERT pretraining approach, 稳健优化型 BERT 预训练方法)。这些模型在文本分类和亲密度计算任务中表现出色, 尤其是 BERT 和 RoBERTa 等预训练模型。BERT 在 11 个经典自然语言处理 (natural language processing, NLP) 任务中超越了此前的最佳模型, 因此选择这些模型进行性能比较, 以验证本文方法的优越性。

具体来说, 本文将 2962 条标注数据分为 2666 条训练数据和 296 条测试数据, 对每个基线模型使用相同数据集, 并进行 10 折交叉验证。对于 BERT 和 RoBERTa 模型, 采用训练数据进行微调, 并分析它们在测试数据中的表现。最终结果如表 4 所示, BERT 和 RoBERTa 在 $F1_{ma}$ 性能上优于传统深度学

习模型,这主要因为其具有较强的文本语义特征提取能力。而本文提出的模型在 $F1_{ma}$ (92.45%)、 $F1_{mi}$ (93.7%) 和 AUC (98.82%) 指标上均超越了所有基线模型,表现最佳。

表 4 模型分类效果对比

模型	$F1_{ma}$	$F1_{mi}$	AUC
BiLSTM	86.45%	87.39%	96.53%
Transformer	88.92%	89.53%	98.37%
TextCNN	88.54%	89.21%	97.86%
BERT	90.43%	89.71%	98.00%
RoBERTa	90.74%	89.27%	97.52%
Interpretability-TextCNN	92.45%	93.70%	98.82%

5.2 消融实验

为了验证本文提出模块的有效性,本文进行了如下两组消融实验:①去除可解释性模块 (Interpretability),保留 TextCNN 模块。②保留可解释性模块,去除 TextCNN 模块。实验结果如表 5 所示,去除可解释性模块和 TextCNN 模块后,模型在大多数标签上的 $F1$ 值和 AUC 值均有显著下降,表明这两个模块在提升模型表现方面均发挥了重要作用。

表 5 消融实验结果对比

模型	$F1$			AUC		
	低	中	高	低	中	高
Interpretability-TextCNN	94.19%	95.72%	92.44%	98.96%	99.45%	98.29%
-Interpretability	93.57%	94.07%	91.49%	98.70%	99.35%	98.21%
-TextCNN	90.47%	92.89%	89.30%	93.42%	94.27%	91.77%

5.3 与扩充后的 LIWC 词典比较

将本文模型与基于 LIWC 词典计算亲密度的方法进行比较。在实验中,变量 LIWC_fre 表示不同亲密度等级词汇的词频。此前在 3.2 节中已扩充 LIWC 词典,包含 1920 个低亲密度、1738 个中亲密度和 2232 个高亲密度词汇。接着统计文本中高、中、低亲密度词汇的出现频率:若高亲密度词频 $\geq$ 中低亲密度词频,则标记为高亲密度;若中亲密度词频 $\geq$ 低亲密度词频,则标记为中亲密度;没有亲密度词汇则标记为低亲密度。结果如表 6 所示,基于 LIWC 的方法在 $F1_{ma}$ 、 $F1_{mi}$ 和 AUC 上均低于本文模型。

表 6 模型分类效果对比

模型	$F1_{ma}$	$F1_{mi}$	AUC
LIWC_fre	70.44%	80.26%	77.83%
Interpretability-TextCNN	92.45%	93.70%	98.82%

5.4 可解释性模块可视化展示

为了展示模型的可解释性，本文从测试集中随机选取了高、中、低三种亲密度的评论，并对可解释性模块中的注意力权重矩阵进行了可视化处理。如图 2 所示，模型能够有效识别“自卑”“自杀”“抑郁”等高亲密度词汇，并赋予较高的权重，显示出模型对情感表达的敏感性。图 3 显示模型能够识别“工作”“老师”等中等亲密度词汇，并赋予较高的权重，表明模型对适度亲密表达的有效捕捉。图 4 展示了模型对“文章”“警察”“破案”等低亲密度词汇的识别，且同样被赋予了较高权重，说明模型在理解语境时能够区分不同亲密度层次的词汇。通过对注意力权重矩阵的可视化分析，模型的标签与亲密度词汇之间展现出合理的关联性，进一步验证了模型在亲密度分类任务中的有效性和可解释性。

