面向心理健康咨询的藏语数据集及大语言模型构建

朱孟笑[1]　　沙九[2]　　冯冲[3*]
1.北方工业大学信息学院
2.百度自然语言处理部
3.北京理工大学计算机学院
zhumx@ncut.edu.cn, shajiu_dn@163.com, fengchong@bit.edu.cn

摘要

焦虑、抑郁已成为人们常见的心理障碍，适度的疏导对于缓解人们精神、心理压力具有重要意义。然而由于病耻感等原因，很多人得不到及时的疏导和治疗。随着人工智能的发展，大语言模型（LLMs）优越的知识融会贯通能力和思维链能力，使得其成为心理疏导的有效工具。然而，现有少量面向心理健康咨询的大语言模型通常针对英文、中文等资源丰富的语种，而对于低资源语言，LLMs在心理咨询领域的应用尚缺少研究。本文以藏语作为低资源语言的代表，研究藏语心理咨询数据集的构建和藏语心理健康大语言模型的构建方法。首先，通过收集现有高质量的中文心理咨询对话数据，并对数据进行处理，生成心理健康多轮对话数据集；其次，构建汉藏翻译工具将其翻译成藏语多轮对话数据，并结合多种机制对数据进行筛选、过滤生成高质量藏语心理健康多轮对话数据；基于构造的数据，采用现有通用大语言模型Baichuan2和LLaMA2模型进行指令调优训练，形成藏语心理健康大语言模型，并将开源用于科学研究。最后通过实验验证了本文发布的藏语心理健康多轮对话数据集以及藏语心理健康咨询大语言模型的有效性。

关键词： 心理健康支持 ； 藏语 ； 大语言模型

Construction of Tibetan Datasets and Large Language Models for Psychological Health Counseling

Mengxiao Zhu[1]　　Shajiu[2]　　Chong Feng[3*]
1. School of Information Science and Technology, North China University of Technology
2. NLP, Baidu Inc.
3. School of Computer Science and Technology, Beijing Institute of Technology
zhumx@ncut.edu.cn, shajiu_dn@163.com, fengchong@bit.edu.cn

Abstract

Anxiety and depression have become prevalent psychological disorders, and moderate counselling plays a critical role in alleviating mental and psychological stress. However, due to reasons such as the sense of shame, many individuals do not receive timely counseling and treatment. With the advancement of artificial intelligence, large language models (LLMs) with their superior abilities in knowledge integration and cognitive chaining have become effective tools for psychological counseling. Nevertheless, existing psychological health LLMs are primarily focused on resource-rich languages like English and Chinese, with limited research on their application in low-resource languages. This paper focuses on Tibetan, a representative low-resource language, to explore the construction of Tibetan psychological counseling datasets and Tibetan psychological health LLMs. Initially, we collect high-quality Chinese psychological counseling dialogue data,

第二十三届中国计算语言学大会论文集，第326页–第339页，太原，中国，2024年7月25日至28日。
卷1：主会论文

process it, and create a multi-turn dialogue dataset for mental health; subsequently, we develop a Chinese-Tibetan translation tool to translate this into Tibetan, using multiple mechanisms to filter and produce high-quality Tibetan psychological health multi-turn dialogue data. Utilizing the constructed data, we fine-tune existing general LLMs, Baichuan2 and LLaMA2, to develop a Tibetan psychological health LLM, which will be open-sourced for scientific research. Finally, experiments validate the effectiveness of the released Tibetan psychological health multi-turn dialogue dataset and the Tibetan psychological health counseling LLM.

Keywords: Psychological health support , Tibetan , Large language model

1 引言

在当下，心理健康成为一个越来越重要的问题 (Liu et al., 2023a)。在快节奏和互联网时代，每个人都面临着很多影响心理健康的挑战。为了确保心理健康，及时的早期心理干预和诊断至关重要。然而，由于高昂的费用、复杂的语言程序和有限的资源，专业心理咨询师往往难以提供及时的援助。

大语言模型（Large Language Models，LLMs）的出现提供了一个可行的解决方案。随着ChatGPT、LLaMA等模型的发布，LLMs显示出了强大的意图理解、逻辑推理等能力，在很多任务上都取得了超出预期的效果。但现有大语言在面对心理咨询领域仍有一些局限性：（1）现有大语言模型是通用模型，缺乏心理健康领域的专业知识。心理健康咨询场景较为复杂，专业心理咨询师通常有丰富的领域知识，通用模型在面对心理咨询时，由于缺乏专业知识，往往出现避而不谈或答非所问的情况。（2）现有大语言模型在面对心理健康咨询时，通常急于解决用户的问题，给出冗长而笼统的建议，缺少追问能力。而专业的心理咨询师通常经过反复的提问来了解用户意图，给出专业、详实的回复。为了解决这些问题，少量研究构建了专门用于心理健康咨询领域的数据集，如SMILE数据集(Qiu et al., 2023)和AUGESC数据集(Zheng et al., 2022)等，还有个别工作采用GLM等模型作为基座模型，利用构建的数据集进行微调，实现心理健康领域的大语言模型 (Chen et al., 2023b)。

然而，现有心理健康咨询大语言模型通常针对英语、汉语等资源丰富的语种，依托丰富的开源数据可以构造高质量的心理健康咨询数据用于LLMs的指令微调。但这些LLMs难以很好地泛化到低资源语言中。对于文本数据充裕的语种，获取大规模数据样本相对容易。然而，在全球7000多种语言中，资源丰富的语言仅占少数，大部分语言的网络资源极为匮乏，许多低资源语言的网络数据占比不足0.1% (Liu, 2022)，这大大增加了收集这些语言数据的难度。此外，由于使用低资源语言的用户相对较少，对这些语言进行数据标注的工作也显得更为困难。标注数据的稀缺和语言间的显著差异意味着LLMs在低资源语言上的研究和应用仍处于起步阶段。因此，在低资源语言上，如何构造高质量的心理健康咨询数据、构建心理健康咨询大模型是一个亟待解决的问题。

本文以藏语作为低资源语言的代表进行研究。在我国，藏语是我国古老的语言文字之一，记载和传播了藏民族优良的文化和博大的精神，也在推动我国社会发展进步，民族团结一致等方面发挥了不可替代的重要作用。然而，藏语语言资源相对匮乏，在算法、算力和数据的平衡中，数据成为了制约藏语LLMs应用发展的关键短板。为推动心理健康大语言模型在藏语中的应用，需要构建藏族心理健康咨询数据集，进而构建藏语心理健康大模型。首先，为了解决数据稀缺的问题，选择中文高质量心理健康单轮对话数据集，并对数据进行自动化处理，生成中文高质量多轮对话数据集；然后，采用构建的汉藏翻译工具对数据集进行翻译，并采用多重筛选机制选择高质量的藏文心理健康多轮对话数据集；最后，基于现有开源的大语言模型，LLaMA-2和Baichuan-2，在构造的数据集上进行指令微调训练，构建藏语心理健康大模型[0]，通过分析大语言模型能力的提升程度验证数据集和模型的有效性。

本文主要贡献如下：

基金项目：北方工业大学科研启动基金项目资助
[0]https://github.com/Shajiu/LLM/tree/main/Tibetan_Mental_Health_Chat

（1）收集高质量中文心理健康单轮对话数据集，基于GPT 4和藏汉翻译工具构造高质量藏语心理健康多轮对话数据集；

（2）基于多重筛选机制选择高质量藏文心理健康多轮对话数据集，并进行开源。

（3）选择LLaMA-2和Baichuan 2大语言模型作为基座模型，采用构建的藏文心理健康多轮对话数据集对基座模型进行指令微调，构建面向心理健康领域的藏语大语言模型MindL-Ti和MindB-Ti，并进行开源。

