

融合角色心理画像的心理健康文本匹配模型

赵芸^{1,2,3} 刘德喜^{1,3} 万常选^{1,3} 刘喜平^{1,3} 廖国琼^{1,3}

¹(江西财经大学信息管理学院 南昌 330013)

²(江西科技师范大学数学与计算机科学学院 南昌 330038)

³(数据与知识工程江西省高校重点实验室(江西财经大学) 南昌 330013)

(zhaoyun_hust@163.com)

Mental Health Text Matching Model Integrating Characters' Mental Portrait

Zhao Yun^{1,2,3}, Liu Dexi^{1,3}, Wan Changxuan^{1,3}, Liu Xiping^{1,3}, and Liao Guoqiong^{1,3}

¹(School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330013)

²(School of Mathematics and Computer Science, Jiangxi Science and Technology Normal University, Nanchang 330038)

³(Jiangxi Key Laboratory of Data and Knowledge Engineering (Jiangxi University of Finance and Economics), Nanchang 330013)

Abstract The global situation regarding mental health problems is a matter of serious concern. Unfortunately, in many countries, the shortage of mental health practitioners often leaves individuals with mental health problems without access to professional services. The retrieval-based community question answering for mental health can provide self-service information for those who are suffering from mental health problems. There are two different matching criteria to consider when matching question post and support response in this task, namely semantic matching and mental health matching. To address the issue of excessive non-ideal support responses in a large pool of candidate support responses, we propose a two-stage text matching model that integrates characters' mental portraits using a multi-layer attention mechanism. This method divides our task into two sequential subtasks: a retrieval task and a matching task, respectively. In the first stage, a retrieval sub-method is used to retrieve a set of candidate support responses with a certain semantic similarity for each question post. In the second stage, the matching sub-method performs matching judgments based on the results generated in the previous stage. This sub-method applies a multi-layer attention mechanism to effectively integrate characters' mental portraits and semantics. This process can modify the results of simple semantic matching. Experimental results show that T2CMP proposed in this paper is superior to state-of-the-art methods.

Key words text matching; 2-stage model; mental health portrait of character; multi-layer attention mechanism; self-service information of mental health

摘要 全球心理健康问题形势严峻,由于心理健康服务的从业人员不足,遭受心理健康困扰的人并不总是能获得专业的心理健康服务。检索式心理健康社区自动问答可以快速地为需要心理健康服务的人提供

收稿日期: 2022-12-07; 修回日期: 2023-10-30

基金项目: 国家自然科学基金项目(62272206, 62272205, 62076112, 62262027); 江西省自然科学基金项目(20212ACB202002); 江西省主要学科科学技术带头人培养计划领军人才项目(20213BCJL22041); 江西省教育厅科学技术研究项目(GJJ2200501, GJJ2201335); 江西科技师范大学博士科研启动基金项目(2022BSQD36)

This work was supported by the National Natural Science Foundation of China (62272206, 62272205, 62076112, 62262027), the Natural Science Foundation of Jiangxi Province (20212ACB202002), the Training Plan for Academic and Technical Leaders of Major Disciplines in Jiangxi Province (20213BCJL22041), the Science and Technology Project of Education Department of Jiangxi Province (GJJ2200501, GJJ2201335), and the PhD Foundation of Jiangxi Science and Technology Normal University (2022BSQD36).

通信作者: 刘德喜(dexi.liu@163.com)

相应的信息自助服务.与传统检索式社区问答中的文本匹配不同,在匹配支持帖和求助帖时,需要考虑2种不同层面的匹配准则:语义层面和心理层面.为了解决该问题,提出融合角色心理画像的2阶段文本匹配模型(two-stage text matching model integrating characters' mental portrait, T2CMP),该模型引入心理特征用于构建角色心理画像,从而辅助模型理解文本心理层面的内容和匹配关系.同时为了提升检索效率以及减少大量负样例带来的噪声问题,将文本匹配任务拆分为2阶段的序列型子任务.首先针对每条求助帖,使用基于语义的筛选模型甄别出候选支持帖;然后依据用户的角色心理画像,使用多层注意力机制将其与语义信息有效融合,提高模型的总体效果.在MHCQA数据集上的实验结果显示,T2CMP比现有优秀算法拥有更高的F1值.

关键词 文本匹配;2阶段模型;角色心理健康画像;多层注意力机制;心理健康信息自助服务

中图法分类号 TP391

在全球心理健康问题日益严峻的背景下,使用自动问答技术为网络社区用户提供面向心理支持的自助信息服务,将该任务称之为面向心理健康的社区问答(community question answering for mental health, MHCQA).相较于其他社区自动问答,关于MHCQA的相关工作比较少见.

虽然有学者尝试使用生成式对话提供心理支持,但要么只关注单一的心理支持信息(例如共情)^[1],心理支持效果有限;要么生成的心理支持信息与求助帖相关性不高,很难起到有效的心理支持作用^[2-3].相较于生成式模型,检索式MHCQA直接从心理健康社区中检索出与心理危机问题相匹配的帖子,这些帖子由人工撰写,能为求助者提供内容和类型更丰富的心理支持信息.

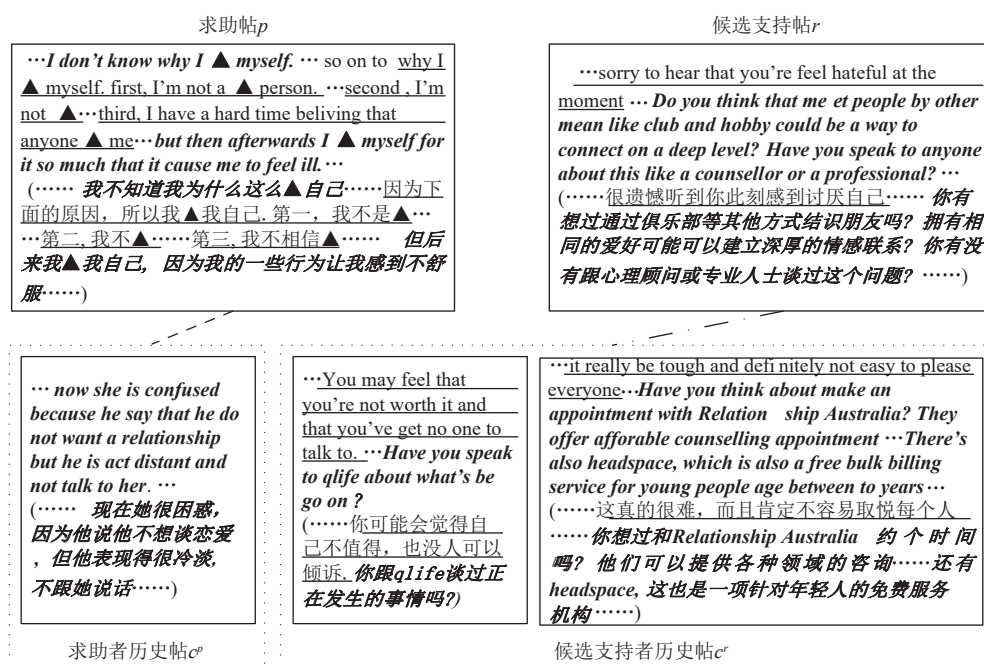
检索式MHCQA的核心是文本匹配,该任务中的求助帖和支持帖的匹配关系复杂,不能仅考虑语义层面的匹配,还需考虑心理层面的匹配.如图1所示,求助帖 p 中的当前心理危机问题(p 中加粗倾斜的内容)和候选支持帖 r 中的专业建议或方法(r 中加粗倾斜的内容)更倾向于心理层面的匹配,另外求助帖中的“原因、背景、情绪”和支持帖中的“共情”具有较强的情绪相关性.现有的检索式问答模型^[4-6]在进行匹配判断时主要基于语义层面,既未融合帖子心理层面的特征,也未挖掘用户的角色心理画像.虽然Zhao等人^[7]尝试引入心理健康相关特征,辅助模型学习不同的匹配关系,但其仅限于当前的求助帖和候选支持帖,没有考虑求助者/支持者的角色心理画像.

当前的心理危机消息往往只包含求助者某一时刻的心理危机表现或情绪感知,仅凭当前心理危机消息无法准确刻画求助者的心理危机状况(包括发生的原因或背景、症状、情绪等),影响心理危机消息和心理支持消息的匹配准确率.如图1所示,在求助帖 p 中,显示的心理危机问题是求助者对自身产

生了一些负面情绪,并列举了一些琐碎的示例来解释其负面情绪的来源.然而若结合他/她的历史帖 c' ,则可能刻画更为完整的心理画像,从图1看出,该求助者发生心理危机的导火索之一是感情受挫.

在心理健康社区中,针对相同类型的心理危机问题,支持者的支持策略是相似或相近的.但是不同的支持者有不同的支持偏好及支持风格,而且支持者的支持偏好和支持风格无法通过单个支持帖充分体现出来,需要从其所发布的多条支持帖中进行挖掘.如图1所示,候选支持者的历史帖 c' 展示了该支持者偏好对自我否定的心理危机问题提供支持(c' 中下划线的内容),且针对该类心理危机问题,其支持策略也类似(c' 中加粗倾斜的内容).

