基于 BERT 的在线心理健康问答社区回答质量预测研究

张 韦[1] 程炜焓[2] 郭富祥[1,3] 张建伟[4]

([1] 华中科技大学同济医学院医药卫生管理学院 武汉 430030
[2] 湖南中医药大学第一附属医院 长沙 410007 [3] 武汉市武昌医院 武汉 430064
[4] 日本岩手大学理工学部 盛冈市 020－8551)

〔摘要〕 **目的/意义** 构建在线心理健康问答社区回答质量预测模型，提高质量评估效率，实现回答质量实时监测和反馈，帮助心理咨询师持续改进服务质量。**方法/过程** 采用 2023 年 1—6 月“壹心理”问答专区 7 110 条问答数据，对 BERT 预训练模型进行微调，提取问答文本的深层语义特征，进而预测在线心理健康问答社区的回答质量。**结果/结论** 经过微调的 BERT 模型准确率和 *F*1 值均为 0.889，其预测性能优于 4 种经典模型，对在线心理健康问答社区的回答质量预测表现优异，具备代替人工标注、内容审核和服务质量实时监测的潜力。

〔关键词〕 深度学习；BERT；在线问答社区；心理健康；服务质量
〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2025.09.005

Study on the Answer Quality Prediction for Online Mental Health Q&A Communities Based on BERT

ZHANG Wei[1], *CHENG Weihan*[2], *GUO Fuxiang*[1,3], *ZHANG Jianwei*[4]

[1] *School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China*; [2] *The First Hospital of Hunan University of Chinese Medicine, Changsha 410007, China*; [3] *Wuhan Wuchang Hospital, Wuhan 430064, China*; [4] *Faculty of Science and Engineering, Iwate University, Morioka 020－8551, Japan*

〔**Abstract**〕 **Purpose/Significance** To construct a prediction model for answer quality in online mental health Q&A communities, so as to improve the efficiency of quality assessment, achieve real－time monitoring and feedback of answer quality, and assist mental health counselors in continuously improving service quality. **Method/Process** Based on the 7 110 pieces of Q&A data from the “Yixinli” Q&A platform from January to June 2023, the BERT pre－trained model is fine－tuned to extract deep semantic features of the Q&A texts, thereby predicting the answer quality in online mental health Q&A communities. **Result/Conclusion** The fine－tuned BERT model achieves an accuracy rate and *F*1－score of 0.889, outperforming four classical models in predictive performance. It demonstrates excellent predictive capability for answer quality in online mental health Q&A communities and shows potential to replace manual annotation, content review, and real－time service quality monitoring.

〔**Keywords**〕 deep learning; BERT; online Q&A community; mental health, service quality

〔修回日期〕 2025－07－30

〔作者简介〕 张韦，副研究员，博士生导师，发表论文 80 余篇；通信作者：程炜焓。

〔基金项目〕 国家自然科学基金项目（项目编号：72104087）。

1 引言

中国科学院心理研究所 2023 年 2 月发布的《中国国民心理健康发展报告（2021～2022）》[1]显示，我国抑郁风险人群检出率达 10.6%，焦虑风险检出率达 15.8%，且 18～24 岁青年群体的抑郁风险检出率达 24.1%，我国国民目前正面临着较大的心理健康风险。《“健康中国 2030”规划纲要》强调“加大全民心理健康科普宣传力度，提升心理健康素养”“加大对重点人群心理问题早期发现和及时干预力度”。随着数字化和网络化的发展，公众越来越倾向于通过线上渠道寻求心理健康信息，如何更好地营造心理健康信息环境日益受到关注[2]。心理疾病多需要长期服用药物、连续进行心理咨询和治疗，因而更适合互联网诊疗模式。通过互联网平台，可将有限的优质心理健康资源进行扩展，提高服务便利性，从而满足用户的咨询和治疗需求。相较于传统心理咨询服务，在线心理健康问答社区（如“壹心理”等）能够及时便捷地为用户提供所需心理健康信息和心理咨询、治疗服务。用户可以在其中相互提问、回答和分享知识[2]，不仅满足了用户的信息需求[3]，还为用户提供了社会支持和情感需求[4]，节省传统心理咨询服务的预约时间与费用。