（4）通过实验分析本文构造的藏文心理健康多轮对话数据集的有效性，以及通过与基座模型进行对比，验证了藏语心理健康大语言模型的有效性。

2 相关研究

焦虑、抑郁等是世界范围内常见的心理障碍，互联网无处不在的特性催生了心理健康支持的新范式。随着大语言模型（LLMs）在各种领域中的广泛应用，其在心理健康领域的应用也正在逐渐展现出巨大的潜力。近期的研究工作表明，LLMs可以被有效地用于提供情感支持对话、精神健康咨询，并具有良好的共情和倾听技巧。而在这其中，高质量的心理健康数据对心理健康咨询模型的效果有着至关重要的作用。下面将对心理健康数据集构建和LLMs在心理健康支持领域的应用进行介绍。

2.1 心理健康数据集构建

近年来，关于心理健康支持的研究在很大程度上取决于公开数据集的可用性 (Qiu et al., 2023)。大规模的心理健康会话数据集对于识别心理健康状况 (Liu et al., 2023c; Srivastava et al., 2022)、决定个性化干预措施 (Golden et al., 2024)等具有重要的意义。

现有心理咨询领域的数据集通常是针对英文、中文等语种。在英文数据集中， Liu et al. (2021)通过众包的方式构建了情感支持会话数据集ESconv，包含1053个情感支持对话。由于该方法的收集时间和成本较高， Zheng et al. (2022)通过大语言模型来扩充ESConv数据集，首先基于ESConv数据集对GPT-J 6B模型进行微调获得从初始对话完成完整对话的能力；然后，收集多样化的对话帖子作为初始对话，基于微调后的模型生成完整对话，据此构建了包含102K个对话的AUGESC数据集；Crisis Text Line (Althoff et al., 2016)提供了一个高质量的心理健康咨询数据集，它包含由经验丰富的志愿咨询师进行的大规模多回合咨询对话，但这个数据集无法公开访问。在中文数据集中，心理咨询领域收个开放的问答语料库 (Hai Liang Wang, 2020)，包括20000条心理咨询数据，该语料库的内容由现实世界的用户提出，回答由志愿者给出，大多数回家比较简短且笼统。 Sun et al. (2021)在公开心理健康支持平台上爬取了对话帖子，构建了高质量的心理健康对话数据集PsyQA。 Qiu et al. (2023)则是将PsyQA中的单轮对话改写为多轮对话，增加了对话的丰富性和多样性。 Chen et al. (2023b)通过众包的方式收集了对话数据集，并基于ChatGPT进行重写，形成了超过200万的心理咨询多轮对话数据集。

2.2 LLMs在心理健康领域的应用

以ChatGPT为代表的大语言模型的出现重新定义了自然语言处理的范式，LLMs不仅显著超越了传统意义的人机对话系统，而且被视作以自然语言为交互方式的通用语言处理平台，具有超出预期的交互体验，包括通用的意图理解能力、强大的连续对话能力、较强的逻辑推理能力等 (Zhao et al., 2023; Wan et al., 2023)。随着LLMs的迅速发展，其应用范围越来越广，在文本生成 (Liebrenz et al., 2023)、医疗 (Jeblick et al., 2023)、机器人 (Vemprala et al., 2023)、编码辅助 (Liu et al., 2023b)等领域都有广泛的应用。

随着LLMs应用范围的扩大，以LLMs技术支撑的聊天机器人在心理健康咨询领域受到越来越多的关注。 Wang et al. (2023a)探讨了LLMs在提供心理健康咨询方面的应用。他们发现LLMs能够表现出一定程度的理解和同理心，提供的回答有助于用户的心理健康支持。 Qin et al. (2023)利用LLMs开发了一种可解释的交互式抑郁检测系统，为检测心理健康指标引入了一种新的范式； Chen et al. (2023a)探索了ChatGPT在模拟精神科医生和患者对话的潜力，该研究肯定了在精神科中部署ChatGPT驱动的聊天机器人的可行性，并深入研究了提示设计对机器人行为和用户参与度的影响。 Ayers et al. (2023)实现了一个基于ChatGPT的聊天机器人，比较医生和聊天机器人在社交媒体中对患者的回应，发现有78.6%的评估者认为聊天机器人更快、更善解人意。 Chen et al. (2023b)基于ChatGLM-6B作为基座模型，对心理健康领

域的多轮对话数据集进行微调，构建支持心理健康咨询的模型SoulChat，使得模型的共情、倾听和安慰能力得到了显著提高。 Yang et al. (2024)基于大语言模型的多智能体系统，提出一种创新性的心理测量范式PsychoGAT，与传统问卷方式不同的是，该研究为每位参与者定制化生成一个可交互的叙事类型游戏，据此测量用户对应的心理特质。Mental-LLM (Xu et al., 2024) 分别基于Alpaca和FLAN-T5模型构建了针对心理健康领域微调的大语言模型Mental-Alpacha和Mental-FLAN-T5，并验证了模型的有效性。ChatCounselor (Liu et al., 2023a)是一个提供心理健康支持的大语言模型，与其他聊天机器人不同的是，其建立在真实对话基础上，使其拥有心理学领域的专业知识和咨询技能。该模型使用的训练数据集是由260次深度访谈构建而成，实验验证了高质量数据集对模型效果的改善。Psy-LLM (Lai et al., 2023)是基于盘古和闻仲大模型，采用心理健康咨询数据集对模型进行微调得到的心理健康支持的大语言模型。

这些最新的研究进展表明LLMs在心理健康领域的应用潜力正被逐步释放，这些方法为提供个性化的心理健康干预措施提供了新的途径。但这些模型都是基于丰富的开源数据集，因此现有关于心理健康支持的大语言模型通常是针对汉语、英语等资源丰富的语种，而对于低资源语种，由于开源心理健康领域数据集的匮乏，面向低资源语种的心理健康支持的大语言模型尚缺少研究。

3 面向心理健康支持的藏语数据集构建

3.1 数据来源

作为低资源语言，藏语数据极度稀缺，本文选择高质量的中文心理健康数据PsyQA (Sun et al., 2021)进行多重处理，生成藏语心理健康支持数据集。PsyQA数据收集自壹心理论坛，该论坛是一个心理健康服务平台，拥有着庞大的用户群体和专业的咨询师团队。在论坛中，匿名用户发布他们在日常生活中的烦恼和问题，由志愿者或专业咨询师给出文本回复。PsyQA包含了多种一般的心理健康障碍问题，设计了9个主题，包括自我成长、情感、爱情问题、人际关系、行为、家庭、治疗、婚姻和职业，其中，一个问题映射到多个答案。原始数据集统计如表1所示。

主题	问题数量	占比	答案数量
成长	4148	18.56%	10585
情绪	3037	13.59%	6804
恋爱	2956	13.23%	8312
人际	2923	13.08%	6911
行为	2490	11.14%	5404
家庭	2466	11.04%	6370
治疗	2304	10.31%	5479
婚姻	1234	5.52%	3962
职业	788	3.53%	2236
汇总	22346	100%	56063

Table 1: PsyQA原始数据

由于真实心理咨询场景中，咨询师往往通过多轮问询和解答来疏解用户的困惑，因此，单轮对话往往难以满足真实场景的需求，本文首先将其转换成多轮对话，然后将其翻译成藏语心理健康支持多轮对话数据，并对数据进行清洗等处理生成高质量的藏语心理健康支持多轮对话数据集。

3.2 多轮对话转换

由于原始数据集是论坛问答数据，为此在转换成多轮对话时，需要去掉一些在多轮对话中不会出现的内容，如“楼主”等 (Qiu et al., 2023)，此外，由于藏汉文化差异，需要删除独特于汉语的一些表达，如谚语等。本文基于规则的方法过滤掉相关内容，并人工进行检验与修正。

由于PsyQA中每个主题的数据量偏少，本文依据SELF-INSTRUCT方法 (Wang et al., 2023b)扩充单轮对话数据。将每个主题的单轮对话作为种子池，通过调用GPT-4生成单轮对