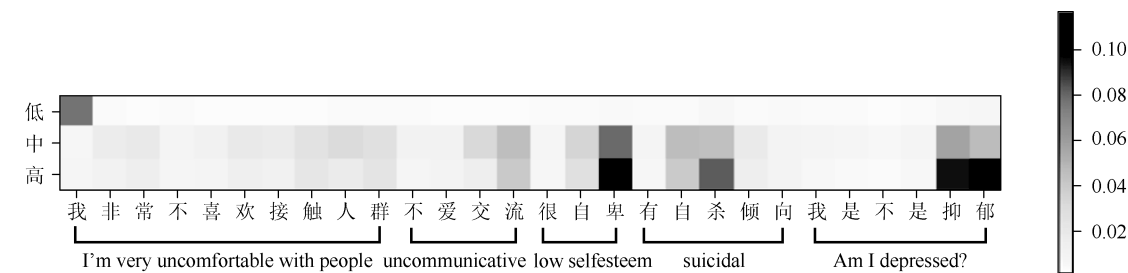


图 2 高亲密度评论的注意力权重矩阵可视化图

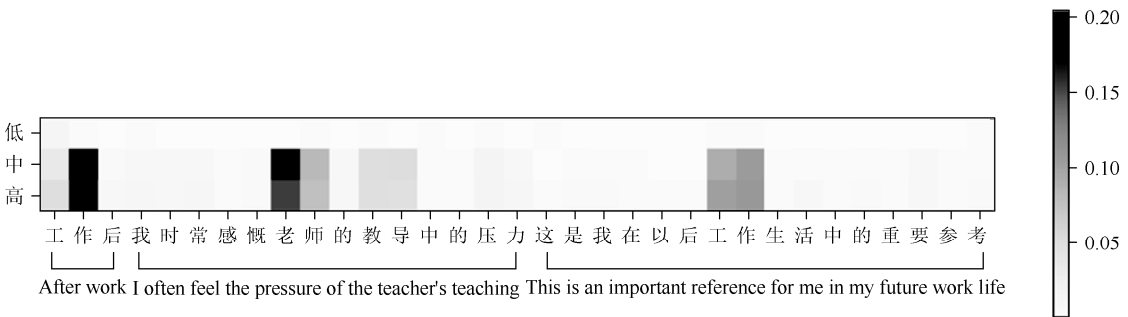


图 3 中等亲密度评论的注意力权重矩阵可视化图

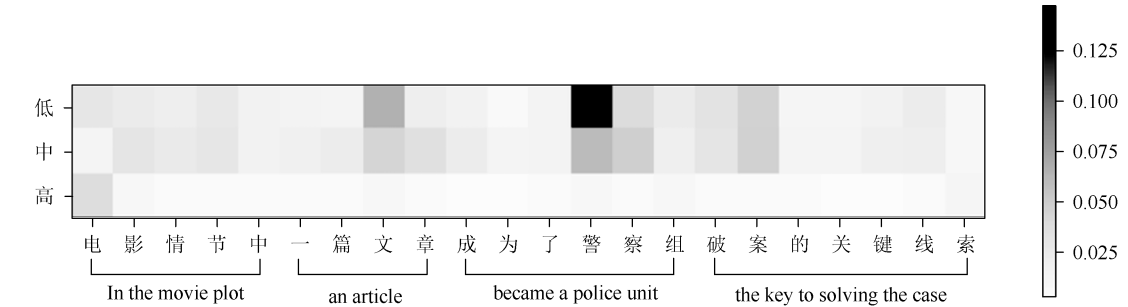


图 4 低亲密度评论的注意力权重矩阵可视化图

6 理论与实践意义

6.1 理论意义

以往对在线心理健康社区评论亲密度的测量方法主要使用 LIWC 词典, 该方法容易丢失语义信息, 并且存在词汇遗漏和准确性低的局限性。本文基于文本挖掘的方法, 构建了一个具有可解释性的用户评论亲密度水平测度模型。该模型不仅考虑了文本的语义信息, 而且具有可解释性。该模型利用中文 BERT 提取文本深层语义特征, 并引入可解释性模块和 TextCNN 模块, 赋予模型高度的可解释性和分类能力。为了与传统方法 (LIWC) 的结果进行比较, 本文使用词向量训练和相似度计算的方法对 LIWC 词典中的亲密度词汇进行扩充, 得到 5890 个亲密度词汇, 这些亲密度词汇可供其他研究者使用。实验结果表明, 本文模型优于 LSTM、Transformer、BERT 等几种深度学习基线方法的分类效果, 优于传统方法 (LIWC) 的分类效果。本文为未来在线心理健康社区文本亲密度预测的研究提供了新的视角。

6.2 实践意义

本文具有以下实践意义。首先, 它有助于提高社区成员的满意度和参与度。高亲密度的评论意味着社区成员更愿意分享个人的心理健康问题和情感经历。通过模型识别出高亲密度的评论, 成员能够感知到这一类评论下的氛围友善且无批判性, 从而更加放心地表达自身的情感与心理健康问题。这种开放的情感氛围有助于营造积极、互助的社区环境, 进一步激发成员参与社区活动的热情。其次, 本文有助于识别潜在的心理健康问题。通过对在线心理健康社区中高亲密度评论的分析, 可以发现一些暗示社区成员心理困扰或情绪问题的线索。这有助于及早发现潜在的心理健康风险, 为后续提供及时且适当的干预和支持奠定基础。最后, 本文有助于改善在线心理健康社区的服务质量。社区管理员可以利用亲密度分类结果, 开发定制化的资源和支持工具, 为社区成员提供个性化、精准化的服务, 从而更好地满足他们的需求。