由此可见,捕获求助者和支持者的心理画像并将其融入语义匹配,能够提高心理危机问题与支持消息之间的匹配精度.基于此,我们提出融合角色心理画像的2阶段文本匹配模型(two-stage text matching model integrating characters' mental portrait, T2CMP).该模型通过求助者发布的历史帖全面刻画求助角色的心理画像,即有效反映求助者的心理危机问题、心理危机的背景、原因、表现以及它们之间的关系.通过支持者发布的历史帖刻画支持角色心理画像,包括支持者的支持专长、支持偏好以及支持风格等.接着使用求助者/支持者的角色心理画像辅助语义匹配,提高检索性能.此外,由于心理健康社区中候选支持帖过多,影响匹配效率,T2CMP将匹配过程划分为2个阶段:第1阶段侧重于语义层面的匹配,即使用基于语义的筛选模型,对大量候选支持帖进行筛选,该阶段可借鉴信息检索的成果,选择较好的检索模型;第2阶段侧重于心理层面的匹配,即通过融合角色心理画像,对筛选后的候选支持帖增加心理层面的匹配判断,修正单纯语义匹配的结果,从而提升模型整体效果.



▲: 用于替代示例中的极端负面词

加粗倾斜的内容: 求助者的心理危机问题或支持者的专业建议或方法

下划线标注的内容: 心理危机问题发生的原因或背景、症状、情绪等, 以及支持者的支持偏好等

Fig. 1 Examples of the question post and the candidate support response in the mental health community

图1 心理健康社区上的心理危机消息和候选心理支持回复示例

本文的创新性包括:

1) 提出并建模角色心理画像. 从求助者/支持者的历史帖抽取适合分析心理的特征, 建模心理特征及其关系, 对求助者/支持者进行角色心理画像.

2) 将角色心理画像与语义匹配融合, 提升心理支持帖检索的质量. 采用多层注意力机制融合帖子的语义与角色的心理画像, 弥补单纯语义匹配的不足.

3) 采用2阶段方法, 提升检索效率. 第1阶段通过语义匹配筛选候选心理支持帖; 第2阶段融合角色心理画像优化检索结果.

4) 基于 CLPsych2017 Workshop (the Workshop on Computational Linguistics and Clinical Psychology) 上的评测任务 (shared task) 构建 MHCQA 数据集, 并在该数据集上评估了所提 T2CMP 的有效性. 实验结果表明, 融合角色心理画像的文本匹配模型能有效修正单纯语义匹配的不足, 提高匹配效果.

1 相关工作

1.1 检索式自动问答

检索式自动问答的核心是问题与候选答案的匹配. 近年来, 神经网络模型被大量应用于文本对

匹配任务, 这些模型可以分为表示模型和交互模型两大类. 表示模型是将文本送入深度神经网络, 由其自主学习文本的表示向量, 然后将表示向量送入全连接网络获得2个文本的匹配得分^[8-11], 例如 ARC-I^[8] 等. 交互模型则是提取文本对中单词间的交互特征, 然后通过深度神经网络聚合交互特征, 得到最终的匹配结果^[4,8,12], 如 ARC-II^[8], MatchPyramid^[4] 等. 然而这些深度神经网络模型并没有完全捕捉到长文本的结构, 因此在长文本对的匹配任务上表现欠佳.

一些研究者通过引入外部知识^[13]、增加深度神经网络的层次^[14]、使用图结构^[5,15-17]等方式改进长文本的匹配效果. 例如 CIG-GCN^[5] 首先将文本对用概念交互图 (concept interaction graph, CIG) 表示, 然后使用图卷积网络 (graph convolutional network, GCN) 聚合各节点特征, 进而获得最终的匹配得分. CIG 是一个无向权重图, 图的节点被称为概念, 节点是由1个或1组高相关性的关键词组成. CIG 将长文本分解成若干个句子子集, 每个子集隶属于图中某个唯一的概念节点, 边的权重则表明2个概念之间相关性的大小. CIG 通过多种相似性计算方法构建节点的特征向量, 用于表达隶属于节点的2个文本间句子集合的相似性, 从而更好地捕获长文本间复杂的语义关系.

1.2 面向心理健康的自动问答

近年来已有学者对心理健康自动问答做了尝试,特别是对共情信息的生成^[1,18]有较高的热情.许多学者在用户情绪的指导下,采用基于 Transformer 的编码-解码框架^[19-20]或预训练语言模型^[21]生成共情信息.然而这些模型只能提供单一的心理支持信息(共情),心理支持效果有限.

另外,高质量数据集的匮乏也是在线心理支持自动问答的一个重要瓶颈.Sun 等人^[3]构建了一个用于心理健康问答的中文数据集,该数据集包含 22 000 个心理危机问题和 56 000 个高质量的回复(心理支持帖).他们还部分回复中的策略进行了标注,用于策略类型预测模型的训练,达到使用预测的策略类型指导回复生成的目的.评测结果显示,虽然生成的文本包含多种支持策略,并保持较好的可读性,但是生成的回复与问题的相关性不高,影响心理支持效果.

鉴于 BERT 在多项自然语言处理任务上的优异表现,He 等人^[22]使用疾病相关的知识(症状、诊断以及治疗信息等)对 BERT 进行再训练,获得融入领域知识的 diseaseBERT,提高模型在健康问答中的表现.然而一般疾病的问答与心理健康问答存在一定的差异,前者的问题非常明确,往往包含特定的疾病名词,且症状描述也有一定的规律可循,回复中也有相关药品名称或检查名称,而后者不具备该特性.因此即使使用融合了一般疾病知识的预训练模型,在面向心理支持的检索式社区自动问答上也存在巨大挑战.

2 融合角色心理画像的文本匹配模型 (T2CMP)

将问题形式化定义为:对于求助者发布的当前求助帖 p 和支持者发布的候选支持帖 r ,给定求助者和支持者发布的历史帖 c^p 和 c^r ,构建模型 $y=f(p, r, c^p, c^r)$ 预测 p 和 r 是否匹配,若匹配则 $y=1$,反之则 $y=0$,其中 c^p 和 c^r 分别用于刻画求助者和支持者的角色心理画像.

融合角色心理画像的文本匹配模型 T2CMP 的框架如图 2 所示,包含 2 个阶段 4 个模块.筛选阶段包括侧重语义层面匹配的候选支持帖筛选模块和角色心理画像生成模块,匹配阶段包括角色心理画像融合模块和决策模块.

2.1 筛选模块

对于求助帖 p ,筛选模块的目的是对心理健康社区上的支持帖进行筛选,过滤出语义相关、规模较小

的候选支持帖集合.

筛选是典型的信息检索问题,可选择使用效果较优的检索模型.从通用性和语义表达优越性 2 方面考虑,再与检索的经典算法 BM25 进行对比,最终选择 BERT 进行筛选,即:对于求助者发布的求助帖 p ,依次从社区中取出支持帖 r ,将 p 和 r 成对送进 BERT 进行相关性分类,类别为“相关”的 r 加入到 p 的候选支持帖集合 R^p 中.

为了尽可能保证理想的支持帖被筛选出来,训练分类模型时,选择带权重的交叉熵作为损失函数,通过设置权重来调整正负样例预测错误时对损失的影响,如式(1)所示:

$$Loss = \frac{\sum_i^N -[a^+ loss^+ + a^- loss^-]}{a^+ N^+ + a^- N^-}, \quad (1)$$

$$loss^+ = y_i \ln(\hat{y}_i),$$

$$loss^- = (1 - y_i) \ln(1 - \hat{y}_i).$$

其中 y_i 为第 i 个帖子对的真实相关性标签, \hat{y}_i 为分类模型对该标签预测出的概率值, N^+ , N^- 分别为训练集中正样例对和负样例对的个数, N 为训练集中总的帖子对个数, a^+ , a^- 分别是“相关”帖和“不相关”帖的损失权重.

设置“相关”帖被错分产生的损失要远大于“不相关”帖被错分所产生的损失,使得“相关”帖尽可能被筛选出来.最终的权重值设置为 $a^+ = 1$, $a^- = 0.01$.

2.2 角色心理画像模块

求助者/支持者 u 的心理画像来自其近期发布的历史帖 $c^u = \{t_1, t_2, \dots, t_n\}$, $u \in \{p, r\}$, 其中 n 是历史帖的个数.以 c^u 中每条帖子为基本单位,提取与心理健康相关的特征生成心理画像.受文献 [23-25] 启发,选择 LIWC 词类、词性和情绪作为心理画像的依据.LIWC 词典很好地涵盖了个体心理构成中许多重要的主题(例如家庭、健康、情感等),是经常被使用且被证明能有效评估个体的心理品质.另一方面许多研究者依据词性类别统计文本中每个类别出现的单词个数,根据词性的分布规律,分析个体的心理品质和写作风格.除此之外情绪是个体心理健康状态的重要组成部分,因此情绪可以作为推断个体心理健康的重要信号.