转换之前：

求助者：成绩不合格久久不能恢复，我怎么那么容易被挫折打倒？

支持者：楼主你好得知自己的成绩不合格，难受，虽然知道这个事情的主要原因是复习的少，但是还是忍不住的难受。楼主其实已经知道了原因，却还是很难受，那么楼主就可以去探究为什么难受？是因为自己没有看书么？是因为这次的题太简单？还是觉得自己没过会受到环境的影响？你需要把这里的关系理清楚，找到自己难受的根源，然后对症下药。楼主可以去听听音乐，做点运动，或者旅游放松自己的心情，其实楼主不必把这种事情放的太重，即便如此那又如何。你要问问自己，了解自己究竟想要什么，没有目标就不会用心对一件事情，有方向能坚持会事半功倍，加油。

转换之后：

求助者：成绩不合格，我久久不能恢复，我怎么那么容易被挫折打倒？

支持者：你好！面对成绩不合格确实很难受，大部分人都会有这样的感受。你已经意识到了复习不足可能是一个原因，这已经是一个很好的开始。你觉得，除了复习时间少，还有其他什么可能导致你这次考试没能达到预期吗？

求助者：可能是因为我没有很好地掌握学习方法，或者是我对这次考试的内容不够感兴趣。

支持者：这是很常见的情况。找到真正的原因可以帮助我们更好地解决问题。对于学习方法，你觉得有哪些方面是可以改进的？同时，对于兴趣的部分，有没有哪些科目或话题是你感觉更加投入的？

求助者：数学我通常比较感兴趣，但这次的章节我觉得特别难，所以就没那么积极。

支持者：那这就是一个很好的线索了。对于那些难的部分，我们可以尝试找一些不同的资源或方法来帮助你理解。同时，保持对自己感兴趣的部分的热情也很重要。你有尝试过和同学一起学习，或者向老师求助吗？

求助者：我试过几次，但感觉效果不是很好。

支持者：那也许我们可以尝试一些其他策略，比如寻找一些在线课程或视频，有时候不同的解释方式会有新的启发。此外，让自己在学习时保持放松也很重要，你有尝试过通过运动、听音乐或其他活动来放松吗？

求助者：有时候我会听音乐来放松。

支持者：那很好，保持心情的平衡对于学习效率是有帮助的。试着在学习和休息之间找到一个合适的平衡，不要太过压迫自己。你已经展现了很多自我反思的能力，我相信你能找到适合自己的学习方式，并逐渐改善成绩的。加油！

Figure 1: 多轮对话转换示例

话加入到原来的数据集中。然后，依据 Qiu et al. (2023)的方法，通过调用GPT-4将单轮对话扩展到多轮对话，采用的提示模版如表2所示。

我想让你作为一个文本重写者。您应该遵循这些要求：1. 每个句子必须以求助者：或支持者为开始：；2.对话必须以求助者：为开始；3. 支持者的反应有适量的情感支持和调节；4. 每个说话人的话语长度应符合对话场景，不宜过长。 你需要将给定的单回合对话改写为心理健康支持对话，对话内容为求助者与支持者之间的10次或更多对话。 单论对话为：XXX 让我们一步一步地分析和重写它。 你重写的多回合对话是:”’

Table 2: 生成心理健康多轮对话的模版提示

图1通过一个例子展示了转换之前的单轮对话和转换之后的多轮对话，通过例子可以看出，以这种方式可以有效地将单轮对话转换成多轮对话。

3.3 多轮对话翻译

由于目前面向藏语的翻译工具比较稀缺，现有的翻译工具缺乏专门针对心理健康领域的翻译能力，尤其是在处理心理健康领域术语方面。为此，本文训练一个专门针对心理健康领域的汉藏翻译模型来对心理健康领域的中文数据进行翻译。然而，由于心理健康领域的汉藏平行语料非常稀缺，难以直接收集平行语料构建汉藏翻译模型。相对来说，获取心理健康领域的藏语单语语料比较容易，为此，本文采用反向翻译 (Sennrich et al., 2016)的方法来构建高质量的心理健康领域的汉藏翻译工具来获得高质量的藏文心理健康多轮对话数据集。

为了构建翻译模型，本文基于实验室收集的汉藏平行语料和CCMT 2020的藏汉翻译数据集，基于OpenNMT的Transformer模型训练翻译模型，构成一个基础的通用汉藏翻译模型。然后采用该模型，对收集到的心理健康领域的藏文单语数据集进行翻译，将其翻译成中文，称为“伪数据集”，该“伪数据集”基于翻译模型翻译得到，其中的中文会存在噪声，以此作为输入对翻译模型进行训练，可以提高翻译模型的泛化性。在得到“伪数据集”后，将其与原始的汉藏平行语料混合，形成更大规模的训练数据，用于训练最终的汉藏翻译模型。通过实验测试发现，通用汉藏翻译模型在心理健康咨询领域的BLUE值为41.86，而添加“伪数据集”后训练的汉藏翻译模型在心理健康咨询领域中多次测试的平均为43.39.

利用训练好的汉藏翻译模型对处理后的中文心理健康数据集进行翻译，获得藏文心理健康数据，并通过人工校正提升数据质量。

3.4 数据筛选

由于本文基于翻译的方式来获得藏文心理健康多轮对话数据集，翻译后的数据存在低质量的内容。本文结合多种机制对数据进行优化和筛选。

首先，基于N-gram语言模型对文本进行评估，过滤掉低质量的内容。本文将收集到的藏文文本数据分成N-gram序列，其中$N=3$，通过BERT-BASE训练藏语的N-gram语言模型。基于该模型对翻译后的藏文心理健康多轮对话中的每个语句进行N-gram序列的评分。针对每个多轮对话，其每个语句中所有N-gram序列评分的均值若低于某个阈值，则将其视为低质量内容。在本文中，通过实验分析发现阈值过大时，容易过滤掉大量正常文本，而阈值过小时，低质量内容无法被过滤掉，平衡过滤的效果，将阈值设置为0.85.

其次，基于指令跟随难度（Instruction-Following Difficulty, IFD）(Li et al., 2023)对多轮对话数据进行筛选，提高模型指令调优的有效性。而由于指令跟随难度需要根据大语言模型的回答进行评估，而现有通用大语言模型在藏语心理健康咨询领域缺乏指令跟随能力，为此，本文采用少量高质量的数据对模型进行指令调优训练，进而利用调优后的模型来评估数据的指令跟随难度。由于本文收集的数据包含9个主题，为了保证质量和多样性，从每个主题中人工挑选等量的数据，构成1000条高质量数据集。在获得数据集后，利用该数据集对LLaMA-2模型进行指令调优训练，获得预经验模型。接着利用该模型对翻译后的心理健康多轮对话数据进行预测，通过指令内容预测答案，并获取预测答案与真实答案之间的差异，采用条件回答分数（Conditioned Answer Score, CAS）来衡量该差异，CAS为调优过程中的平均交叉熵损失。在调优过程中，给定问答对(Q,A)，其损失计算为：

$$L_\theta(A|Q)=\frac{1}{N}\sum_{i=1}^{N}\log P(w_i^A|Q,w_1^A,...,w_{i-1}^A;\theta) \tag{1}$$

其中，θ为模型的参数，N为回答A中单词的数量，w_i^A表示回答A序列中的第i个词，$P(w_i^A|Q,w_1^A,...,w_{i-1}^A;\theta)$是在问题$Q$和回答序列$w_1^A,...,w_{i-1}^A$的条件下得到词$w_i^A$的概率。CAS值的大小反应了模型基于指令生成答案的难易程度，但其可能受到模型生成答案A难易程度的影响。为此，采用直接答案分数（Direct Answer Score, DAS）来衡量模型生成答案的难易程度：

$$s_\theta(A)=\frac{1}{N}\sum_{i=1}^{N}\log P(w_i^A|w_1^A,...,w_{i-1}^A;\theta) \tag{2}$$