7 结论与展望

随着互联网技术的不断发展, 越来越多的人倾向于在在线心理健康社区上发表评论为他人提供情感支持或信息支持。而评论的亲密度水平直接影响社区成员之间的人际关系, 对社区成员的心理健康及社区活力具有深远影响。以往研究主要使用 LIWC 词典测量文本的亲密度水平, 但该方法未充分考虑文本的语义信息, 并且存在亲密度词汇遗漏和准确率低的问题。为了弥补现有方法的不足, 本文基于文本挖掘的方法, 构建了一个具有可解释性的用户评论亲密度水平测度模型, 旨在推动在线心理健康社区的亲密度测量方法创新。首先, 利用中文 BERT 模型获取评论的词向量表示, 并将其输入到可解释性模块与 TextCNN 模块进行分类器训练。引入可解释性模块不仅提升了模型的分类效果, 还能够可视化展示对分类标签贡献度较高的亲密度词汇, 使模型更具可解释性。此外, 为更有效地将本文模型与基于 LIWC 词典的亲密度测量方法进行对比, 本文对 LIWC 词典中的亲密度词汇进行扩展, 最终获得 5890 个亲密度词汇。实验结果表明, 本文所提出的模型在分类效果上优于 BiLSTM、Transformer 和 BERT 等几个主流文本分类模型。相比传统的 LIWC 方法, 该模型在 $F1_{ma}$ 、 $F1_{mi}$ 和 AUC 指标上表现更加出

色,证明其在分类和识别文本亲密度水平方面更加准确,为在线心理健康社区提供了一种更高效的亲密度分析工具。

尽管该模型为亲密度预测做出了贡献,但仍存在一些局限性。例如,知乎“心理健康”模块仅为在线心理健康社区的一部分,无法代表所有的在线心理健康社区。对更多类型在线心理健康社区的研究将更具实用价值。未来,将通过实际应用实验进一步验证该模型的有效性和实用性,并分析研究过程中收集的帖子,以确定影响用户发表低、中、高亲密度评论的因素。这样不仅能够改进模型,还能更深入地理解用户行为,为在线心理健康服务提供更有效的支持。