本文选择基于 LIWC 词典的词类频次特征(LF)、基于词性的词类频次特征(PF)和基于 Emo-BERT 的情绪特征(EF)刻画角色心理画像.

1) 基于 LIWC 词典的词类频次特征(LF).依据

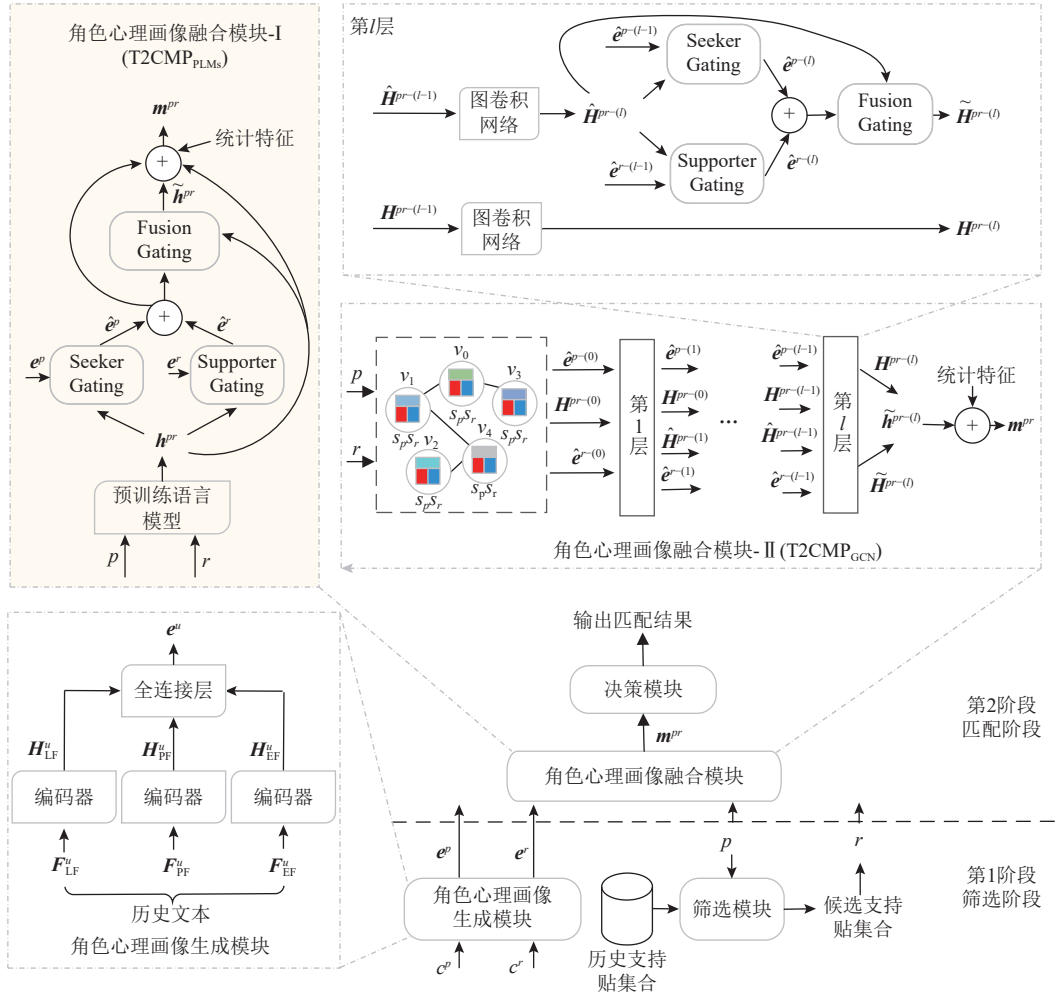


Fig. 2 The architecture of T2CMP

图2 T2CMP 架构

2007 版的 LIWC 词典, 统计 c^u 内每条帖子中各词类的频次, LF 的每一维对应 LIWC 词典中 1 种词类的频次, 共 64 维. 为了避免文本长度的影响, 使用归一化指数函数 (softmax) 对其进行处理, 形成最终的特征矩阵 F_{LF}^u .

2) 基于词性的词类频次特征 (PF). 与 LF 使用 LIWC 中的词类不同, PF 中使用词性表示不同的词类 (本文使用 NLTK 包的词性标注工具, 共 37 种词类), 得到特征矩阵 F_{PF}^u .

3) 情绪特征 (EF). 无论是求助帖还是支持帖, 用户表达的情绪是含蓄的、多样的、复杂的. 为此, 选择使用 Emo-BERT 模型来理解帖子中的情绪表达. 基于 plutchik-2 情绪轮^[26] 理论, 在 Twitter 数据集上标注了 8 类情绪 (喜悦、信任、恐惧、惊讶、悲伤、厌恶、愤怒和期待), 然后使用该标注数据集对 BERT 模型进行微调, 获得 Emo-BERT 模型. 在用户 u 的历史帖上使用 Emo-BERT, 得到情绪特征矩阵 F_{EF}^u .

使用 3 个独立的编码器对这 3 类特征进行编码, 并通过一个全连接层 (MLP) 对各类特征之间的关系进行编码, 获得最终的用户角色心理画像表示, 如式 (2) 所示. 根据不同种类特征的特点和历史经验, 用于 LF 和 PF 的编码器选择 CNN, 而用于 EF 的编码器选择 Transformer, 其中 $[]$ 表示向量的拼接, 用户的心理画像表示为 e^u .

$$\begin{aligned} H_{LF}^u &= f_{CNN}(F_{LF}^u), \\ H_{PF}^u &= f_{CNN}(F_{PF}^u), \\ H_{EF}^u &= f_{Transformer}(F_{EF}^u), \\ e^u &= f_{MLP}([H_{LF}^u, H_{PF}^u, H_{EF}^u]). \end{aligned} \quad (2)$$

2.3 用户角色心理画像融合模块

在对帖子进行表示时, 求助帖和支持帖均较长, 相较于短文本, 拥有更加复杂多变的语义、主题和关系等. 鉴于以上特点, 采用 2 种不同的表示方法, 即基于预训练语言模型 (pretrained language models, PLMs)

的表示和基于图的表示. 前者得益于大规模的预训练数据, 使其有较好的文本表示能力; 而后者将长文本拆分为局部信息(词集合或子句集合), 更加灵活有效地刻画长文本之间的局部匹配能力.

在融合角色心理画像时, 与大部分社区问答模型不同, 不能无差别地使用用户角色心理画像. 这是因为人是一种非常复杂的生物, 尤其是在心理和情绪方面. 人的心理危机问题不会一成不变, 甚至随着时间的推移, 可能会有新的心理危机问题产生. 换言之, 求助者的角色心理画像可能与多种心理危机问题相关, 但当前的求助帖是具有时效性的, 其可能只与求助者整体画像的一部分有关, 因此模型必须有选择、有限制地使用角色心理画像, 而非全盘无差别地让其影响求助帖和支持帖的文本表示. 同理, 心理支持者也可能拥有多种不同的专业偏好或兴趣, 角色心理画像中仅有部分信息与当前的帖子对相关.

鉴于以上考虑, 我们使用多层注意力机制分步融合角色心理画像和帖子对语义表示, 即: 先使用2种过滤门 Seeker Gating 和 Supporter Gating 分别对求助者和支持者的角色心理画像进行过滤, 然后再使用融合门(Fusion Gating)将过滤后的角色心理画像与帖子对的语义表示进行融合. 结合2种不同的语义表示机制, 提出2种不同的融合架构, 分别是基于预训练语言模型表示的 T2CMP_{PLMs} 和基于图表示的 T2CMP_{GCN}.

2.3.1 T2CMP_{PLMs}

首先使用属于 PLMs 的 BERT 获得帖子对的语义表示向量, 如式(3)所示:

$$\mathbf{h}^{pr} = f_{\text{BERT}}(\text{CLS}, p, \text{SEP}, r, \text{SEP}), \quad (3)$$

其中 CLS 是帖子对的起始符号, SEP 用于标识每条帖子的结束.