其中，$P(w_i^A|w_1^A,...,w_{i-1}^A;\theta)$是在回答序列$w_1^A,...,w_{i-1}^A$的条件下得到词$w_i^A$的概率。DAS得分越高，表示模型生成该答案任务本身更为复杂。为了获取更好的指令数据，需要去除答案本身的影响，因此采用IFD分数来评价指令数据对模型调优的影响，计算如下：

$$IFD=\frac{CAS}{DAS} \tag{3}$$

较高的IFD分数表示模型难以将答案与给定的指令内容进行对齐，表示指令的难度更高，对模型调优更为有利。通过计算每个问答对的IFD指标，选择分数更高的数据作为藏语心理健康多轮对话数据集，最终数据集如表3所示。

主题	规模（条数）	占比
成长	9296	18.592%
情绪	8074	16.148%
恋爱	6912	13.824%
人际	6846	13.69%
行为	4980	9.96%
家庭	4932	9.86%
治疗	4608	9.22%
婚姻	2468	4.94%
职业	1884	3.77%
汇总	50000	100%

Table 3: 藏语心理健康多轮对话数据集分布表

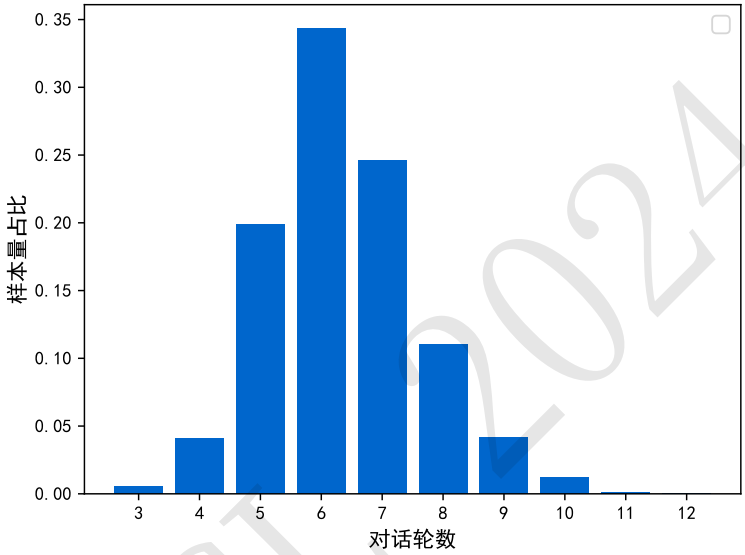

Figure 2: 藏语心理健康多轮对话轮数分布

基于以上方式最终形成了50000个藏语心理健康多轮对话数据，每个多轮对话的轮数在3-12轮之间，轮数的分布如图2。图 3是其中一个多轮对话示例。对最终的藏文心理健康多轮对话数据进行统计，结果如表4所示。其中，平均长度是去除“求助者”和“支持者”之后的文本按照音节计算出来的长度，这里的音节通过藏语文本中的”་”进行分割得到。

对话数目	平均轮数	平均长度	平均长度（求助者）	平均长度（支持者）
50000	6.34926	57.493	41.298	73.688

Table 4: 数据统计信息

4 藏文心理健康咨询大语言模型构建

4.1 基座模型选择

本文选择两个通用大语言模型LLaMA-2和Baichuan-2作为基座模型，其中，LLaMA-2（Large Language Model Meta AI） (Touvron et al., 2023)是由Meta AI发布的一个开放且高效的大型基础语言模型，作为完全开源的模型，被广泛用作AI模型的通用基座；Baichuan-2 (Yang et al., 2023)是国内百川智能发布的一个开源大语言模型，多次在国内外测评榜单中脱

Q: ང་རང་ཉེ་ལམ་མགོ་འཁོར་ཡོད།མཐོ་གསུམ་རྗེས་ཀྱི་ང་རང་མཐོ་རྒྱུགས་སྔོན་གྱི་འཇིགས་སྐྲག་ལས་གཞན་ད་དུང་རང་ཉིད་སྐྱེས་མ་ཐག་ནས་བཟུང་ལྷག་མ་ཞིག་ཡིན་པར་འདོད།དོན་སྙིང་ཅི་ཡང་མེད།ང་ལ་གྲོགས་པོ་མེད།ཁེར་རྐྱང་གིས་བཟོད་མི་ཐུབ།ཁྱེད་ཀྱིས་ང་ལ་བསམ་འཆར་འགའ་འདོན་ཐུབ་བམ། (我最近有点迷茫，高三之后的我除了高考前的恐惧，还觉得自己生来就是多余的，没有什么意义，我没有朋友，孤独难耐，你能给我一些建议吗？)
A: ཁྱོད་ཀྱི་སྐད་ཆ་མཐོང་རྗེས།ངས་ཁྱོད་ལ་སེམས་ཁྲལ་ཆེན་པོ་སྐྱེས་བྱུང་།འདི་ནི་རྒྱུན་མཐོང་གི་མཐོ་འབྲིང་འཚོ་བའི་རྣམ་པ་རེད།ཐོག་མར་བཤད་འདོད་པ་ནི།ངས་ཁྱོད་ཀྱི་ཆོད་སེམས་མཐོང་བྱུང་།འདི་ནི་མི་མང་པོ་ཞིག་ལ་མེད་པ་རེད།མཐོ་རྒྱུགས་ཀྱི་ཆེས་གལ་ཆེ་བ་ནི་སེམས་ཁམས་རེད།ཁྱོད་ལ་རྒྱལ་ཁ་ཐོབ་པའི་སེམས་ཁམས་དགོས་པ་རེད། (看到你的话，我很担心你，这是常见的高中生活状态，首先想说的是，我看到了你的决心，这是很多人都没有的，高考最重要的是心态，你要有成功的心态。)
Q: ཁྱོད་ཀྱིས་སྐུལ་མ་བྱས་པར་བཀའ་དྲིན་ཆེ།ངས་དངོས་གནས་ལྷོད་གཡེང་བྱེད་འདོད།འོན་ཀྱང་རང་ཉིད་ཀྱི་དམིགས་འབེན་དང་སྐུལ་ཤུགས་ཤོར་མི་རུང་བར་འདོད། (谢谢你的鼓励，我真的很想放松，但我觉得自己不能失去目标和动力。)
A: གལ་ཏེ་ཁྱོད་ཀྱིས་འཆར་གཞི་ཅི་ལྟར་བཟོ་དགོས་པ་མི་ཤེས་ན།ངས་བསམ་འཆར་འགའ་འདོན་ཐུབ། (如果你不知道如何制定计划，我可以提些建议。)
Q: འོ་ན་ངས་འཆར་གཞི་ཅི་ལྟར་བཟོ་དགོས་སམ། (那我应该怎么制定计划呢？)
A: མཐོ་གཅིག་ནས་འགོ་བརྩམས་ཏེ་གོམ་པ་གང་རེ་གང་རེ་སྤོས་ན།འགོ་རྩོམ་ས་དམའ་ཞིང་།འདྲི་གཞིའི་འཐབ་ཇུས་ལ་ཚོད་ལྟ་བྱེད་དགོས།སྔོན་ལ་རྨང་གཞིའི་གཤར་སྦྱང་འགའ་བྱེད་པ་དང་།རིམ་བཞིན་དཀའ་ཚད་ཇེ་ཆེར་བཏང་ན།ཁྱོད་ལ་རིམ་བཞིན་གྲུབ་འབྲས་ཐོབ་པའི་ཚོར་བ་ཡོད་སྲིད།དེ་དང་མཉམ་དུ་རིགས་སོ་སོའི་ཁ་བྱང་ཚང་མ་བསྒྲུབ་དགོས།འདི་ལྟར་བྱས་ན་ཁྱོད་ཀྱིས་མྱུར་ལས་མགྱོགས་པའི་སྒོ་ནས་འདྲི་གཞི་ལ་གོ་བ་ལེན་པ་དང་དྲིས་ལན་འདེབས་ཐུབ། (从高一开始一步一步得走，从低起点开始，对于题型战术，先做一些基础练习，逐步加大难度，你会慢慢有成就感，同时，每一类题目都要做。 这样你能更快地理解和回答问题。)
Q: གལ་ཏེ་ངས་ཁ་བྱང་ཁ་ཤས་ལ་གོ་བ་ལེན་མི་ཐུབ་ན་ཅི་ཞིག་བྱ། (如果我不理解一些题目该怎么办。)
A: ངོ་ཚ་མི་དགོས།དགེ་རྒན་ལ་འདྲི་རུ་སོང་།སྤོབས་པ་ཆེན་པོས་ཁོ་ཚོར་ཁྱོད་ཀྱིས་གནད་དོན་འདིར་གོ་བ་ལེན་མི་ཐུབ་ཅེས་ཤོད།དགེ་རྒན་གྱིས་ཁྱོད་ལ་རོགས་བྱས་ནས་རྒྱུས་ལོན་བྱེད་ངེས།དེ་མ་ཟད་ཁྱོད་ཀྱི་དྲི་བར་ལན་འདེབས་སྲིད།འདི་ནི་དགེ་རྒན་གྱིས་སྒྲུབ་འོས་པ་ཞིག་རེད།མཐར་ཕྱིན་རྗེས་བཤད་མི་དགོས།མྱུར་དུ་སོང་ནས་འདྲི། (不要害羞，去问老师，大胆地告诉他们你不理解这个问题，老师会帮助你理解，并回答你的问题，这是老师应该做的，毕业了就不要说了，快去问吧。)
Q: ངས་བལྟས་ན་རང་ཉིད་ཀྱི་དབྱིན་ཡིག་དང་དངོས་ལུགས་ཀྱི་རྒྱུགས་བླངས་པ་མི་བཟང་།གནས་ཚུལ་འདི་རི་ལྟར་ཐག་གཅོད་བྱེད་དགོས་སམ། (我觉得自己英语考得不好，怎么解决这种情况呢？)
A: ཁྱོད་ཀྱིས་དབྱིན་ཡིག་མང་པོ་ཞིག་བློར་བཟུང་ན་ཆོག་པ་རེད། (你可以多背一些英语。)
Q: ཁྱོད་ཀྱིས་ང་ལ་རྒྱབ་སྐྱོར་དང་རྒྱབ་སྐྱོར་བྱས་པར་བཀའ་དྲིན་ཆེ།ཁྱོད་ཀྱི་བསམ་འཆར་དང་རྒྱབ་སྐྱོར་ཡོད་ན།ངས་ཕྱར་བས་སློབ་སྦྱོང་ལ་འབད་པ་དང་ཡིད་ཆེས་བརྟན་པོ་བྱས་ཏེ།མཐོ་རྒྱུགས་ལེགས་འགྲུབ་ཡོང་བར་བརྩོན་ལེན་བྱེད་སྲིད། (谢谢你对我的支持和支持，有你的建议和支持，我会更加努力学习，坚定信心，争取高考成功。)
A: གནད་དོན་མེད།ངས་ནམ་ཡང་ཁྱོད་ལ་རྒྱབ་སྐྱོར་བྱེད།རང་ཉིད་ལ་ཡིད་ཆེས་ཐོགས།རང་ཉིད་ཀྱི་མི་ཚེའི་དམིགས་འབེན་མངོན་འགྱུར་བྱེད་ཐུབ་ངེས་ཡིན།འབད་པ་བྱོས། (没问题，我永远支持你，相信自己，一定能实现自己的人生目标，加油！)