参 考 文 献

- [1] 中国麻醉药品协会精神卫生分会. 2023 年度中国精神心理健康蓝皮书[R]. 2023.
- [2] Tan Y X, Teng Z W, Qiu Y, et al. Potential of mobile technology to relieve the urgent mental health needs in China: web-based survey[J]. JMIR MHealth and UHealth, 2020, 8 (7): e16215.
- [3] Diethei D, Colley A, Wienert J, et al. Different length, different needs: qualitative analysis of threads in online health communities[C]//2022 IEEE 10th International Conference on Healthcare Informatics (ICHI) . Rochester: IEEE, 2022: 348-356.
- [4] Kang Y B, McCosker A, Farmer J. Leveraging stylometry analysis to identify unique characteristics of peer support user groups in online mental health forums[J]. Scientific Reports, 2023, 13: 22979.
- [5] Peng Z, Ma X, Yang D, et al. Effects of support-seekers' community knowledge on their expressed satisfaction with the received comments in mental health communities[C]//Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2021: 1-12.
- [6] Yang X J, Gu D X, Wu J, et al. Factors influencing health anxiety: the stimulus-organism-response model perspective[J]. Internet Research, 2021, 31 (6): 2033-2054.
- [7] Chen Y, Xu Y. Exploring the effect of social support and empathy on user engagement in online mental health communities[J]. International Journal of Environmental Research and Public Health, 2021, 18 (13): 6855.
- [8] Rimé B, Bouchat P, Paquot L, et al. Intrapersonal, interpersonal, and social outcomes of the social sharing of emotion[J]. Current Opinion in Psychology, 2020, 31: 127-134.
- [9] Sprecher S, Treger S, Wondra J D. Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions[J]. Journal of Social and Personal Relationships, 2013, 30 (4): 497-514.
- [10] Ogbe E, Harmon S, Van den Bergh R, et al. A systematic review of intimate partner violence interventions focused on improving social support and/mental health outcomes of survivors[J]. PLOS One, 2020, 15 (6): e0235177.
- [11] Motta M, Liu Y N, Yarnell A. "Influencing the influencers: " a field experimental approach to promoting effective mental health communication on TikTok[J]. Scientific Reports, 2024, 14: 5864.
- [12] Gilbert E, Karahalios K. Predicting tie strength with social media[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2009: 211-220.
- [13] Reuel A K, Peralta S, Sedoc J, et al. Measuring the language of self-disclosure across corpora[C]//Findings of the Association for Computational Linguistics: ACL 2022. Dublin: Association for Computational Linguistics, 2022: 1035-1047.
- [14] Zeng X K, Yang C, Tu C C, et al. Chinese LIWC lexicon expansion via hierarchical classification of word embeddings with sememe attention[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Cambridge: AAAI Press, 2018, : 5650-5657.
- [15] Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence

- classification[J]. arXiv preprint arXiv: 1510.03820, 2015.
- [16] 杨善林, 丁帅, 顾东晓, 等. 医疗健康大数据驱动的知识发现与知识服务方法[J]. 管理世界, 2022, 38 (1): 219-228.
- [17] 杨雪洁, 顾东晓, 梁昌勇, 等. 在线健康社区中慢性病用户知识采纳行为研究[J]. 信息系统学报, 2020, (2): 67-76.
- [18] 杨善林, 丁帅, 顾东晓, 等. 医联网: 新时代医疗健康模式变革与创新发展[J]. 管理科学学报, 2021, 24 (10): 1-11.
- [19] Win Myint P Y, Lo S L, Zhang Y H. Unveiling the dynamics of crisis events: sentiment and emotion analysis via multi-task learning with attention mechanism and subject-based intent prediction[J]. Information Processing & Management, 2024, 61 (4): 103695.
- [20] 顾东晓, 赵旺, 王晓玉, 等. 大模型“涌现”视域下的智能健康医疗知识管理体系变革与创新发展[J]. 图书情报知识, 2025, 42 (1): 44-56.
- [21] 沈彦琪, 宋小康, 朱庆华. 在线健康信息替代搜寻对老年人健康素养的影响研究: 基于社会认知理论[J]. 现代情报, 2024, 44 (10): 115-125.
- [22] 邓朝华, 邓子豪, 樊国睿, 等. 在线健康平台能否减少健康不公平? : 基于价值共创视角[J]. 管理世界, 2024, 40 (1): 119-134.
- [23] Ke Q, Du J T, Geng Y X, et al. Studying health anxiety related attentional bias during online health information seeking: impacts of stages and task types[J]. Information Processing & Management, 2023, 60 (5): 103453.
- [24] 孙士伟, 刘思宇, 颜志军. 医生在线医疗绩效影响因素研究: 信任传递视角[J]. 信息系统学报, 2024, (1): 1-16.
- [25] 邓朝华, 邓子豪, 樊国睿. 在线医疗平台患者择医行为研究: 医生团队的作用[J]. 信息技术与管理应用, 2023, 2 (2): 94-106.
- [26] Fan H M, Lederman R. Online health communities: how do community members build the trust required to adopt information and form close relationships? [J]. European Journal of Information Systems, 2018, 27 (1): 62-89.
- [27] Ma X, Hancock J, Naaman M. Anonymity, intimacy and self-disclosure in social media[C]//Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2016: 3857-3869.
- [28] Ogba K T U, Onyishi I E, Chukwuorji J C. Self-disclosure in intimate relationships: moderating role of online moral foundations[J]. Current Psychology, 2021, 40 (10): 4785-4790.
- [29] Sheng J T, Kairam S R. From virtual strangers to IRL friends: relationship development in livestreaming communities on Twitch [J]. Proceedings of the ACM on Human-Computer Interaction, 2020, 4 (CSCW2): 1-34.
- [30] Seol K, Kim J D, Baik D K. Common neighbour similarity-based approach to support intimacy measurement in social networks[J]. Journal of Information Science, 2016, 42 (2): 128-137.
- [31] Spiliotopoulos T, Oakley I. Altruistic and selfish communication on social media: the moderating effects of Tie strength and interpersonal trust[J]. Behaviour & Information Technology, 2021, 40 (3): 320-336.
- [32] Pei J X, Jurgens D. Quantifying intimacy in language[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) . Stroudsburg: Association for Computational Linguistics, 2020: 5307-5326.
- [33] Gu D X, Liu H, Zhao H M, et al. A deep learning and clustering-based topic consistency modeling framework for matching health information supply and demand[J]. Journal of the Association for Information Science and Technology, 2024, 75 (2): 152-166.
- [34] Kilanioti I, Papadopoulos G A. A knowledge graph-based deep learning framework for efficient content similarity search of sustainable development goals data[J]. Data Intelligence, 2023, 5 (3): 663-684.
- [35] Gu D X, Wang Q, Chai Y D, et al. Identifying the risk factors of allergic rhinitis based on Zhihu comment data using a topic-enhanced word-embedding model: mixed method study and cluster analysis[J]. Journal of Medical Internet Research, 2024, 26: e48324.
- [36] Zhang Y Y, Chen Y, Yu S K, et al. Bi-GRU relation extraction model based on KeywordsAttention[J]. Data Intelligence, 2022, 4 (3): 552-572.
- [37] Ji D H, Gao J, Fei H, et al. A deep neural network model for speakers coreference resolution in legal texts[J]. Information Processing & Management, 2020, 57 (6): 102365.
- [38] Qiu Y, Ding S, Tian D, et al. Predicting the quality of answers with less bias in online health question answering