然后基于 \mathbf{h}^{pr} 对当前求助者和候选支持者的角色心理画像表示进行过滤, 保留与当前帖子对相关的用户角色心理画像信息 $\hat{\mathbf{e}}^p$ 和 $\hat{\mathbf{e}}^r$, 如式(4)(5)所示. 接着, 使用过滤后的角色心理画像 $\hat{\mathbf{e}}^p$ 和 $\hat{\mathbf{e}}^r$ 影响待匹配帖子对的匹配向量 \mathbf{h}^{pr} , 获得融合了角色心理画像的匹配向量 $\tilde{\mathbf{h}}^{pr}$, 如式(6)所示, 其中 \circ 表示 Hadamard 乘积, $\sigma(\cdot)$ 是激活函数, 此处选择 ReLU 作为激活函数.

$$\boldsymbol{\alpha}^p = \text{softmax}(\sigma(\mathbf{w}^p \mathbf{h}^{pr} + \mathbf{b}^p)), \quad (4)$$

$$\hat{\mathbf{e}}^p = \boldsymbol{\alpha}^p \circ \mathbf{e}^p,$$

$$\boldsymbol{\alpha}^r = \text{softmax}(\sigma(\mathbf{w}^r \mathbf{h}^{pr} + \mathbf{b}^r)), \quad (5)$$

$$\hat{\mathbf{e}}^r = \boldsymbol{\alpha}^r \circ \mathbf{e}^r,$$

$$\begin{aligned} \boldsymbol{\beta}^{pr} &= \text{softmax}(\sigma(\mathbf{w}^{pr} [\hat{\mathbf{e}}^p, \hat{\mathbf{e}}^r] + \mathbf{b}^{pr})), \\ \tilde{\mathbf{h}}^{pr} &= \boldsymbol{\beta}^{pr} \circ \mathbf{h}^{pr}. \end{aligned} \quad (6)$$

最后将匹配向量 (\mathbf{h}^{pr} 和 $\tilde{\mathbf{h}}^{pr}$)、过滤后的用户角色心理画像表示 ($\hat{\mathbf{e}}^p$ 和 $\hat{\mathbf{e}}^r$) 和基于句子的统计特征 SF 拼接成 \mathbf{m}^{pr} , 作为用户角色心理画像融合模块的输出. 统计特征 SF 的加入源于对数据集的观察, 即求助帖和候选支持帖中一些重要信息(心理危机问题、原因或背景、共情信息等)往往存在于疑问句或感叹句中, 因此我们选择一些关于句子的统计特征 SF 作为辅助特征. 这些特征包括: 帖子中疑问句、感叹句以及陈述句的个数, 帖子中句子总数和帖子中单词总数.

2.3.2 T2CMP_{GCN}

T2CMP_{GCN} 包括 CIG 的构建和角色心理画像的融合2个部分. 前者构建 p 和 r 的初始关系图, 后者通过多层图卷积融合角色心理画像.

1) CIG 的构建

①节点的构建. 首先, 以 TextRank^[27] 抽取 p 和 r 中的关键词为节点, 以词在句子中的共现为边, 构建基于词的文本图. 其次, 文本图上使用社群算法^[28], 将相关的关键词聚合形成概念, 作为 CIG 中的节点 v , 其中每个概念包含1个或多个相关的关键词. 最后, 将 p 和 r 以子句为单位, 计算子句与每个概念之间的 TF-IDF 相似度, 依据分值的大小决定子句的归属, 归属于节点 v 下的所有子句称之为节点 v 的内容. 节点的内容根据子句的来源分为2个子内容, 即 $S^p(v)$ 和 $S^r(v)$, 它们是局部匹配的基础. 需要注意的是, 如果某个子句与所有的概念节点的相似度均不高(即低于阈值), 则将其归属于不包含任何关键词的空概念节点.

②节点初始特征的构建. 使用2类语义匹配特征, 即基于词袋的局部匹配特征和基于神经网络的局部匹配特征作为 CIG 的节点初始特征. 第一, 基于词袋的局部匹配特征. 我们选择了5种经典的基于词袋的相似度计算方法, 即 TF-IDF、TF、BM25、基于 unigram 的 Jaccard 和 Ochiai, 使用它们计算 $S^p(v)$ 和 $S^r(v)$ 的语义相似度, 并将这些相似度分值拼接为基于词袋的局部匹配特征 $\mathbf{x}_{\text{TB}}^{pr}(v)$. 除此之外, 额外增加2维指示特征用于指示该概念节点 v 是否是空概念节点, 以及该节点的2个子内容是否都不为空. 第二, 基于神经网络的局部匹配特征. 神经网络强大的表示能力可以在一定程度上弥补传统基于词袋的文本表示的不足. 首先根据 $S^p(v)$ 和 $S^r(v)$ 的内容, 以预训练的 Word2Vec 词向量为输入, 使用卷积神经网络(CNN)学习它们的潜在语义表示 $\mathbf{h}^p(v)$ 和 $\mathbf{h}^r(v)$. 然后根据式(7)计算基

于神经网络的局部匹配特征 $\mathbf{x}_{\text{NN}}^{pr}(v)$. 考虑到运算效率, 所有节点共享同一个卷积神经网络. 将 $\mathbf{x}_{\text{TB}}^{pr}(v)$ 和 $\mathbf{x}_{\text{NN}}^{pr}(v)$ 拼接为节点 v 的初始特征 $\mathbf{x}^{pr}(v)$. 所有节点的初始特征构成 CIG 的特征矩阵 \mathbf{X} .

$$\mathbf{x}_{\text{NN}}^{pr}(v) = [\mathbf{h}^p(v) - \mathbf{h}^r(v), \mathbf{h}^p(v) \circ \mathbf{h}^r(v)]. \quad (7)$$

③边及其权重的构建. 依据 2 个节点的内容计算它们的 TF-IDF 相似度, 若该相似度分值超过阈值, 则为这 2 个节点生成边, 相似度分值作为该边的权重.

2) 角色心理画像的融合

每层角色心理画像的融合包括卷积和融合 2 个操作, 卷积操作整合 CIG 中不同节点的信息, 融合操作引入角色心理画像, 其中卷积操作需分别对未融合角色心理画像的节点特征和融合角色心理画像的节点特征进行卷积.

①卷积操作. 使用 GCN 对节点特征进行更新, 其输入为 CIG, 即 $G=(V, E)$, 其中 $v_i \in V$ 表示第 i 个概念节点, $e_{ij} \in E$ 表示节点 i 和 j 之间的边, 边的权重记为 w_{ij} . 将节点特征矩阵 \mathbf{X} 作为图的输入, 如式(8)所示:

$$\begin{aligned} \mathbf{H}^{pr-(0)} &= \mathbf{X}, \quad \hat{\mathbf{H}}^{pr-(0)} = \mathbf{X}, \\ \mathbf{H}^{pr-(l)} &= f_{\text{GCN}}(\mathbf{H}^{pr-(l-1)}), \\ \hat{\mathbf{H}}^{pr-(l)} &= f_{\text{GCN}}(\hat{\mathbf{H}}^{pr-(l-1)}). \end{aligned} \quad (8)$$

为区分节点向量是否融合角色心理画像, 将未融合角色心理画像的第 l 层隐藏向量输出命名为 $\mathbf{H}^{pr-(l)}$, 而融合角色心理画像的第 l 层隐藏向量输出命名为 $\hat{\mathbf{H}}^{pr-(l)}$, 2 个 GCN 网络共享可训练参数.

②融合操作. 使用多层注意力将用户角色心理画像与节点特征进行融合, 具体过程如下: 首先根据第 l 层输出 $\hat{\mathbf{H}}^{pr-(l)}$ 计算出不同节点的局部信息对用户角色心理画像的注意力得分 $\alpha^{p-(l)} \in \mathbb{R}^{|V| \times 1}$, $\alpha^{r-(l)} \in \mathbb{R}^{|V| \times 1}$, 如式(9)(10)所示; 然后根据注意力得分对用户角色心理画像进行调整, 如式(11)(12)所示; 接着将过滤后的 $\hat{\mathbf{e}}^{p-(l)}$ 和 $\hat{\mathbf{e}}^{r-(l)}$ 拼接后作为输入, 用于计算注意力权重 $\beta^{pr-(l)}$, 如式(13)所示, 揭示每个局部匹配特征与求助者和候选支持者角色心理画像的相关性; 最后获得融合了用户角色心理画像的局部匹配特征矩阵 $\tilde{\mathbf{H}}^{pr-(l)}$, 如式(14)所示, 其中 \otimes 表示 Kronecker 乘积.

经过 L 层获得节点特征的高阶表示, 即 $\mathbf{H}^{pr-(L)}$ 和 $\hat{\mathbf{H}}^{pr-(L)}$. 其中 $|L|$ 表示 GCN 网络的层数. $\tilde{\mathbf{h}}^{pr-(L)}$ 是 $\mathbf{H}^{pr-(L)}$ 和 $\hat{\mathbf{H}}^{pr-(L)}$ 拼接后按列最大池化的结果, 再将其与 SF 拼接形成匹配表示向量 \mathbf{m}^{pr} .

