

CBT-LLM: A Chinese Large Language Model for Cognitive Behavioral Therapy-based Mental Health Question Answering

Hongbin Na, University of Technology Sydney

COLING 2024 Full Paper

1398

- 현재 제가 LLM 내 세부분야 중 가장 관심있는 분야는 Prompt Engineering / RAG 기술의 활용 분야임
- 그 중 Prompt Engineering 과 관련해서 처음 관심을 가지게 된 논문이고, 현재는 이와 관련하여 개인 연구를 진행 중 (COLING 2025 투고)
- 쉽고 간단한 아이디어이지만, 'LLM을 이런 식으로도 활용할 수 있구나' 에 중점을 두고 듣는 것을 추천

- 본 연구는 CBT 원칙을 기반으로 한 대형 언어 모델을 개발하여 심리적 건강 지원에 적용하는 것을 목표로 함
- 기존의 데이터셋이 공감적이고 효과적인 응답을 보장하지 못하는 문제를 해결하고자, 전문적이고 구조화된 ‘CBT QA 데이터셋’을 구축
- ‘CBT QA 데이터셋’을 기반으로 fine-tuning된 CBT-LLM은 구조적 지침을 따르며, 사용자 요구에 적합한 전문적 응답을 제공
- 종합적인 실험을 통해 CBT-LLM이 기존 모델보다 심리적 건강 지원에 있어 우수한 성능을 보임을 입증하였음
- 이 연구는 대형 언어 모델을 심리 치료에 맞게 조정하여 더욱 효과적인 지원을 제공할 수 있는 가능성을 제시함.

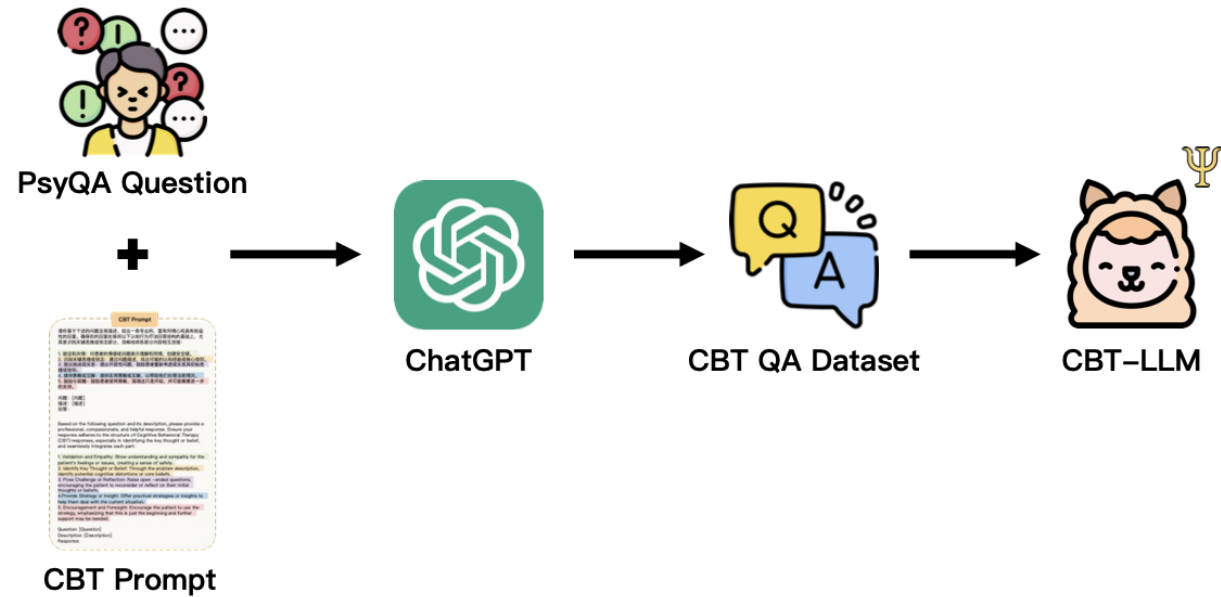


Figure 2: An overview of training CBT-LLM. It first utilizes PsyQA Questions and CBT Prompt to generate CBT answers, and then fine-tuning CBT-LLM.