Figure 3: 藏语心理健康多轮对话示例

颖而出，夺得榜首位置。虽然两个模型的词汇表中都缺少藏语词汇，但都采用了回退到字节的方式来支持多种语言，故而可以进行藏语数据的指令调优。

4.2 指令微调方案

基于构造的藏文心理健康多轮对话数据集，采用基座模型进行指令微调可以得到藏文心理健康咨询大语言模型。由于构造的数据集是一个多轮对话数据集，在进行指令微调时，需要将数据进行特殊处理后输入到大语言模型中进行训练。

假定一条n轮藏文心理健康多轮对话数据，提问为Q，回答为A。在指令微调时，一般只有回答A部分的经验损失会用于梯度回传，更新模型参数，而提问Q部分的经验损失则不会用于更新参数。目前对于多轮对话数据的训练方式主要有：

（1）A中最后一次回答的经验损失参与更新模型。例如现有三轮对话数据$Q_1, A_1, Q_2, A_2, Q_3, A_3$，采用该方式，则将$Q_1, A_1, Q_2, A_2, Q_3$作为模型的输入部分，$A_3$作为模型的预测部分，即只有$A_3$部分的经验损失参与模型参数的更新。该方法没有充分利用多轮对话的训练数据，A_1和A_2部分都没有参与模型训练。并且对于心理健康咨询领域，中间的回复往往具有更丰富的信息，最后的回复可能是比较简短的内容，若只使用最后的回复训练模型，会影响模型的训练效果。

（2）将多轮对话拆分成多条数据进行训练。对于三轮对话数据$Q_1, A_1, Q_2, A_2, Q_3, A_3$，采用该方式将数据拆分成三条数据；$\{Q_1, A_1\}$, $\{Q_1, A_1, Q_2, A_2\}$, $\{Q_1, A_1, Q_2, A_2, Q_3, A_3\}$，对每条数据，仍然以$A$的最后一部分作为模型的预测部分，其他部分作为模型的输入部分。相比第一种方法，这种方法能够更加充分地利用对话中每个回复的内容，但需要将n轮对话拆分成n条数据，降低了训练的效率。

本文采用一种更加高效的方法 (Yang, 2023)，方案如图 4所示。对于三轮对话数

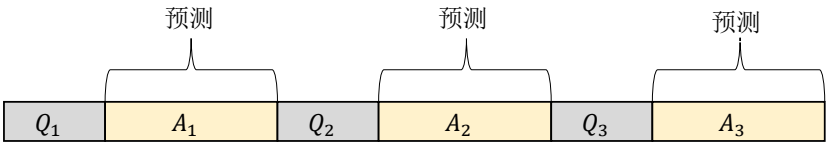

Figure 4: 藏文心理健康多轮对话数据训练方法示意图

Token_id

$< s >$	Q_1	$</s >$	A_1	$</s >$	Q_2	$</s >$	A_2	$</s >$	Q_3	$</s >$	A_3	$</s >$

A_mask

0	0	0	1	1	0	0	1	1	0	0	1	1

Figure 5: 掩码向量示意图

据$Q_1, A_1, Q_2, A_2, Q_3, A_3$，将该多轮对话数据进行拼接后输入模型，并计算每个位置的经验损失，只有A_1, A_2, A_3部分的经验损失参与模型参数的更新。由于以GPT为代表的因果语言模型都采用了一种注意力掩码机制，即对角掩码矩阵，在该机制中，每个token在编码时，只能看到它之前的token信息，而看不到它之后的token信息，因此在该三轮对话中，Q_1部分的编码输出只考虑了Q_1的内容，可以用来预测A_1的内容；而Q_2部分的编码输出是考虑了Q_1, A_1, Q_2的内容，可以用来预测A_2的内容；Q_3部分的编码输出是考虑了Q_1, A_1, Q_2, A_2, Q_3的内容，可以用来预测A_3的内容。对于整个序列，只需要输入一次就可以获得每个位置的经验损失，从而更新模型参数。

为了采用这种方法进行训练，需要选择属于因果语言模型的大语言模型。因此本文选择的是LLaMA2-7B和Baichuan2-7B作为基线模型进行训练。在训练时，为每个回复后面添加结束标记符，如$< /s >$作为此轮对话生成结束的标识符。因此，将n轮的多轮对话拼接成如下格式作为模型的输入：

$$< s > Q_1 < /s > A_1 < /s > Q_2 < /s > A_2 < /s > ... \tag{4}$$

此外，为了标记文本中的token属于提问还是回复，生成一个掩码向量A_mask，取值为0或1，用来标记文本中的token是否属于回复A部分，即是否需要模型进行预测，如图 5，其中$A < /s >$部分的掩码全为1，其余部分均为0.