值得注意的是, 在 T2CMP_{GCN} 中, 由于用户的角色心理画像与待匹配帖子对的各局部信息经过了多层 GCN 的融合, 所以不再像 T2CMP_{PLMs} 一样将用户

的角色心理画像拼接进 \mathbf{m}^{pr} .

$$\alpha^{p-(l)} = \text{softmax}(\mathbf{w}^{p-(l)} \tanh(\mathbf{W}^{p-(l)} \hat{\mathbf{H}}^{pr-(l)} + \mathbf{b}^{p-(l)})), \quad (9)$$

$$\alpha^{r-(l)} = \text{softmax}(\mathbf{w}^{r-(l)} \tanh(\mathbf{W}^{r-(l)} \hat{\mathbf{H}}^{pr-(l)} + \mathbf{b}^{r-(l)})), \quad (10)$$

$$\hat{\mathbf{e}}^{p-(0)} = \mathbf{e}^p, \quad \hat{\mathbf{e}}^{p-(l)} = \alpha^{p-(l)} \otimes \hat{\mathbf{e}}^{p-(l-1)}, \quad (11)$$

$$\hat{\mathbf{e}}^{r-(0)} = \mathbf{e}^r, \quad \hat{\mathbf{e}}^{r-(l)} = \alpha^{r-(l)} \otimes \hat{\mathbf{e}}^{r-(l-1)}, \quad (12)$$

$$\beta^{pr-(l)} = \text{softmax}(\mathbf{w}^{pr-(l)} \tanh(\mathbf{W}^{pr-(l)} [\hat{\mathbf{e}}^{p-(l)}, \hat{\mathbf{e}}^{r-(l)}] + \mathbf{b}^{pr-(l)})), \quad (13)$$

$$\tilde{\mathbf{H}}^{pr-(l)} = \beta^{pr-(l)} \otimes \hat{\mathbf{H}}^{pr-(l)}. \quad (14)$$

2.4 决策模块

决策模块由 2 个全连接层和激活函数组成, 其中每个线性层的输出均需经过激活函数. \mathbf{m}^{pr} 送入决策模块后, 输出最终的预测结果.

3 实验结果与分析

3.1 数据集

实验数据集来自 CLPsych2017 Workshop 上的 shared task. 该任务的目标是自动检测帖子所反映的用户心理危机程度, 使用的数据集取自 2012 年 1 月到 2017 年 3 月期间在线论坛 ReachOut 上的帖子, 该论坛为受心理健康困扰的网络用户提供帮助, 用户发布求助帖后, 会由心理咨询师或有经验的用户回复, 为其提供帮助. 因此, 在采集的数据中, 以 1 次回复中的首个帖子为求助帖, 对求助帖的回复为相应的支持帖. 数据集中部分帖子被人工标注了 crisis, red, amber, green 这 4 类标签, 其中非 green 类 (crisis, red, amber) 类标签表示该帖子作者存在心理危机, 需要干预, 是本文需要关注的求助帖.

尽管该数据集中有 657 条帖子被标注为 crisis, red, amber, 但仅有 86 条为求助帖 (1 次会话中的首条帖子). 为扩展数据集, 本文使用 F³TMH 模型^[23]对人工标注的帖子进行分类, 该模型在 CLPsych2017 Workshop 的 shared task 中, 对 green 和非 green 帖子的识别准确率取得最高值, 达到 0.955, 得到求助帖 2 584 条, 过滤非 green 类标签的回复后, 得到相应的支持帖 (回复) 17 848 条.

对于 MHCQA 数据集的构建, 针对每条求助帖, 人工回复的支持帖作为正样例, 随机抽取其他支持帖作为负样例. 参照一般社区问答的负样例数量^[29], 设定每条求助帖的负样例数量是其正样例数量的 16 倍.

考虑到自动标注与人工标注仍然存在细微的差距, 为检测模型在真实场景下的性能, 将人工标注的

86 条求助帖及其正样例、负样例作为测试集, 而将扩展的求助帖及其正样例、负样例作为训练集. 数据集的统计数据如表 1 所示.

Table 1 Statistics of MHCQA Dataset

表 1 MHCQA 数据集统计

统计信息	训练集	测试集
正样例对数	13 808	555
负样例对数	187 054	8 736
求助帖长度单词数 (最短/最长/平均)	9/1 880/269	7/808/200
支持帖长度单词数 (最短/最长/平均)	4/1 899/115	4/1 393/113

3.2 模型参数设置

1) 用户角色心理画像生成模块

c' 使用的历史帖个数为 10. 由于统计特征和语言特征为 one-hot 向量, 因此用于这 2 种特征的编码器是单层 CNN, 使用多个卷积核, 卷积核大小分别是 1, 2, 3. 考虑到求助者心理健康状态对时间更为敏感, 故针对求助者只使用大小为 1 和 2 的卷积核, CNN 的输出维度均为 128. 而情绪特征是由 Emo-BERT 获得的 768 维特征, 我们采用 Transformer 作为情绪特征的编码器. 其注意力个数和 dropout 分别为 2 和 0.5, 输出维度设置为 128. 全连接层输出 128 维的用户角色心理画像表示向量.

2) T2CMP_{PLMs}

T2CMP_{PLMs} 使用 12 层的 BERT 模型, 输出 768 维的表示向量. 训练中学习率设置为 0.001, batchsize 和最大训练代数分别设置为 32 和 100.

3) T2CMP_{GCN}

T2CMP_{GCN} 的图构建中, 节点内容隶属的相似度阈值和边构建阈值均设置为 0.001, GCN 的层数 $L=2$, 每层 GCN 的输出维度均为 16, dropout 为 0.1. 初始词向量使用维度为 200 的预训练 Word2Vec, CLPsych2017 shared task 和 CLPsych2019 shared task 数据集共同参与了该词向量的预训练过程. 对于模型中用于获取局部语义匹配特征的 CNN, 其输出为 32 维, 同时将 ReLU 作为激活函数并且添加最大池化层, 各节点共用 CNN 的参数.

与 T2CMP_{PLMs} 相比, T2CMP_{GCN} 对用户角色心理画像的使用有一定的区别, 因此对部分参数做了相应的调整, 每个用户历史帖个数增加为 15. CNN 采用单个卷积核, 其大小为 1. 鉴于图表示的输出维度为 32, 因此 CNN 的输出维度和情绪特征的编码器维度分别调整为 16 和 32, 且最终输出 16 维的用户角色心

理画像表示. 采用动态的学习率, 即在每次迭代的前 1 000 步中, 学习率从 0.001 开始以指数方式递减. 最大训练迭代次数为 100.

4) 决策器

决策器由 2 层全连接层和 sigmoid 层组成, 其中第 1 层的全连接输出为 16 维, 使用 ReLU 激活后送入下一层.

在实验中使用参数 $\beta_1=0.8$, $\beta_2=0.999$, $\epsilon=108$ 的 ADAM 优化器.

3.3 对比方法

本文选择了 6 个具有竞争力的模型作为对比方法.

1) ARC-I^[8]. 该模型使用 CNN 分别学习待判断的 2 个文本表示特征, 然后使用学习到的文本表示特征进行相关性判断.

2) ARC-II^[8]. 与 ARC-I 不同, 提前让待匹配的 2 个文本进行交互, 即在学习文本表示时就考虑了 2 个文本的交互.

3) MatchPyramid^[4]. 该模型是经典的交互型神经网络模型, 与表示型神经网络模型不同, 该模型以待匹配文本词级别的相关矩阵作为输入.

4) CIG-GCN^[5]. 用 CIG 表达长文本以及长文本间的语义关系, 然后使用 GCN 进行特征转换, 提高模型对于长文本的匹配判断.

5) BERT^[6]. 针对英文数据集的 12 层基础 BERT 模型, 通过训练集对其进行微调 (fine-tune). 在训练时, 输入为求助帖和候选支持帖, 输出为该帖子对的预测标签, 学习率设为 0.000 05, 训练时的 batchsize 与测试时的 batchsize 不同, 分别为 32 和 8.

6) diseaseBERT^[22]. 在 BERT 中融入专业领域知识用于医疗保健的问答. 在使用 diseaseBERT 进行测试前, 使用当前数据集的训练集对其进行 fine-tune, fine-tune 时使用的参数与 BERT 相同.

3.4 实验结果与分析

鉴于实验任务是二分类任务, 并且任务的目的是为求助帖检索合适的支持帖, 并不关心负样例的预测效果, 因此采用正样例的精确率 (precision)、召回率 (recall) 和 F1 值作为实验的评价指标.