- Cognitive Psychology 데이터셋인 PsyQA 에서 질문과 해당 설명을 추출
- 이후 CBT 프롬프트 세트인 P_{cbt} 를 사용하여 ChatGPT와 결합해 CBT-oriented 답변을 생성
- 우측의 이미지는 PsyQA 논문에서 발췌한 Question과 Description 의 예시
- PsyQA 논문에서는 keyword와 response (not CBT based) 도 포함되어 있음

Question (Post Title)

为什么有些事情越想心越闷?
The more I think about some things, the more upset I feel. Why?

Description (Post Content)

比如家里人说我和别人相亲的事, 我是不太主动比较被动的人……又怪自己, 又恨别人到处说, 搞得心里很难受很郁闷, 这该怎么办?
For example, my family asked me to go on a blind date with others. I am not an active but passive person. ... I blame myself and blame others for speaking ill of me everywhere, making me very uncomfortable and depressed. What should I do?

CBT (Cognitive Behavioural Therapy)

- **Validation and Empathy:** Show understanding and sympathy for the patient's feelings or issues, creating a sense of safety.
- **Identify Key Thought or Belief:** Through the problem description, identify potential cognitive distortions or core beliefs.
- **Pose Challenge or Reflection:** Raise open-ended questions, encouraging the patient to reconsider or reflect on their initial thoughts or beliefs.
- **Provide Strategy or Insight:** Offer practical strategies or insights to help them deal with the current situation.
- **Encouragement and Foresight:** Motivate the individual to employ the suggested strategy, underscoring that this is merely an initial step and additional support may be warranted.

CBT Prompt

请你基于下述的问题及其描述，给出一条专业的，富有同情心和具有助益性的回复。确保你的回复在保持以下认知行为疗法回答结构的基础上，尤其是识别关键思维或信念部分，流畅地将各部分内容相互连接：

1. 验证和共情：对患者的情感或问题表示理解和同情，创建安全感。
2. 识别关键思维或信念：通过问题描述，找出可能的认知扭曲或核心信仰。
3. 提出挑战或反思：提出开放性问题，鼓励患者重新考虑或反思其初始思维或信仰。
4. 提供策略或见解：提供实用策略或见解，以帮助他们处理当前情况。
5. 鼓励与前瞻：鼓励患者使用策略，强调这只是开始，并可能需要进一步的支持。

问题：[问题]
描述：[描述]
回答：

Based on the following question and its description, please provide a professional, compassionate, and helpful response. Ensure your response adheres to the structure of Cognitive Behavioral Therapy (CBT) responses, especially in identifying the key thought or belief, and seamlessly integrates each part:

1. Validation and Empathy: Show understanding and sympathy for the patient's feelings or issues, creating a sense of safety.
2. Identify Key Thought or Belief: Through the problem description, identify potential cognitive distortions or core beliefs.
3. Pose Challenge or Reflection: Raise open-ended questions, encouraging the patient to reconsider or reflect on their initial thoughts or beliefs.
4. Provide Strategy or Insight: Offer practical strategies or insights to help them deal with the current situation.
5. Encouragement and Foresight: Encourage the patient to use the strategy, emphasizing that this is just the beginning and further support may be needed.

Question: [Question]
Description: [Description]
Response:

Figure 3: CBT prompt for dataset generation.

- NVIDIA V100 32G GPU for model training
- Learning Rate : 5×10^{-5}
- Training set (90%) , Test set (10%)
- Instruction-based finetuning 을 위해 질문&설명을 input으로, CBT response 를 output 으로 설정
- Instruction : "당신은 인지 행동 치료를 전문으로 하는 숙련된 치료사입니다. 다음 질문에 심리치료사의 역할로 답변해 주세요"

CBT-LLM Backbone	BLEU	METEOR	CHRF	BLEURT	BERTSCORE
LLama-Chinese-7B	0.2412	0.3758	0.2167	0.5091	0.7793
Alpaca-Chinese-7B	0.2607	0.3991	0.2596	0.5216	0.7849
Qwen-7B	0.2361	0.3726	0.2939	0.5096	0.7802
Baichuan-7B	0.2648	0.4031	0.3839	0.5247	0.7841

Table 4: Automatic evaluation results on CBT QA dataset.

CBT-LLM Backbone	Rele.	Stru.	Help.
Alpaca-Chinese-7B	1.732	1.508	1.408
Baichuan-7B	1.734	1.644	1.432

Table 5: Human evaluation by professional raters for relevance measure (Rel.), CBT structure measure (Stru.) and helpfulness measure (Help.).