然而，由于我国心理咨询专业人员供给不足、缺乏专业培训，在线心理健康问答社区可能存在信息质量良莠不齐的问题。不同于传统的医生－患者面对面咨询，在线健康社区难以保证所有信息的准确性、时效性，甚至真实性，因为这些信息可能来自未经医学教育的用户[5]。一项针对我国 40 个知名在线健康咨询平台的信息质量调查显示，其健康信息质量平均得分仅为 70 分，标准差为 12，表明在线健康咨询平台信息质量普遍不佳且不够稳定[6]。同时，在在线心理健康服务实践中，也暴露出隐私泄露、缺乏门槛、信息延迟等方面的问题[7-9]，极大影响了在线心理健康服务的效果。在线心理健康社区主要由同伴支持者驱动，因此必须确保共享内容的质量、可信度和支持性[10]，以便促进健康和颠覆性行为改变[11]。为避免骚扰、垃圾信息[12]、不良行为[13]等，国外 Reddit 社区中的在线心理健康社区制定了成员行为的规范和准则，并设立审核机制[14]。与常规在线社区相比，在线心理健康社区中会存在一系列关于自我伤害和自杀念头的帖子，因此其内容审核伴有更强烈的责任感和心理冲击，审核人员可能会因阅读过多负面内容而产生心理问题[10]。心理咨询师的工作压力日益增加，可能会影响其心理健康和幸福感[15]，产生同情疲劳[16]，导致咨询过程中问答服务质量下降。因此，对心理咨询师回答内容质量评估和预测的需求日益突显[17]，需要具备专业知识的心理咨询师通过大量工作人工对回答进行标注。这种回答质量的评估和预测具有滞后性，无法实现对回答质量的实时监测以及对心理健康服务需求方和提供方的实时反馈。

以往在线健康社区的回答主要利用文本特征（主题、情感等）和非文本特征（社会特征、写作风格、时间特征等）进行质量评估和预测[18-20]。随着深度学习方法的发展，研究者发现医学领域非文本特征似乎并不像文本特征那样突出，原因可能是医学领域知识的特殊性[21-23]。心理咨询同样具有很强的专业性[24]，同时在线心理健康问答社区中存在更多的非专业人士回答，质量预测更复杂，需要先进的深度学习模型来提取深层语义特征。

因此，构建在线心理健康问答社区的回答质量自动化预测模型不仅能够成为专业心理咨询师在回答质量标注中的助手，减轻内容审核人员压力，提高回答质量评估效率，而且还能够实现对心理健康问答服务质量的实时监测，为心理健康服务需求方提供快速质量评估参考。同时，对于心理健康服务提供方而言，快速的回答质量反馈，也能帮助心理咨询师持续改进其所提供的服务质量。

本研究基于双向编码器表征（bidirectional encoder representations from transformers，BERT）预训练模型，利用在线心理健康问答平台上的问答数据，针对回答质量预测任务，对模型参数进行微调，构建自动化质量预测模型，并与常用文本分类模型进行性能比较，以期得到效果最佳的模型，为在线心理健康服务平台提供能够快速确定回答质量

的预测工具，为高效筛选高质量回答提供支持。

2 对象与方法

2.1 对象

以“壹心理”网站的在线心理健康问答文本作为模型训练语料。“壹心理”是国内最大的在线心理健康服务平台，拥有广泛的市场影响力和用户基础。其问答板块专注于提供专业解答，帮助用户应对日常心理问题，并提供心理健康科普与咨询服务。“壹心理”问答社区的特色在于，平台会组织专业审核团队对平台上产生的回答进行实时审核，根据平台共情、科普、表达、分享、分析、梳理和解答等方面的标准决定是否给予“星标”标志（满足 2 条以上则算作高质量回答），“星标”决定回答者是否有资格参与付费提问的赏金分配。本研究使用该标准作为区分高质量回答与低质量回答的标准。