根据这种方式，并行计算每个位置的经验损失，只有掩码A_mask=1的部分位置的损失函数才会参与模型参数更新。该方法充分利用了模型并行计算的优势，并且多轮对话中每个回复A部分都参与了训练，提高了数据的利用率。

本文采用Firefly (Yang, 2023)作为大语言模型指令调优训练的工具，选择QLORA (Dettmers et al., 2024)作为指令微调技术，使用本文的指令微调数据集分别对基座模型LLaMA2-7B和Baichuan2-7B进行指令微调，得到藏文心理健康咨询大语言模型MindL-Ti和MindB-Ti。

5 实验

5.1 实验实现

对基座模型LLaMA2-7B和Baichuan2-7B模型进行指令调优训练的大部分相关参数设置遵从Firefly的默认设置。由于采用QLoRA进行训练，QLoRA矩阵的秩$lora_rank$设置为64，QLoRA中的缩放参数$lora_alpha$设置为16，训练时的最大长度max_length设置为1024，学习率$learning_rate$设置为2e-4，训练步数$trainingstep$设置为1500。本实验的模型都是在拥有4块32G NVIDIA Tesla V100的服务器上进行调优训练与测试。

为了验证构造的藏文心理健康多轮对话数据集及藏文心理健康咨询大语言模型的有效性，采用定量与定性分析结合的方法进行。为了方便评估，本文开发了一个在线藏语心理健康咨询平台，该平台让专业咨询师能够向每位客户提供免费的文本形式咨询服务，平台如图6所示。

Figure 6: 在线藏语心理健康咨询平台

定量分析采用BLEU-4、ROUGE-1、GOUGE-2、GOUGE-L作为评价指标，来测试本文构造的数据集对于大语言模型能力的提升效果，以及构造的大语言模型MindL-Ti和MindB-Ti相比基座模型在心理健康咨询领域效果的提升。本文采用5折交叉验证的方式进行实验，将本文构造的藏文心理咨询多轮对话数据集随机分成5份，各占比20%，采用其中四份作为训练集进行训练，另外一份作为测试集。取5次测试结果的平均值作为最终结果进行展示。

5.2 定量分析

模型	ROUGE-1	ROUGE-2	ROUGE-L	BLUE-4
Baichuan2-7B-base	55.07	28.55	21.64	2.94
Baichuan2-7B-chat	48.57	28.34	19.78	8.14
Llama2-7B-base	40.55	23.99	27.76	10.96
Llama2-7B-chat	25.99	15.63	7.38	5.76
MindB-Ti	83.14	41.54	44.58	28.27
MindL-Ti	65.04	34.14	40.42	27.36

Table 5: 各模型在测试集上的实验结果

对于定量分析，利用本文构造的心理健康多轮对话数据集进行指令调优训练的模型MindB-Ti和MindL-Ti与对应的基座模型Baichuan2-7B-base、Baichuan2-7B-chat、LLaMA2-7B-base和LLaMA2-7B-chat在测试集上的实验结果如表 5所示。根据表中的结果可以看出，本文构建的藏文心理健康多轮对话数据集对于提升大语言模型的能力具有显著的效果，通过在两个基座模型上进行调优训练，得到的模型MindB-Ti和MindL-Ti相比基座模型在各指标上都有较大的提升，在（ROUGE-1、ROUGE-2、GOUGE-L、BLUE-4）上相比base模型提升的平均值分别为（28.07，12.99，22.94, 25.33）、（24.49, 10.15, 12.66, 16.4），相比chat模型提升的平均值分别为（34.57，13.2，24.84, 20.13）、（39.05, 18.51, 33.04, 21.6）.

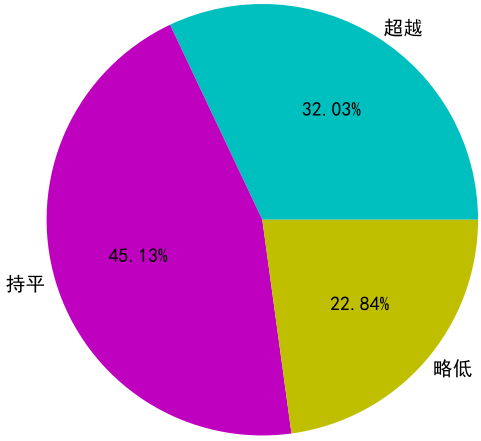

Figure 7: 人工评估结果

Q: ངའི་ཕྱིས་པ་སློབ་གྲྭར་འགྲོ་བའི་སྐབས་གུ་རྒྱན་དུ་དཔལ་འགོར་ཁེད་ལ་དང་།ཁོ་སློབ་གྲྭར་སྐྱེལ་ཞིང་ཤ་ཚོད་ཟེལ་འཚུབ་སྐྱེལ་གུ་བཏུག་དས་ཀྱང་ཁོ་ལ་སྐད་ཆ་ཙམ་འདོད་ཐིབ།་འདི་ཅི་ཤར་ཁེད་ལ་ཐག་ཆོད་ཡོག་ཡིན།་འདི་ཅི་ཤར་ཁེད་དགོས་སམ། (我的孩子上学时总是磨蹭，每次送他上学都让我很着急，我也会对他大喊大叫，这应该怎么办呢？)

A: དས་ཁྱོད་ཀྱི་ཚོར་བ་ཤེས།ཁྱོད་ཀྱིས་བཤད་པ་ཤར་ན།ཁྱོད་ཀྱི་ཕྱིས་པར་རང་ཉིད་ཀྱི་ཚོར་བ་སྐྱེན་པོ་དང་རྒྱུན་ལྡན་རེད།འབའ་ཞིག་ཏུ་ཁྱོད་ཀྱི་ཚོར་བ་ཙམ་མིན་པའི་ཁེད་ཀྱི་ཐབས་ལམ་ཡང་སྟོང་ཁྱེད་དགོས་པ་རེད། （我知道你的感受，据你说，你的孩子有自己的敏感和正常，这是很多家长的想法，你不能仅表达自己的感受，你要注意自己的情绪。）

Q: ཨེན།་ངས་བསམ་ན་ཁྱོས་པ་དང་ཁྱེམ་བདག་ཀ་ཤས་ཀྱི་ཐེམས་ཁམས་རྒྱུ་རྐྱེན་ཁྱོས་གནད་དོན་འདི་དག་བཟོས་པ་རེད།ཏ་ལའི་དག་ཡིན་ན་ངས་ཅི་ཤར་ཁང་ཁེད་ཀྱི་ཐེམས་ཁམས་ལ་ཚོད་འཛིན་ཁྱེད་དགོས། (嗯，我想是有一些孩子和家长的心理因素导致了这些问题。那我该如何控制自己的情绪呢？)

A: ཐོག་མར།ཁྱོད་ཀྱིས་རང་གི་ཐེམས་ཁམས་ལ་རྒྱས་པོན་ཁྱེད་པ་དང་ཉེ་ནས་ཐབས་བསོར།འཚེན་ནས་ཐག་གཅོད་ཁྱེད་པར་ཚོད་རྟེན་ཅིག་ཁྱོད་ཀྱིས་དཔྱག་རྩལ་ཁྱེད་ཚོགས་ཤ་ཟན་ཡིན་ན་རང་ཉིད་ཀྱི་ཐེམས་ཁམས་བརྩོལ་ཏོག་ཙམ་ལ་ཉེན་ཁྱེད་ཀྱི་ཐབས་ཤེས་ཡོད་པ་ཨིན་ན།ཁྱོད་ཀྱིས་ཐེམས་ཁམས་ལ་ཚོད་འཛིན་ཁྱེད་མི་ཐུབ་པར་འདོད་ན།ཁེད་ཀྱིས་ཐེམས་ཁམས་ལ་ཚོད་འཛིན་ཁྱེད་ཚོགས། （首先，你可以试着了解你的情绪，然后想办法解决，比如，你可以深呼吸、听音乐或者表达自己的情绪，如果你觉得情绪失控，可以找一些专业人士咨询。）