- communities[J]. *Information Processing & Management*, 2022, 59 (6): 103112.
- [39] 韩普, 叶东宇. 基于语义增强的在线健康社区情感分析研究[J]. *科技情报研究*, 2024, 6 (2): 88-99.
- [40] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//*Proceedings of the 26th International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc, 2013: 3111-3119.
- [41] Peters M, Neumann M, Iyyer M, et al. Deep contextualized word representations[C]//*Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Louisiana: Association for Computational Linguistics, 2018: 2227-2237.
- [42] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//*Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis: Association for Computational Linguistics, 2019: 4171-4186.
- [43] 熊回香, 向瀛泓, 孟璇, 等. 在线健康社区中基于用户评论挖掘的医生服务评价研究[J]. *情报科学*, 2024, 42 (8): 118-125, 163.
- [44] Shang F J, Ran C F. An entity recognition model based on deep learning fusion of text feature[J]. *Information Processing & Management*, 2022, 59 (2): 102841.
- [45] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc, 2017: 6000-6010.
- [46] Zhang D S, Zhou L N, Tao J, et al. KETCH: a knowledge-enhanced transformer-based approach to suicidal ideation detection from social media content[J]. *Information Systems Research*, 2024, 36 (1): 572-599.
- [47] Gu D X, Li M, Yang X J, et al. An analysis of cognitive change in online mental health communities: a textual data analysis based on post replies of support seekers[J]. *Information Processing & Management*, 2023, 60 (2): 103192.
- [48] Kumar A, Srinivasan K, Cheng W H, et al. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data[J]. *Information Processing & Management*, 2020, 57 (1): 102141.
- [49] Pennington J, Socher R, Manning C. GloVe: global vectors for word representation[C]//*Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha: Association for Computational Linguistics, 2014: 1532-1543.
- [50] Yang T T, Li F, Ji D H, et al. Fine-grained depression analysis based on Chinese micro-blog reviews[J]. *Information Processing & Management*, 2021, 58 (6): 102681.

Research on Measuring User Comment Intimacy Levels in Online Mental Health Communities: A Text Mining Approach

SUN Jiakun¹, GU Dongxiao^{1,2}, HUANG Zhengxing³, SU Kaixiang¹, ZHU Kaixuan¹, LI Min¹

(1. School of Management, Hefei University of Technology, Hefei, 230009, China)

(2. Laboratory of Data Science and Smart Society Governance of the Ministry of Education, Hefei, 230009, China)

(3. Institute of Artificial Intelligence, Zhejiang University, Hangzhou, 310008, China)

Abstract Intimacy is a key element of language and emotional communication between members of online mental health communities. Existing studies mainly use the LIWC dictionary to calculate intimacy, but this method is prone to losing semantic information and has the limitations of vocabulary omission and low

accuracy. This study constructed an interpretable user comment intimacy level measurement model based on text mining methods. The model used Chinese BERT to extract text semantic features and introduced an interpretability module and a TextCNN module to give the model a high degree of interpretability and classification capabilities. The effectiveness of the model was verified by experimental evaluation of comments under the topic of “Mental Health” on Zhihu.

Key words Online mental health community, User comment, Text mining, Intimacy, Interpretability

作者简介

孙佳坤（1998—），男，合肥工业大学管理学院硕士研究生，研究方向包括医疗健康大数据分析、自然语言处理等。E-mail: 2022110739@mail.hfut.edu.cn。

顾东晓（1980—），男，合肥工业大学管理学院教授，博士生导师，研究方向为医疗大数据分析、互联网医疗等。E-mail: dongxiaogu@yeah.net。

黄正行（1980—），男，浙江大学人工智能研究所教授，博士生导师，研究方向为医学人工智能。E-mail: zhengxinghuang@zju.edu.cn，

苏凯翔（1996—），男，合肥工业大学管理学院博士研究生，研究方向包括信息系统、知识管理等。E-mail: 2018110745@mail.hfut.edu.cn。

朱凯旋（1993—），男，合肥工业大学管理学院博士研究生，研究方向包括医疗健康、数据挖掘与智能决策等。E-mail: kaixuan_zhu@mail.hfut.edu.cn。

李敏（1998—），女，合肥工业大学管理学院硕士研究生，研究方向包括数据挖掘、自然语言处理等。E-mail: 2020110828@hfut.edu.cn。