表 2 展示了三大类模型的实验评价结果, 即单阶段模型 R1(), 2 阶段模型 R2() 和融合角色心理画像的 2 阶段模型 T2CMP, 其中 T2CMP_{1a} 只使用单层注意力机制, 即没有对求助者/支持者的角色心理画像进行过滤, 直接融合原始角色心理画像和待匹配的文本对匹配向量. R1(BERT), R1(diseaseBERT) 分别表示使用 BERT 和 diseaseBERT 的单阶段匹配模型. 从

Table 2 Assessment Results of Different Models**表 2 各模型的评测结果**

模型	精确率	召回率	F1 值
R1(BERT)	0.664 5	0.364 0	0.470 3
R1(diseaseBERT)	0.707 3	0.313 5	0.434 5
R2 (ACR-I)	0.476 4	0.182 0	0.263 4
R2 (ACR-II)	0.350 0	0.378 4	0.363 6
R2 (MatchPyramid)	0.392 3	0.475 7	0.430 0
R2 (CIG-GCN)	0.676 2	0.383 8	0.489 7
R2 (BERT)	0.612 0	0.457 7	0.523 7
R2 (diseaseBERT)	0.642 9	0.454 1	0.532 2
T2CMP _{GCN_1a} (本文)	0.604 1	0.423 4	0.497 9
T2CMP _{PLMs(BERT)_1a} (本文)	0.635 4	0.452 3	0.528 4
T2CMP _{PLMs(diseaseBERT)_1a} (本文)	0.653 7	0.455 9	0.537 2
T2CMP _{GCN} (本文)	0.605 8	0.454 1	0.519 1
T2CMP _{PLMs(BERT)} (本文)	0.643 4	0.464 9	0.539 7
T2CMP _{PLMs(diseaseBERT)} (本文)	0.696 0	0.441 4	0.540 2

注：加粗数字表示最优结果。

表 2 可以看出, 当大量的非相关负样例存在时, 带来了大量的噪声, 基于单阶段的匹配模型对正样例的预测具有不错的精确率, 但召回率较低, 使得模型的整体匹配效果一般, 其中 R1(BERT) 的实验结果最好, F1 值为 0.4703。

对于所有的 2 阶段匹配模型 R2(), 它们的筛选模块均选用 BERT, 括号内为第 2 阶段使用的模型。第 1 阶段中, 首先使用 MHCQA 数据集中的训练集对 BERT 进行 fine-tune, 选择带权重的交叉熵作为训练时的损失函数。然后使用 fine-tune 后的 BERT 分别对 MHCQA 数据集中的训练集和测试集进行帖子筛选, 形成用于第 2 阶段的候选支持帖集合 (MHCQA-II) 的训练集和测试集, 如表 3 所示。从表 3 中可以看出, 筛选模块已经筛选了大量的负样例对。

Table 3 Statistics of MHCQA-II Dataset**表 3 MHCQA-II 的数据集统计**

统计信息	训练集	测试集
正样例对数	15 651	360
负样例对数	15 840	688

表 2 中的评测结果显示, 经过第 1 阶段的筛选后, 去除了大量的噪声, 其最佳匹配效果相比于单阶段匹配模型得到了明显的改善。例如 R2 (BERT), R2 (diseaseBERT) 比 R1(BERT), R1(diseaseBERT) 的 F1 值分别提高了 5.34 个百分点和 9.77 个百分点。在众多

对比方法中, 擅长长文本语义理解和匹配的 R2(CIG-GCN) 的预测精确率最高, 达到了 0.676 2。

使用多层注意力机制引入用户角色心理画像可以进一步提高模型性能, 其中 T2CMP_{GCN} 与未融合角色心理画像的 R2(CIG-GCN) 相比, 其 F1 值从 0.489 7 提高到 0.519 1; 相比于 R2(BERT), T2CMP_{PLMs(BERT)} 则将 F1 值提高了 1.6 个百分点 (0.523 7 到 0.539 7)。除此之外, 还使用 diseaseBERT 替代 BERT, 即 T2CMP_{PLMs(diseaseBERT)}, 其效果最好, F1 值达到了 0.540 2, 虽然只比未融合角色心理画像的 R2(diseaseBERT) 略高了 0.8 个百分点, 但其对正样例预测的精确率提高了 5.31 个百分点。使用单层注意力机制时 T2CMP 的效果有所下降, T2CMP_{GCN_1a} 比 T2CMP_{GCN} 的 F1 值下降了 2.12 个百分点, 相比于 T2CMP_{PLMs(BERT)} 和 T2CMP_{PLMs(diseaseBERT)}, T2CMP_{PLMs(BERT)_1a}, T2CMP_{PLMs(diseaseBERT)_1a} 分别下降了 1.13 个百分点、0.3 个百分点。这进一步验证了过滤求助者/支持者角色心理画像的必要性。

附录 A 列举了 2 个示例, 更直观地表明求助者/支持者角色心理画像对于模型的匹配具有一定的帮助。

3.5 消融实验

用户的角色心理画像已被证明能有效地提高模型的匹配效果。为了进一步考察各种心理健康相关特征对用户角色心理画像质量的影响, 以及基于句子的统计特征 SF 的作用, 分别对 T2CMP_{PLMs(BERT)} 和 T2CMP_{GCN} 进行消融实验。

表 4 展示了基于 LIWC 词典的词类频次特征 (LF)、基于词性的词类频次特征 (PF)、情绪特征和统计特征 (SF) 对 T2CMP_{PLMs(BERT)} 模型的影响。从结果来看, 减少各特征会影响角色心理画像刻画的质量, 从而对于那些语义相关性不大且心理健康状态比较

Table 4 Ablation Experiment Results of T2CMP_{PLMs(BERT)} Model**表 4 T2CMP_{PLMs(BERT)} 模型的消融实验结果**

模型	精确率	召回率	F1 值
T2CMP _{PLMs(BERT)} (本文)	0.643 4	0.464 9	0.539 7
w/o LF	0.654 5↑	0.450 5↓	0.533 6
w/o PF	0.642 1↓	0.455 9↓	0.533 2
w/o EF	0.650 5↑	0.459 5↓	0.538 5
w/o LF, PF	0.651 2↑	0.454 1↓	0.535 0
w/o LF, EF	0.639 6↓	0.454 1↓	0.531 1
w/o PF, EF	0.647 8↑	0.454 1↓	0.533 9
w/o SF	0.647 2↑	0.459 5↓	0.537 4

注：“↑”“↓”分别代表其前面的数值与本文所提模型相应结果的比较趋势是上升和下降。

隐晦的正样例无法正确识别,导致最终预测的正样例总数量减少,召回率下降.与此同时,由于模型仅识别出语义匹配和心理健康匹配较明显的支持帖,因此精确率普遍上升.

从表 5 可以看出,LF,PF,EF 对 $T2CMP_{GCN}$ 也有影响,剔除任何 1 个或 2 个特征,模型最终的 $F1$ 值均有较为明显的下降.对于与心理健康相关的特征,从依次减少的单个特征可以看出,EF 和 LF 相较于 PF 对模型的影响更大.若不使用 SF,模型的 $F1$ 值下降了 1.18 个百分点.

Table 5 Ablation Experiment Results of $T2CMP_{GCN}$

表 5 $T2CMP_{GCN}$ 模型的消融实验结果

模型	精确率	召回率	$F1$ 值	$F1$ 值下降百分点
$T2CMP_{GCN}$ (本文)	0.605 8	0.454 1	0.519 1	
w/o LF	0.606 1↑	0.427 0↓	0.501 1	1.8
w/o PF	0.606 3↑	0.452 3↓	0.518 1	0.1
w/o EF	0.585 0↓	0.434 2↓	0.498 4	2.07
w/o LF, PF	0.619 6↑	0.410 8↓	0.494 0	2.51
w/o LF, EF	0.617 0↑	0.432 4↓	0.508 5	1.06
w/o PF, EF	0.612 2↑	0.398 2↓	0.482 5	3.66
w/o SF	0.677 7↑	0.405 4↓	0.507 3	1.18

注:“↑”“↓”分别代表其前面的数值与本文所提模型相应结果的比较趋势是上升和下降.

3.6 候选支持帖规模对 $T2CMP$ 的影响

2 阶段模型 $T2CMP$ 的效果还受到候选支持帖集合 MHCQA-II 规模的影响,为此,本节进一步考察不同的筛选力度对 $T2CMP$ 的影响.通过使用不同权重的损失函数来调节模型对正样例的召回能力,形成不同规模的候选集合,如表 6 所示.从表 6 可以看出,当 $a^+ = 0.1$ 时,第 1 阶段的 BERT 模型在测试集中只召回了 255 对正样例.减小 a^+ ,第 1 阶段的模型可以找回更多正样例,但不幸的是,负样例数量增长得更快.

Table 6 Statistics of MHCQA-II formed by Loss Functions with Different Weights

表 6 不同权重的损失函数形成的 MHCQA-II 的统计

数据集	训练集对数		测试集对数	
	正样例	负样例	正样例	负样例
MHCQA-II ($a^+ = 1, a^- = 0.01$)	15 651	15 840	360	688
MHCQA-II ₂ ($a^+ = 1, a^- = 0.1$)	12 069	5 336	255	236
MHCQA-II ₃ ($a^+ = 1, a^- = 0.001$)	13 471	40 134	418	2 055

对于筛选阶段选出的不同规模 MHCQA-II,各模型的表现如表 7 所示.