Q: བོད་ཀྱང་ངས་ཁྱོས་པར་ང་གཞའ་ནན་མེན་པར་འདོད་ནས་ང་ལ་བརྡེ་བཀུར་མི་ཁྱེད་པར་སེམས་ཁྱོད། (但是我也担心孩子会觉得我不够严格，不尊重我。)

A: ངས་ཁྱོད་ཀྱི་སེམས་ཁྱོ་ཤེས།ཁོས་ཀྱང་ཁྱོས་ཀྱིས་ཁྱོས་པ་ཚོར་ཁྱོད་ཀྱི་སེམས་ཁྱོ་བཤད་ཚོགས།ཁྱོས་ཀྱི་སེམས་ཁམས་བརྩེ་ཚོགས།ཁྱོས་པ་འདི་དག་ལ་སེམས་ཁམས་བརྩེ་ཚོགས།ཁྱོས་པ་ཚོར་ཁྱོད་ཀྱི་སེམས་ཁམས་བརྩེ་ན།ཁོ་ཚོས་ཁྱོད་ཀྱི་སེམས་ཁམས་ལ་རྒྱུས་པོན་ཁྱེད་ཤུལ་པ་མ་རེད།ཁྱོད་ཀྱི་སེམས་ཁམས་ལའང་རོགས་རམ་དང་རྒྱབ་སྐྱོར་ཁྱེད་ཤུལ། （我知道你的关注点，但是你可以告诉孩子们你的意思，你可以告诉孩子们你的心情，你在关心这些孩子，告诉他们你的心情，他们不仅能理解你的情绪，也能帮助和支持你的情绪。）

Figure 8: 采用MindL-Ti模型的案例

5.3 定性分析

对于定性分析，为了评估本文提出的藏语心理健康咨询大语言模型的性能，本文采用专业人员评估的方式进行。首先，从测试集中随机抽取了100个多轮会话，每个会话包含至少9轮对话。接着，采用基座模型和本文构造的藏语心理健康咨询大语言模型中生成了相应的回应。然后，将历史对话、由基线模型生成的回应、本文构造的模型生成的回复和真实回复一起提供给专业心理咨询师，由他们根据回应的自然度、类人特性、正确性等方面选择出历史对话的最佳回应。本文采用多数投票原则，由5位具有心理健康咨询背景的人员给出最终决策，结果如图 7所示。根据结果显示，采用本文构造的心理健康咨询大语言模型展现出了良好的性能。在所有回复中，约有32.03%回复在某种程度上超出了真实回应；达到真实回应相当水平的回应约占45.13%，而约有22.84%的回复略低于真实回复的水平。

此外，为了直观展示本文提出的藏语心理健康咨询大语言模型的有效性，选取了一个心理健康咨询的案例进行展示，使用相同的问题作为模型的输入，分析不同模型对问题的回答是否合理。由于未使用本文构造的藏语心理健康多轮对话数据微调的模型往往给出冗长的回复，本文仅展示LLama2-7B-chat模型针对不同问题的回复情况，如图9所示。对于微调后的模型，采用与之对应的MindL-Ti模型的结果进行展示，结果如图8所示。根据图中的结果可以看出，未使用本文构建的藏语心理健康多轮对话数据进行微调的模型在回答藏语心理健康问题时，往往给出无意义的回复。例如，在回答第一个问题时，Llama2-7B-chat模型对问题进行了复述，并重复了很多无意义的短句；在回答第二个和第三个问题时，直接用英语分析了藏语句子，无法直接对藏语问题进行直接回答。而使用本文构建的藏语心理健康多轮对话数据进行微调的模

Q:ངའི་བྱིས་པ་སློབ་གྲྭར་འགྲོ་བའི་སྐབས་སུ་རྒྱུན་དུ་དལ་འགོར་བྱེད་པ་དང་། ཁོ་སློབ་གྲྭར་སྐྱེལ་ཐེངས་རེ་རེར་ང་ཧ་ཅང་བྲེལ་འཚུབ་སྐྱེས་སུ་བཅུག །ངས་ཀྱང་ཁོ་ལ་སྐད་ཆེན་པོས་འབོད་སྲིད། འདི་ཅི་ལྟར་བྱེད་དགོས་སམ།

A:ངའི་བྱིས་པ་སློབ་གྲྭར་འགྲོ་བའི་སྐབས་སུ་རྒྱུན་དུ་དལ་འགོར་བྱེད་པ་དང་། ཁོ་སློབ་གྲྭར་སྐྱེལ་ཐེངས་རེ་རེར་ང་ཧ་ཅང་བྲེལ་འཚུབ་སྐྱེས་སུ་བཅུག །ངས་ཀྱང་ཁོ་ལ་སྐད་ཆེན་པོས་འབོད་སྲིད། འདི་ཅི་ལྟར་བྱེད་དགོས་སམ།ཡོངས་བྱེད་དགོས་སམ།ཡོངས་བྱེད་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དེ་དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེད་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི་ལོ་བྱེན་དགོས་སམ།དོན་མི（我的孩子上学总是磨蹭，每次送他上学都让我很紧张，我也会对他大喊大叫，这该怎么办呢？该办吗？该办吗？不该办？不该办吧？ 到底该不该当？该不该当？该当吗？该当事吗？该当事吗？该当吗？该当么？不当回事吗。该当么？不当吧？ 到底该不该当？该不该当？该当当？该打人还是该打人？该不该打事？该不该打事？该做事？不该做事？不该当事吗？该当事吗？应该做人吗？ 到底该不该当？该不该当？该当吗？该当事吗？该当事吗？该当吗？该当么？不当回事吗。该当么？不当吧？ 到底该不该做，到底该不该做，到底是不是该做。要不要？事要做？事要做？人要做事吗？事需要做吗？事需要做吗？不需要做什么吗？不需要做什么吗？ 到底该不该做，事不该做，事不该做。）

Q:ཨེན། ངས་བསམས་ན་བྱིས་པ་དང་ཁྱིམ་བདག་ཁ་ཤས་ཀྱི་སེམས་ཁམས་རྒྱ་རྒྱུན་གྱིས་གཞན་དོན་འདི་དག་བཟོས་པ་རེད་དེ་ན་ངས་ཅི་ལྟར་རང་ཉིད་ཀྱི་སེམས་ཁམས་ལ་ཚོད་འཛིན་བྱེད་དགོས།

A:This is a Tibetan text that contains a list of 108 different types ofatial beings or "lower realms" that are said to exist in Buddhist teachings. These beings are believed to inhabit various planes of existence, including hells, hungry ghosts, animals, and humans, and are said to be trapped in a cycle of suffering and rebirth due to their negative actions in past lives. The text provides detailed descriptions of each of the 108 types of beings, including their physical characteristics, their suffering, and the causes of their suffering. The text also includes advice on how to help these beings through the practice of Buddhism and the performance of rituals and offerings.

Q: ཁོན་ཀྱང་ངས་ཀྱང་བྱིས་པས་ང་གཞབ་ནན་མིན་པར་འདོད་ནས་ང་ལ་བརྩི་བཀུར་མི་བྱེད་པར་སེམས་ཁུར་བྱེད།

A:This is a Tibetan sentence, which can be translated to English as: "The one who has the power to do something, but does not do it, is like a person who has a treasure but does not use it.“ In this sentence, the word "ངས" (ngas) means "power" or "ability", and "ངང་བྱིས་པས" (ngyang bgyis pa) means "to have the power to do something". The word "ང་གཞབ་ནན་མིན" (ngag zhabs ngan mi) means "a treasure", and "རྩི་བཀུར་མི་བྱེད་པར" (rtsis bkuur mi byed pa) means "to use it". The sentence is using a metaphor to compare the person who has the power to do something but does not do it, to someone who has a treasure but does not use it.