利用 Python 编写爬虫代码，抓取 2023 年 1 月 1 日—6 月 30 日“壹心理”在线心理健康服务平台问答专区的全部问答文本数据。共包括 5 812 条提问，29 323 条回答，其中 3 条回答由于违反平台守则或答主规范被删除，剩余 29 320 条回答。这些提问共获得 56 万余次浏览，得到 2 515 人次关注和 9 700 次“共情”。上述回答平均字数为 706 字，共获得 55 673 次点赞和 3 147 次评论，其中获得“星标”的回答为 25 765 条，未获得“星标”的回答为 3 555 条。将获得星标的回答编码为 1，非星标回答编码为 0。

在获取的数据中，高质量回答远多于低质量回答，前者是后者的 7 倍多。在模型训练过程中，如果分类模型获得的训练数据中各类别样本极不均衡，可能会带来如下问题：一是过拟合导致的泛化能力差；二是由于某部分样本占据主导而对少数类别的预测性能下降；三是评估指标无法完全衡量模型性能。因此，按照 5∶5 的两类别比例构建训练数据集，编码为 0 的 3 555 条数据全部保留，编码为 1 的 25 765 条数据中，为保证文本长度分布的一致，按文本长度进行排序后系统抽取 3 555 条，最终数据量为 7 110 条。回答数据的文本长度分布，见图 1，其均值和中位数分别为 690 字与 600 字，40% 的回答长度在 510 个词元内，80% 的回答长度在 988 个词元内。

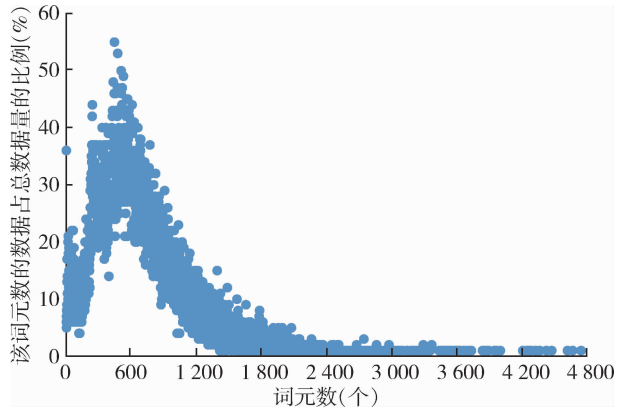

图 1 回答数据的文本长度分布

将 7 110 条回答创建为数据集 A，为测试提问文本对模型预测的影响，将 7 110 条回答及其对应的提问文本创建为数据集 QA。机器学习模型将数据集按 8∶2 的比例随机划分训练集和测试集，对训练集进行 5 折交叉验证；深度学习模型将数据集按 6∶2∶2 的比例划分训练集、验证集和测试集，其中测试集与机器学习模型测试集一致，训练集与验证集基于机器学习模型的训练集随机划分。验证集用于模型性能的实时监控和超参数调整，测试集用于评估模型的最终性能。

2.2 特征提取与模型选择

自然语言处理（natural language processing，NLP）所面临的文本数据往往是杂乱无章的非结构化数据，而机器学习模型或深度学习模型只能接受固定长度的数据输入，因此必须将文本转化为数字，即文本特征。2018 年问世的 BERT 模型[25]以 Transformer 模型[26]为基础提取文本特征，与 Transformer 模型相比，主要有两项改进：一是摒弃以往单向语言模型或浅层拼接单向语言模型的做法，使用掩码语言模型对双向 Transformer 模型进行预训练，以生成深层的双向语言表征；二是完成预训练后，只须在模型下层添加一个输出层，并结合小规模数据进行微调，就可以胜任各种不同的下游任

务，无须修改 BERT 本身的结构。BERT 模型可以将整段文本直接转化为特征，相比传统机器学习需要人工提取特征而言，信息丢失少，特征提取全面，具备对文本深层次语义的提取能力。因此，选择 BERT 模型提取回答的语义特征。

2.3 BERT 预测模型构建

2.3.1 参数设置与模型微调 在模型训练前须指定 BERT 模型的超参数。经测试，batch_size = 32，epoch_num = 2，learning_rate = 2e - 5 是计算资源消耗较小、性能较高的超参数组合。参照 BERT 模型的原始论文[25]，其他参数与预训练期间保持一致。确定超参数后，即可将数据输入模型进行微调。为计算模型的损失并进行参数优化，定义损失函数和优化器。本研究采用 BERT - Base - Chinese 模型内部逻辑机制计算损失，选用 Adam 优化器，对 BERT 中的参数进行优化。