Table 7 Assessment Results of Different Models Based on MSOIP-II₂, MSOIP-II₃

表 7 基于 MSOIP-II₂, MSOIP-II₃ 的各模型评测结果

模型	MHCQA-II ₂			MHCQA-II ₃		
	精确率	召回率	$F1$ 值	精确率	召回率	$F1$ 值
R2 (ACR-I)	0.519 4	0.457 7	0.486 6	0	0	0
R2 (ACR-II)	0.546 1	0.437 8	0.486 0	0.177 0	0.327 9	0.229 9
R2 (MatchPyramid)	0.530 3	0.457 7	0.491 3	0.182 5	0.228 8	0.203 0
R2 (CIG-GCN)	0.621 1	0.392 8	0.481 2	0.694 6	0.209 0	0.321 3
R2 (BERT)	0.641 5	0.412 6	0.502 2	0.684 6	0.367 6	0.478 3
R2 (diseaseBERT)	0.604 9	0.400 0	0.481 6	0.687 9	0.389 2	0.497 1
$T2CMP_{GCN}$ (本文)	0.657 2	0.376 6	0.478 8	0.714 3	0.216 2	0.332 0
$T2CMP_{PLMs(BERT)}$ (本文)	0.686 9	0.387 4	0.495 4	0.678 0	0.384 6	0.498 9
$T2CMP_{PLMs(diseaseBERT)}$ (本文)	0.705 3	0.362 2	0.478 6	0.668 7	0.385 6	0.489 1

注:加粗数字表示最优结果.

对于规模较小的 MHCQA-II₂,虽然所有模型的 $F1$ 值波动性不大(0.478 6~0.502 2),但各模型的精确率还是有较大差异.尤其是使用多层注意力融合用户角色心理画像的模型,即 $T2CMP_{PLMs(BERT)}$, $T2CMP_{GCN}$, $T2CMP_{PLMs(diseaseBERT)}$,精确率相较于其他模型均有提升,但受限于筛选阶段返回的正样例个数, $F1$ 值不如在 MHCQA-II 上的表现.

综上所述,无论筛选阶段返回的候选支持帖的规模如何,融入用户角色心理画像均能改善模型的性能.从各模型在 3 个不同规模的数据集上的表现来看,当第 1 阶段返回的候选支持帖中正样例和负样例的数量比较均衡时(MHCQA-II),可以充分发挥 2 阶段模型的优势,使得 $T2CMP$ 获得较优的结果.

由于 BM25 在信息检索方面的高效率,我们还尝试使用其作为第 1 阶段的检索模型,其检索效果如表 8 所示.从表 8 可以看出,当返回相同数量的候选支持帖时, BM25 的正样例返回数远低于 BERT 第 1 阶段过低的正样例返回数会极大地限制模型的最终匹配效果.

Table 8 Retrieval Results of BM25 and BERT Models

表 8 BM25 和 BERT 模型的检索结果

模型	训练集对数		测试集对数	
	正样例	负样例	正样例	负样例
BERT ($a^+ = 1, a^- = 0.01$)	15 651	15 840	360	688
BM25	7 102	24 389	227	821

4 总 结

当存在大量候选支持帖时,本文提出的融合角

色心理画像的2阶段文本匹配模型(T2CMP)能有效地提高模型的匹配效果,它将面向心理健康文本的匹配任务拆分为2个连续的子任务.2个子任务的目的不同,第1阶段的子任务是为了最大限度地筛除语义不相关的支持帖,形成一个较小的候选支持帖集合.为了避免过多的理想支持帖被筛除,并且由于它的判断结果不是最终的结果,因此通过减小模型误将非理想支持帖判定为理想支持帖的惩罚力度,使得模型对理想支持帖有更高的召回率.第2阶段子任务的目的是对第1阶段筛选出的候选支持帖增加心理层面的匹配判断,修正语义匹配的结果.由于候选支持帖均与求助帖有一定的相关性,单纯依赖语义匹配是不够的,T2CMP根据求助者和候选支持者的历史帖,刻画求助者/候选支持者的角色心理画像,了解更多的信息(例如心理健康问题的原因,心理支持者的专业偏好、策略偏好等),利用这些信息辅助模型对求助帖和支持帖进行精确匹配判断以获得更好的匹配效果.

实验显示,T2CMP的效果不仅优于基于单阶段的匹配模型,而且对于第2阶段的子任务,通过多层注意力融合用户角色心理画像的模型要优于其他当前较为先进的对比模型.除此之外,我们还通过适应性实验验证了第1阶段的检索效果对第2阶段模型的影响.

面向心理健康的文本匹配依然面临巨大挑战,例如支持帖中仅包含通用性的“支持方法或建议”会导致语义匹配失效、求助者求助信息有效内容少或支持者支持偏好的缺失,使T2CMP无法利用上下文等.

以ChatGPT为代表的大语言模型具有较强的文本生成能力,能生成自然流畅且逻辑性好的文本.在社区问答任务中,面对广泛且多样的问题,大语言模型能较好地理解语境,快速生成相关性高的回复,为用户带来良好的体验感.然而,当面临关于心理健康问题的求助文本时,ChatGPT通常返回的是笼统的、普适的、缺乏个性化或共情的建议或方法.此外,直接将ChatGPT生成的结果返回给被心理问题困扰的求助者,也面临伦理挑战.虽然ChatGPT主要用于文本生成,无法直接应用于面向心理健康的文本匹配,然而由于其良好的对话能力,使用ChatGPT可以提升用户的倾诉欲望,获得更为丰富和完整的心理画像,提升心理健康文本的匹配效果.此外,心理支持者还可以先利用ChatGPT自动生成相应的支持消息,然后对生成的消息进行评估和修改,提升回复效率.

作者贡献声明:赵芸收集资料,负责研究方案的构思、设计和实现,以及论文撰写及修订;刘德喜提供研究思路、论文组织结构的设计、论文审阅及修订、全过程监督;万常选、刘喜平和廖国琼负责论文审阅及修订.

参 考 文 献

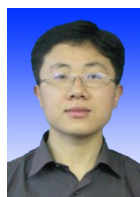
- [1] Sharma A, Lin I, Miner A S, et al. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach[C]//Proc of the 30th Web Conf. New York: ACM, 2021: 194–05
- [2] Liu Siyang, Zheng Chujie, Demasi O, et al. Towards emotional support dialog systems[C]//Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2021: 3469–3483
- [3] Sun Hao, Lin Zhenru, Zheng Chujie, et al. PsyQA: A Chinese dataset for generating long counseling text for mental health support[C]//Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2021: 148–1503
- [4] Pang Liang, Lan Yanyan, Guo Jiafeng, et al. Text matching as image recognition[C]//Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 2793–2799
- [5] Liu Bang, Niu Di, Wei Haojie, et al. Matching article pairs with graphical decomposition and convolutions[C]//Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 6284–6294
- [6] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1 : Long and Short Papers). Stroudsburg, PA: ACL, 2019: 4171–4186
- [7] Zhao Yun, Liu Dexi, Wan Changxuan, et al. Find supports for the post about mental issues: More than semantic matching[J]. Transactions on Asian and Low-Resource Language Information Processing, 2022, 21(6): 1–14
- [8] Hu Baotian, Lu Zhengdong, Li Hang, et al. Convolutional neural network architectures for matching natural language sentences[C]//Proc of the 28th Conf on Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2014: 2042–2050
- [9] Qiu Xipeng, Huang Xuanjing. Convolutional neural tensor network architecture for community-based question answering[C]//Proc of the 24th Int Joint Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2015: 1305–1311
- [10] Wan Shengxian, Lan Yanyan, Guo Jiafeng, et al. A deep architecture for semantic matching with multiple positional sentence representations[C]//Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 2374–3468
- [11] Liu Bang, Zhang Ting, Han F, et al. Matching natural language sentences with hierarchical sentence factorization[C]//Proc of the 27th Web Conf. New York: ACM, 2018: 1237–1246