Figure 9: 采用Llama2-7B-chat模型的案例

型MindL-Ti在回答藏语心理健康问题时，能够给出比较合理且人性化的回复。例如，在回答第一个和第三个问题时，该模型首先肯定了求助者的感受，然后给出了合理的建议。这样的回复方式更符合心理健康咨询场景的问答方式，有助于舒缓求助者的心情，并且更容易让求助者接受合理的建议。

6 结论

基于大语言模型的心理健康咨询对个性化的心理健康干预具有重要意义。针对现有心理健康咨询大语言模型难以泛化到低资源语言的问题，本文以藏文作为低资源语言的代表，基于GPT 4和藏汉翻译工具构造高质量的藏文心理健康多轮对话数据集，并基于多重筛选机制选择高质量的数据构成最终的藏文心理健康多轮对话数据；然后，利用该数据构造藏文心理健康大语言模型，选择LLaMA 2和Baichuan 2大语言模型作为基座模型，利用构建的心理健康咨询多轮对话数据对基座模型进行指令微调，构建了面向心理健康咨询的大语言模型MindL-Ti和MindB-Ti。最终的实验验证了本文构造的藏文心理健康多轮对话数据集和藏文心理健康大语言模型的有效性。

然而，由于藏文具有特殊的语法语义结构，现有大语言模型中的原始词汇表中缺少藏文词汇，虽然一些大语言模型，如LLaMA模型通过回退到字节来支持其他语言的字符，但这种回退策略增加了序列长度，并且难以很好地体现语言的语义信息，因此如何针对藏文数据构造词表，进行二次预训练对于提升大语言模型在藏文上的理解和生成能力具有重要意义。

参考文献

Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, 4:463–476.

John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine*, 183(6):589–596.

Siyuan Chen, Mengyue Wu, Kenny Q Zhu, Kunyao Lan, Zhiling Zhang, and Lyuchun Cui. 2023a. Llm-empowered chatbots for psychiatrist and patient simulation: application and evaluation. *arXiv preprint arXiv:2305.13614*.

Yirong Chen, Xiaofen Xing, Jingkai Lin, Huimin Zheng, Zhenyu Wang, Qi Liu, and Xiangmin Xu. 2023b. Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1170–1183.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2024. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 36.

Grace Golden, Christina Popescu, Sonia Israel, Kelly Perlman, Caitrin Armstrong, Robert Fratila, Myriam Tanguay-Sela, and David Benrimoh. 2024. Applying artificial intelligence to clinical decision support in mental health: What have we learned? *Health Policy and Technology*, page 100844.

Jia Yuan Lang Hai Liang Wang, Zhi Zhi Wu. 2020. 派特心理：心理咨询问答语料库.

Katharina Jeblick, Balthasar Schachtner, Jakob Dexl, Andreas Mittermeier, Anna Theresa Stüber, Johanna Topalis, Tobias Weber, Philipp Wesp, Bastian Oliver Sabel, Jens Ricke, et al. 2023. Chatgpt makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *European radiology*, pages 1–9.

Tin Lai, Yukun Shi, Zicong Du, Jiajie Wu, Ken Fu, Yichao Dou, and Ziqi Wang. 2023. Psy-llm: Scaling up global mental health psychological services with ai-based large language models. *arXiv preprint arXiv:2307.11991*.

Ming Li, Yong Zhang, Zhitao Li, Jiuhai Chen, Lichang Chen, Ning Cheng, Jianzong Wang, Tianyi Zhou, and Jing Xiao. 2023. From quantity to quality: Boosting llm performance with self-guided data selection for instruction tuning. *arXiv preprint arXiv:2308.12032*.

Michael Liebrenz, Roman Schleifer, Anna Buadze, Dinesh Bhugra, and Alexander Smith. 2023. Generating scholarly content with chatgpt: ethical challenges for medical publishing. *The lancet digital health*, 5(3):e105–e106.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144*.

June M Liu, Donghao Li, He Cao, Tianhe Ren, Zeyi Liao, and Jiamin Wu. 2023a. Chatcounselor: A large language models for mental health support. *arXiv preprint arXiv:2309.15461*.

Yue Liu, Thanh Le-Cong, Ratnadira Widyasari, Chakkrit Tantithamthavorn, Li Li, Xuan-Bach D Le, and David Lo. 2023b. Refining chatgpt-generated code: Characterizing and mitigating code quality issues. *ACM Transactions on Software Engineering and Methodology*.

Yuhan Liu, Anna Fang, Glen Moriarty, Robert Kraut, and Haiyi Zhu. 2023c. Agent-based simulation for online mental health matching. *arXiv preprint arXiv:2303.11272*.

Zihan Liu. 2022. *Effective Transfer Learning for Low-Resource Natural Language Understanding*. Hong Kong University of Science and Technology (Hong Kong).

Wei Qin, Zetong Chen, Lei Wang, Yunshi Lan, Weijieying Ren, and Richang Hong. 2023. Read, diagnose and chat: Towards explainable and interactive llms-augmented depression detection in social media. *arXiv preprint arXiv:2305.05138*.

Huachuan Qiu, Hongliang He, Shuai Zhang, Anqi Li, and Zhenzhong Lan. 2023. Smile: Single-turn to multi-turn inclusive language expansion via chatgpt for mental health support. *arXiv preprint arXiv:2305.00450*.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 86–96.

Aseem Srivastava, Tharun Suresh, Sarah P Lord, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. Counseling summarization using mental health knowledge guided utterance filtering. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3920–3930.

Hao Sun, Zhenru Lin, Chujie Zheng, Siyang Liu, and Minlie Huang. 2021. Psyqa: A chinese dataset for generating long counseling text for mental health support. *arXiv preprint arXiv:2106.01702*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Sai Vemprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. 2023. Chatgpt for robotics: Design principles and model abilities. *arXiv preprint arXiv:2306.17582*.

Zhongwei Wan, Xin Wang, Che Liu, Samiul Alam, Yu Zheng, Zhongnan Qu, Shen Yan, Yi Zhu, Quanlu Zhang, Mosharaf Chowdhury, et al. 2023. Efficient large language models: A survey. *arXiv preprint arXiv:2312.03863*, 1.

Jiaqi Wang, Enze Shi, Sigang Yu, Zihao Wu, Chong Ma, Haixing Dai, Qiushi Yang, Yanqing Kang, Jinru Wu, Huawen Hu, et al. 2023a. Prompt engineering for healthcare: Methodologies and applications. *arXiv preprint arXiv:2304.14670*.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023b. Self-instruct: Aligning language models with self-generated instructions. In *ACL*, pages 13484–13508.

Xuhai Xu, Bingsheng Yao, Yuanzhe Dong, Saadia Gabriel, Hong Yu, James Hendler, Marzyeh Ghassemi, Anind K Dey, and Dakuo Wang. 2024. Mental-llm: Leveraging large language models for mental health prediction via online text data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(1):1–32.

Aiyuan Yang, Bin Xiao, Bingning Wang, Borong Zhang, Ce Bian, Chao Yin, Chenxu Lv, Da Pan, Dian Wang, Dong Yan, et al. 2023. Baichuan 2: Open large-scale language models. *arXiv preprint arXiv:2309.10305*.

Qisen Yang, Zekun Wang, Honghui Chen, Shenzhi Wang, Yifan Pu, Xin Gao, Wenhao Huang, Shiji Song, and Gao Huang. 2024. Llm agents for psychology: A study on gamified assessments. *arXiv preprint arXiv:2402.12326*.

Jianxin Yang. 2023. Firefly(流萤): 中文对话式大语言模型. https://github.com/yangjianxin1/Firefly.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223*.

Chujie Zheng, Sahand Sabour, Jiaxin Wen, and Minlie Huang. 2022. Augesc: Large-scale data augmentation for emotional support conversation with pre-trained language models. *arXiv preprint arXiv:2202.13047*.