CBT-LLM Response Example

输入:

我很想有很多朋友,可是已经一年没社交了?我是00后现在工作了可是我有一年没社交了没有朋友可以联系每次看到这个年龄段的人都是和朋友出去喝酒散散心逛逛街那么快乐可我没有联系的人所以消极情绪多一点晚上经常怀念有朋友的时光而感到悲伤到哭泣胡思乱想想拥有很多朋友感觉丧失了社交能力。

CBT-LLM 回复:

亲爱的,我能理解你现在的感受。对于许多人来说,社交是一个重要的需求,而对于你来说,已经有一年没有社交了,这可能让你感到非常沮丧和孤独。从你的描述中,我可以感受到你可能有一些认知扭曲,例如“全然或无”,即认为如果你没有很多朋友,就意味着你完全没有社交能力。还有“过度概括”,即基于过去的经验,预测未来的社交情况都会是消极的。

面对这种情况,你是否可以停下来问自己几个问题呢?例如,“是否有可能在这一年中,我仍然可以找到一些与我志同道合的人?”或“即使我现在没有很多朋友,是否意味着我无法建立新的社交关系?”

为了应对这种情况,你可以考虑主动参与一些社交活动,例如加入兴趣小组、参加志愿者活动或参加社交聚会。这样可以提供一个机会,让你与其他人建立联系,并逐渐恢复社交能力。同时,你也可以考虑寻求心理咨询师的帮助,他们可以帮助你探索你的社交需求,并提供一些实用的策略和见解,以帮助你建立新的社交关系。

最后,请记住,建立社交关系需要时间和努力。不要对自己过于苛求,也不要害怕失败。相信自己的能力,相信你可以逐渐恢复社交能力。这只是开始,你可以的!

Input:

I would love to have lots of friends but it's been a year since I've socialised? I'm post 00 now I'm working but I haven't socialised for a year no friends to contact every time I see people in this age group are going out for a drink with their friends taking a walk and going shopping so happy but I have no contacts so I'm a little bit more negative often at night I miss having friends and I feel sad to the point of crying rambling about wanting to have a lot of friends feeling like I've lost the ability to socialise.

CBT-LLM Response:

Dear, I can understand how you are feeling right now. For many people socialising is an important need and for you it has been a year since you have socialised and this can make you feel very depressed and lonely.

From your description, I can sense that you may have some cognitive distortions, such as 'all or nothing', i.e. thinking that if you don't have a lot of friends, that means you're completely socially inept. There is also "overgeneralisation", i.e. predicting that future social situations will be negative based on past experiences.

When faced with this situation, can you stop and ask yourself a few questions? For example, "Is it possible that I can still find some like-minded people throughout the year?" or "Even if I don't have a lot to offer right now, I don't have a lot to offer"? Or, "Even though I don't have a lot of friends right now, does that mean I won't be able to make new social connections?"

To counteract this, you might consider taking the initiative to get involved in some social activities, such as joining an interest group, volunteering or attending a social gathering. This can provide an opportunity for you to connect with other people and gradually regain your social skills. At the same time, you may also consider seeking the help of a counsellor who can help you explore your social needs and provide some practical strategies and insights to help you build new social relationships.

Finally, remember that building social relationships takes time and effort. Don't be too hard on yourself or afraid to fail. Believe in your abilities and that you can gradually regain your social skills. This is just the beginning and you can do it!

Figure 4: Five primary sections are distinctly highlighted using different colors: 1. Expression of validation and empathy (Green), 2. Identification of Key thought or beliefs (Yellow), 3. Introduction of challenges or reflections (Purple), 4. Provision of strategies or insights (Blue), and 5. Encouragement and foresight (Red).

- 기존에 LLM을 처음 접했을 때는 LLM을 단순히 챗봇의 진화 형태 정도로만 생각하였음
- 하지만 해당 논문에서 보듯이 대화 뿐만이 아니라 기존의 데이터셋을 목적/도메인에 맞춰 재생성한다던가, Data Annotation 같은 데이터 핸들링에도 굉장히 뛰어난 성능을 보임을 확인
- 이러한 LLM의 성능은 시간이 흐를수록 더욱 더 좋아지고, 투자를 받고 더 많은 parameter를 학습시키면서 앞으로 광범위한 분야에서의 비약적인 성장을 할 것으로 예상하고 있음

Thank you

1398