- [12] Zhang Xiaodong, Li Sujian, Sha Lei, et al. Attentive interactive neural networks for answer selection in community question answering[C]//Proc of the 31st AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2017: 3525–3531
- [13] Wu Yu, Wu Wei, Xu Can, et al. Knowledge enhanced hybrid neural network for text matching[C]//Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2018: 5586–5593
- [14] Jiang J, Zhang Mingyang, Li Cheng, et al. Semantic text matching for long-form documents[C]//Proc of the 28th Web Conf. New York: ACM, 2019: 795–806
- [15] Nikolentzos G, Meladianos P, Rousseau F, et al. Shortest-path graph kernels for document similarity[C]//Proc of the 22nd Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2017: 1890–1900
- [16] Paul C, Rettinger A, Mogadala A, et al. Efficient graph-based document similarity[C]//Proc of the 13th European Semantic Web Conf. Berlin: Springer, 2016: 334–349
- [17] Sun Yawei, Cheng Gong, Li Xiao, et al. Graph matching network for interpretable complex question answering over knowledge graphs[J]. *Journal of Computer Research and Development*, 2021, 58(12): 2673–2683 (in Chinese)
(孙亚伟, 程龚, 厉肖, 等. 基于图匹配网络的可解释知识图谱复杂问答方法[J]. *计算机研究与发展*, 2021, 58(12): 2673–2683)
- [18] Kim H, Kim B, Kim G. Perspective-taking and pragmatics for generating empathetic responses focused on emotion causes[C]//Proc of the 26th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2021: 2227–2240
- [19] Xie Yubo, Pu P. Empathetic dialog generation with fine-grained intents[C]//Proc of the 25th Conf on Computational Natural Language Learning. Stroudsburg, PA: ACL, 2021: 133–147
- [20] Li Qintong, Li Piji, Ren Zhaochun, et al. Knowledge bridging for empathetic dialogue generation[C]//Proc of the 36th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2022: 10993–11001
- [21] Zheng Chujie, Liu Yong, Chen Wei, et al. CoMAE: A multi-factor hierarchical framework for empathetic response generation[C]//Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2021: 813–824
- [22] He Yun, Zhu Ziwei, Zhang Yin, et al. Infusing disease knowledge into BERT for health question answering, medical inference and disease name recognition[C]//Proc of the 25th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2020: 4604–4614
- [23] Liu Dexi, Xia Xianyi, Wan Changxuan, et al. Mental health assessment for online forum users based on multi-feature fusion[J]. *Chinese Journal of Computers*, 2019, 42(7): 1553–1569 (in Chinese)
(刘德喜, 夏先益, 万常选, 等. 基于多特征融合的在线论坛用户心理健康自动评估[J]. *计算机学报*, 2019, 42(7): 1553–1569)
- [24] 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): 1041–4347
- [25] Aladağ A E, Muderrisoglu S, Akbas N B, et al. Detecting suicidal ideation on forums: Proof-of-concept study[J]. *Journal of Medical Internet Research*, 2018, 20(6): e215
- [26] Abdul-Mageed M, Ungar L. EmoNet: Fine-grained emotion detection with gated recurrent neural networks[C]//Proc of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2017: 718–728
- [27] Mihalcea R, Tarau P. TexTrank: Bringing order into text[C]//Proc of the 9th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2004: 404–411
- [28] Sayyadi H, Raschid L. A graph analytical approach for topic detection[J]. *ACM Transactions on Internet Technology*, 2013, 13(2): 1–4
- [29] Shen Yikang, Rong Wenge, Jiang Nan, et al. Word embedding based correlation model for question/answer matching[C]//Proc of the 31st AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2017: 2374–3468



Zhao Yun, born in 1985. PhD, lecturer. Her main research interests include social media processing and natural language processing.

赵芸, 1985年生. 博士, 讲师. 主要研究方向为社会媒体处理和自然语言处理.



Liu Dexi, born in 1975. PhD, professor, PhD supervisor. His main research interests include social media processing, information retrieval, and natural language processing.

刘德喜, 1975年生. 博士, 教授, 博士生导师. 主要研究方向为社会媒体处理、信息检索、自然语言处理.



Wan Changxuan, born in 1962. PhD, professor, PhD supervisor. His main research interests include Web data management and sentiment analysis.

万常选, 1962年生. 博士, 教授, 博士生导师. 主要研究方向为Web数据管理、情感分析.



Liu Xiping, born in 1981. PhD, professor, PhD supervisor. His main research interests include Web data management and text mining.

刘喜平, 1981年生. 博士, 教授, 博士生导师. 主要研究方向为Web数据管理、文本挖掘.



Liao Guoqiong, born in 1969. PhD, professor, PhD supervisor. His main research interests include social network mining and IoT data management.

廖国琼, 1969年生. 博士, 教授, 博士生导师. 主要研究方向为社交网络挖掘、物联网数据管理.

附录 A. 实验示例分析.

从图 A1 的例子可以看出, 求助者披露的关于心理危机问题描述(加粗倾斜的内容)与支持帖中的专业建议或方法(加粗倾斜的内容)很难产生关联. 若是结合求助者的历史帖, 可以获知他/她感觉压抑的原因是因为与心理咨询师的沟通出了问题, 无法从咨询师那里获得支持和安慰, 同时学校近期的一些活动(例如毕业旅行)要与同学沟通相处, 也给他/她带来很大的压力. 充分使用这些历史帖可以获得求助者的角色心理画像, 在求助者角色心理画像的帮助下, 有助于模型跨越当前求助帖和候选支持帖中的语义鸿沟, 获得正确的判断.

图 A2 的例子略有不同, 虽然求助帖中关于心理危机问题的描述以及背景比较详细, 但是支持帖中的专业建议或方法仍然与其存在一定的语义鸿沟. 模型可以从当前心理支持者的历史帖中获得其角色心理画像, 该画像包含其策略偏好的刻画. 从历史帖中可以看出, 该支持者针对求助者的负面情绪, 其通常给予一定的共情用于安抚求助者, 例如情绪支持和鼓励之类的内容(下划线标注的内容), 然后再给予专业的建议或方法(加粗倾斜的内容). 针对某一类心理危机问题的策略偏好可以弥补语义匹配上的不足, 使得模型做出正确的判断.

求助帖: <i>...It seem that my mission to supress myself and stay at hidden at all cost be turn out to be hard than I expected. ... I'm not do an excursion, I want to avoid be a round school peer as m uch as possible...</i> (.....我压抑自己, 不惜一切代价隐藏自己的使命似乎比我想象的要艰难.....我没有参加旅行, 我想尽量避免和同学在一起.....)
候选支持贴: <i>...also especially that communication lead to understanding. If this be a requirement of your study and you discuss any aspect you be struggle with, teacher can often provide solution so you can still meet the same learning outcomes. Where be the excursion to anyway?</i> (.....尤其是沟通会带来理解. 如果这是你学习的要求, 并且你讨论了你遇到的任何困难, 老师通常可以提供解决方案, 这样你仍然可以达到相同的学习结果. 旅行要去哪里?)
求助者的历史帖: <i>c₁: ...recently they lecture me about unleash my true potential...I find this odd as they barely let me do anything with them. ...the late one I'm seen be not very helpful...</i> (.....最近他们教导我要释放我真正的潜力.....他们几乎不让我和他们做任何事情.上一个咨询师并没有让我觉得有帮助.....) <i>c₂: ...I don't want to go on any of them. ...Our school have also say student will have to or encourage to have some involvement with the formal how do I explain I don't want to do it?...</i> (.....我不想去任何一个.我们学校也说学生必须或鼓励参与一些有关毕业的正式活动, 我该如何解释我不想这么做?.....)

加粗倾斜的内容: 求助者的心理危机问题或支持者的专业建议或方法

下划线标注的内容: 心理危机问题发生的原因或背景、症状、情绪等

Fig. A1 The experimental cases analysis 1

图 A1 实验案例分析 1

求助帖: <i>...so today have my first psych apt, and it don't go so well. I'm in a ▲ from the start and find it incredibly hard to come out of that, ... I'm scar about all this. I see her next monday I'm wonder how others find their first apt with a psychcounsellor. Do you have trouble open up? if so how do you manage that?...</i> (.....今天我第一次有心理辅导, 但进展不太顺利. 我从一开始就感到▲, 然后发现很难从中走出来,我对这一切感到很恐惧, 我下周一还会见到她, 我想知道其他人第一次心理辅导是怎样的. 他们刚开始也有困难吗?如果有, 他们是怎么做到的?.....)
支持帖: <i>...I'm really proud of your for take that step to email your psych! ...it's so great to see that your psych be cool with your email her as well as come up with a middle ground which suit you both. ...</i> (.....我真为你感到骄傲, 你竟然给你的心理医生发邮件!很高兴看到你的心理医生在你给她发邮件的时候很冷静, 并且想出了这个适合你们双方的中间安全地带.....)
支持者的历史帖: <i>c₁: ...it can be a difficult situation when you'd like to distance yourself from someone you're not get along with, but also don't want to hurt their feelings. ...if you decide to do so. but try to remember that it might be really tough for if she see you distance yourself from her and her feeling may be hurt, so if you do decide to distance yourself from her, let her know in a compassionate way could help limit how upset it may make her. ...</i> (.....当你想要和无法相处的人保持距离但又不想伤害他们的感情时, 这是一个很困难的情况.....如果你决定这么做的话, 但要记住, 如果她看到你远离她, 她可能会觉得受伤. 所以一旦你决定和她保持距离, 以富有同情心的方式让她知道, 这可以帮助减少这件事让她难过的程度.....) <i>c₂: I'm really glad to hear you've already speak to your school counsellor. that's a great step to take! even though it sounds like you don't quite get what you be hop for out of it, it might still be worth have a few more chat with them, because they could still be able to help. In the meantime, ...there's heap of really helpful technique, info and tip that you might want to try for school and studying...</i> (.....我很高兴你已经和学校辅导员谈过了, 这是一个伟大的进步! 尽管听起来你还没有从中得到你想要的东西, 但是和他们多聊几句还是值得的, 因为他们还能帮上忙. 同时,这里有一堆非常有用的技巧、信息和提示, 你可能想在学校或学习中尝试一下.....)

▲: 用于替代示例中的极端负面词

加粗倾斜的内容: 求助者的心理危机问题或支持者的专业建议或方法

下划线标注的内容: 心理危机问题发生的原因或背景、症状、情绪等, 以及支持者的支持偏好等

Fig. A2 The experimental cases analysis 2

图 A2 实验案例分析 2