2.3.2 模型构建 将在线心理健康问答的回答质量评价问题转换成文本二分类问题。总结和对比传统浅层机器学习分类方法和深度学习分类方法，选择支持向量机（support vector machine，SVM）和全连接神经网络作为分类器，以 BERT 模型对文本特征进行表征，利用数据集 A 进行有监督分类任务训练。选取 BERT 模型 + 线性全连接层评估数据集训练效果。模型结构，见图 2。

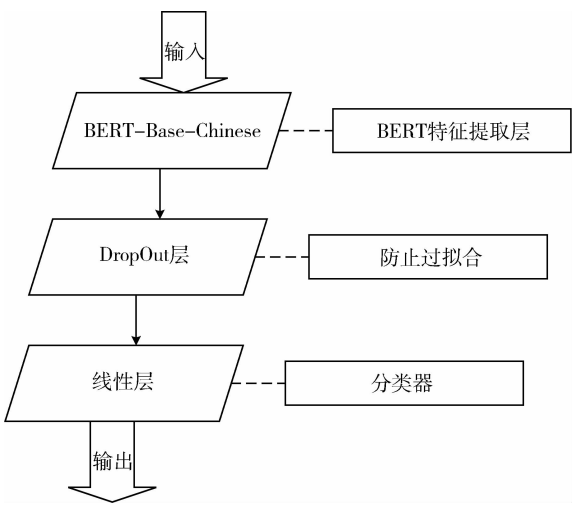

图 2 模型结构

BERT 模型最大输入为 512 个词元，其中有两个固定掩码，占 2 个词元，因此对超过 510 个词元的数据进行预处理。Sun C 等[27]研究认为，文本分类任务常用的长文本处理方法包括截断法和池化法，其中截断法的微调效果最好。本研究场景中超长文本较少，可对超过 510 个词元的数据进行截断处理。截断法包括前截断（head）、后截断（tail）和前后截断（H&T）3 种形式，前（后）截断是指截取输入数据的前（后）510 个词元进行输入，前后截断是指从输入数据的开头和结尾各截取一部分输入模型。按照设计思路，使用 3 种截断方法完成对 BERT - Base - Chinese 模型的微调，超参数则与 BERT 模型[25]进行对比。依据实验结果，选择微调效果最佳的截断方法作为后续实验的长文本处理方法。为确定提问文本对模型微调的影响，将数据集 A 和 QA 分别输入模型进行微调。依据实验结果，选择是否纳入提问文本来微调模型。

2.4 融合特征的“BERT +”模型

使用经过微调的 BERT - Head 模型提取文本特征，添加文本长度和词频 - 逆文本频率（term frequency - inverse document frequency，TF - IDF）主题特征，将 3 种特征输入极端梯度提升（extreme gradient boosting，XGBoost）或双向长短期记忆（bidirectional long short - term memory，BiLSTM）模型，训练“BERT +”分类模型。通过网格搜索，获取各模型最佳超参数组合：BERT + XGBoost 最佳超参数组合为 gamma = 1，learning_rate = 0.05，max_depth = 3，n_estimators = 300；BERT + BiLSTM 最佳超参数组合为 epoch = 4，batch_size = 32，learning_rate = 0.001。

2.5 常用文本分类模型性能对比

为探究 BERT 模型对于在线心理健康问答的回答质量预测是否具备性能优势，选择 SVM、XGBoost[28]、卷积神经网络（convolutional neural network，CNN）[29]和长短期记忆网络（long short - term memory，LSTM）4 种常用于文本分类的机器学习或深度学习模型，进行对比。通过网格搜索，获取各预测模型最佳超参数组合：SVM 最佳超参数组

合为 C = 1，gamma = 10，kernel = rbf；XGBoost 最佳超参数组合为 gamma = 0.05，learning_rate = 0.05，max_depth = 7，n_estimators = 100；TextCNN 最佳超参数组合为 epoch = 3，batch_size = 32，learning_rate = 0.01；BiLSTM 最佳超参数组合为 epoch = 2，batch_size = 32，learning_rate = 0.01。

2.6 模型评价指标

使用准确率、*F*1 值和 Macro – *F*1 值等指标评估分类器的效果和泛化能力，以全面反映模型在不同方面的表现，确保评价结果的准确性和实用性。这些指标均适用于不平衡样本，其中准确率易于理解和计算，*F*1 值综合考虑了精确率和召回率，而 Macro – *F*1 能有效评估模型预测阴性和阳性结果的性能。

3 结果

前截断、后截断和前后截断方法的 BERT 训练模型结果，见表 1，3 种截断方法在性能上无明显差异，前截断方法得到的 BERT – Head 模型性能在各分类上更为均衡，因此采取前截断方法。

表 1 3 种截断方法性能比较

类别	准确率	*F*1	Macro – *F*1
BERT – Tail	0.884	0.886	0.875
BERT – H&T	0.889	0.884	0.889
BERT – Head	0.889	0.889	0.889

将数据集 A 和 QA 分别输入模型进行微调，所得到的模型性能，见表 2，加入提问文本与否对于回答质量的预测无显著影响。因此选择仅包含回答的数据集 A 作为研究数据。

表 2 数据集 A 和 QA 微调模型性能比较

类别	数据集	准确率	*F*1	Macro – *F*1
BERT_QA	QA	0.89	0.89	0.89
BERT_A	A	0.89	0.89	0.89

汇总采用前截断方法的 BERT 训练模型、融合两种特征的“BERT +”模型与 4 个常用文本分类模型共 7 组实验所得的性能指标，综合 3 种性能指标进行排序，以便对比和分析模型性能差异的原因，见表 3。

表 3 预测模型性能比较

类别	准确率	*F*1	Macro – *F*1
SVM	0.809	0.816	0.808
TextCNN	0.828	0.810	0.826
XGBoost	0.850	0.854	0.850
BERT + BiLSTM	0.861	0.863	0.861
BiLSTM	0.866	0.876	0.866
BERT + XGBoost	0.886	0.888	0.886
BERT – Head	0.889	0.889	0.889

总体来看，深度学习模型在在线心理健康回答质量预测任务中表现优于传统机器学习模型。BERT – Head 模型准确率为 0.889，*F*1 分数为 0.889，Macro – *F*1 为 0.889，优于 SVM、XGBoost、TextCNN 和 BiLSTM 这 4 种经典分类模型。融合特征的 BERT + BiLSTM、BERT + XGBoost 模型并未得到高于 BERT 模型的性能。

4 讨论

在学术社区[30]、知识社区[31]、计算机专业社区[32]等问答社区的回答质量评估领域，主要基于提问、回答和用户 3 方面特征，使用机器学习方法预测回答质量。通过加入部分医疗健康领域特征或邀请医疗健康领域专家，有研究[19-20]建立了一些在线健康社区回答质量预测模型。深度学习方法的使用提升了预测模型的性能，研究场景也逐渐扩展到具有一定专业门槛的在线健康社区[19,21]。心理类咨询文本评估与预测往往需要专家级的注释者来标记数据[33]，且需要大量数据进行模型训练和调整。研究发现，经过训练的 BERT 模型在心理咨询类文本中的编码注释甚至比人类新手表现得更好[34]，在理解和推理文本含义方面具备强大能力[35]。本研究基于 BERT 预训练模型，从语义层面提取回答的特征，使用国内最大在线心理健康服务平台的问答数据对模型进行微调，实现了在线心理健康问答的回答质量自动化评价。

实验表明，BERT 模型能够实现对回答文本深

层次语义信息的表征，BERT－Head 模型准确率和 *F*1 值均为 0.889，相比基于传统文本特征提取算法的机器学习模型，能够得到更好的分类效果。

在 3 种截断方法中，前截断性能最佳。这表明，在线心理健康问答的回答中，关键的语义信息往往集中在文本开头。加入提问文本后对于回答质量的预测并无显著影响，这与其他问答社区的质量预测不同，可能是在线心理健康问答更注重回答本身，回答的深层语义是其质量预测最主要的特征，显示出心理健康领域的特殊性。融合特征的 BERT＋BiLSTM、BERT＋XGBoost 模型并未得到高于 BERT 模型的性能，可能是因为心理健康问答的回答质量评估重视深层语义解析和双方信息交互，而表层特征的融合并未带来预期的性能提升。这与 Qiu Y 等[21]直接用 BERT 模型预测在线健康问答社区回答质量得到的结果一致，语义特征对于识别回答质量更为重要[36]。

综上，基于预训练模型的微调模型已经具备取代人工标注的能力。后续问答社区可考虑从小范围开始试点人工智能模型的自动化质量标注，并将其用途从质量标注拓展到提问分类、抄袭审查、回答规范性审查等人力资源消耗较大的工作。

由于研究对象和研究方法的局限，本研究仍存在不足，有待改进。首先，由于爬取“壹心理”问答社区的高质量回答与低质量回答比例差距悬殊，本研究最终纳入 7 110 条问答数据，未来可纳入平台更长时间阶段的问答数据并预测其质量，以获得 BERT 模型更佳的预测效果。其次，本研究选取的 BERT 模型对输入存在 512 个词元的限制，无法处理超过 512 个词元的回答，在后续研究中，可考虑使用输入量更大的预训练模型进行训练。最后，本研究在融合特征的“BERT＋”模型中，仅纳入文本特征、文本长度、TF－IDF 主题特征，未考虑其他特征，如写作风格[19]、时间特征[20]、客观性、可读性[21]、用户特征等，后续可增加更多融合特征，以探索提升融合特征的“BERT＋”模型分类效果的途径。

作者贡献：张韦负责研究设计、论文撰写与修订；程炜焓负责文献调研、研究设计与实施、论文撰写与修订；郭富祥负责文献调研、论文撰写与修订；张建伟负责研究设计、论文修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 傅小兰，张侃，陈雪峰，等．中国国民心理健康发展报告（2021～2022）［M］．北京：社会科学文献出版社群学出版分社，2023.

2 姚宛京．基于 LDA 主题模型的用户心理健康信息需求研究——以社会化问答社区“知乎”为例［J］．现代信息科技，2024，8（1）：175－179，184.

3 刘烁，陈盼，杨冰香，等．基于知乎抑郁症问答社区的用户健康信息需求分析［J］．护理研究，2021，35（13）：2273－2279.

4 SHANG J，WEI S，JIN J，et al. Mental health Apps in China：analysis and quality assessment［J］. JMIR mhealth and uhealth，2019，7（11）：e13236.

5 LEDERMAN R，FAN H，SMITH S，et al. Who can you trust? Credibility assessment in online health forums［J］. Health policy and technology，2014，3（1）：13－25.

6 钱明辉，徐志轩，连漪．在线健康咨询平台信息质量评价及其品牌化启示［J］．情报资料工作，2018（3）：57－63.

7 杨晶，余林．网络心理咨询的实践及其存在的问题［J］．心理科学进展，2007（1）：140－145.

8 何珍妮，陈振华．网络心理咨询的优势、局限以及咨访双方对网络心理咨询的态度［J］．心理学进展，2021，11（9）：2172－2181.

9 靳宇倡，张政，郑佩璇，等．远程心理健康服务：应用、优势及挑战［J］．心理科学进展，2022，30（1）：141－156.

10 SAHA K，ERNALA S K，DUTTA S，et al. Understanding moderation in online mental health communities［C］. Copenhagen：22nd HCI International Conference，2020.

11 CHANCELLOR S，HU A，DE CHOUDHURY M. Norms matter：contrasting social support around behavior change in online weight loss communities［C］. Montreal：2018 CHI Conference on Human Factors in Computing Systems，2018.

12 JHAVER S，CHAN L，BRUCKMAN A. The view from the other side：the border between controversial speech and harassment on Kotaku in action［EB/OL］.［2025－04－19］. https：//arxiv. org/abs/1712. 05851.

13 CHENG J，BERNSTEIN M，DANESCU－NICULESCU－MIZIL C，et al. Anyone can become a troll：causes of troll-

ing behavior in online discussions [C]. Portland: 2017 ACM conference on computer supported cooperative work and social computing, 2017.

14 JUNEJA P, RAMA SUBRAMANIAN D, MITRA T. Through the looking glass: study of transparency in Reddit's moderation practices [J]. Proceedings of the ACM on human - computer interaction, 2020, 4 (1): 1 - 35.

15 ZHANG L, REN Z, JIANG G, et al. Self - oriented empathy and compassion fatigue: the serial mediation of dispositional mindfulness and counselor's self - efficacy [J]. Frontiers in psychology, 2021 (1): 613908.

16 ZHANG L, ZHANG T, REN Z, et al. Predicting compassion fatigue among psychological hotline counselors using machine learning techniques [J]. Current psychology, 2023, 42 (5): 4169 - 4180.

17 HUANG Y, LIU H, LI S, et al. Effective prediction and important counseling experience for perceived helpfulness of social question and answering - based online counseling: an explainable machine learning model [J]. Frontiers in public health, 2022 (12): 817570.

18 OH S, YI Y J, WORRALL A. Quality of health answers in social Q&A [J]. Proceedings of the American society for information science and technology, 2012, 49 (1): 1 - 6.

19 HU Z, ZHANG Z, YANG H, et al. A deep learning approach for predicting the quality of online health expert question - answering services [J]. Journal of biomedical informatics, 2017, 71 (7): 241 - 253.

20 ZHANG Z, HU Z, YANG H, et al. Factorization machines and deep views - based co - training for improving answer quality prediction in online health expert question - answering services [J]. Journal of biomedical informatics, 2018, 87 (11): 21 - 36.

21 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.

22 AMANCIO L, DORNELES C F, DALIP D H. Recency and quality - based ranking question in CQAs: a stack overflow case study [J]. Information processing & management, 2021, 58 (4): 102552.

23 DONG S, MAO J, KE Q, et al. Decoding the writing styles of disciplines: a large - scale quantitative analysis [J]. Information processing & management, 2024, 61 (4): 103718.

24 吴慕窈, 马向真. 咨询师视角下网络心理咨询伦理问题的定性研究 [J]. 中国心理卫生杂志, 2023, 37 (8): 694 - 700.

25 DEVLIN J, CHANG M W, LEE K, et al. BERT: pre - training of deep bidirectional transformers for language understanding [C]. Minneapolis: Association for Computational Linguistics, 2019.

26 VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2025 - 04 - 19]. https://dl.acm.org/doi/pdf/10.5555/3295222.3295349.

27 SUN C, QIU X, XU Y, et al. How to fine - tune bert for text classification [C]. Kunming: Chinese Computational Linguistics, 2019.

28 CHEN T, GUESTRIN C. Xgboost: a scalable tree boosting system [C]. San Francisco: 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016.

29 CHEN Y. Convolutional neural network for sentence classification [D]. Waterloo: University of Waterloo, 2015.

30 LI L, HE D, JENG W, et al. Answer quality characteristics and prediction on an academic Q&A site: a case study on researchgate [C]. New York: 24th International Conference on World Wide Web, 2015.

31 FU H, WU S, OH S. Evaluating answer quality across knowledge domains: using textual and non - textual features in social Q&A [J]. Proceedings of the association for information science and technology, 2015, 52 (1): 1 - 5.

32 NESHATI M. On early detection of high voted Q&A on stack overflow [J]. Information processing & management, 2017, 53 (4): 780 - 798.

33 LI A, MA J, MA L, et al. Towards automated real - time evaluation in text - based counseling [EB/OL]. [2024 - 06 - 21]. https://arxiv.org/abs/2203.03442.

34 GRANDEIT P, HABERKERN C, LANG M, et al. Using BERT for qualitative content analysis in psychosocial online counseling [C]. Online: 4th Workshop on Natural Language Processing and Computational Social Science, 2020.

35 赵铁军, 许木璠, 陈安东. 自然语言处理研究综述 [J]. 新疆师范大学学报 (哲学社会科学版), 2025, 46 (2): 89 - 111, 2.

36 HE X, WANG L, ZHANG W, et al. Research on the quality prediction of online Chinese question answering community answers based on comments [C]. New York: 2nd International Conference on Big Data Technologies, 2019.