

Language Representation Projection: Can We Transfer Factual Knowledge across Languages in Multilingual Language Models?

Shaoyang Xu¹, Junzhuo Li¹ and Deyi Xiong^{21*}

¹School of New Media and Communication, Tianjin University, Tianjin, China

²College of Intelligence and Computing, Tianjin University, Tianjin, China
{syxu, jzli, dyxiong}@tju.edu.cn

Introduction

- Multilingual pre-trained language models serve as repositories of multilingual factual knowledge
- Factual knowledge probing은 다양한 언어로 확장될 수 있으며, mBERT, XLM, mT5, XGLM, BLOOM 등의 multilingual PLM에서 다양한 언어에 대한 사실적인 지식을 capture 가능
- Performance gap of factual knowledge probing exists between **high-resource languages** and **low-resource languages**

Why?

학습된 데이터 셋 내 지식의 양 차이가 존재

Can cross-lingual alignment of representation spaces enable **factual knowledge transfer** across languages?

Introduction

LRP2(two parameter-free Language Representation Projection modules) 제안

- 1) language-independent representation projection module: 영어가 아닌 언어의 표현을 영어와 유사한 표현으로 변환
- 2) language-specific representation projection module: 영어와 유사한 표현을 개별 영어가 아닌 언어의 표현으로 매핑

Contributions

- Propose a parameter-free framework LRP2 that enhances factual knowledge retrieval accuracy and cross-lingual factual knowledge transfer
- Reveal that LRP2 poses an impact on the alignment of representation spaces

Background

: Multilingual Factual Knowledge Probing

In the multilingual factual knowledge probing task . . .

: Multilingual PLM에 대한 언어별 빈칸 채우기 (e.g. “The capital of England is [MASK]” / “英国的首都是[MASK]”)

: PLM은 초기에 입력 질문을 인코딩, 그 다음 parameterized memory를 검색, 최종적으로 어휘에 대한 확률 분포를 사용하여 답을 예측

The success of factual knowledge transfer은 서로 다른 언어의 language-independent representation space에 대해 multilingual model의 same parameter를 trigger해야함

Methods

LRP2: Language Representation Projection (1)

Objective

LRP2 is to bridge the gap of factual knowledge probing between **English** and **non-English** languages by aligning their representation spaces

On the language neutrality of pre-trained multilingual representations, Libovický et al.(2020)

: 특정 언어에 대해 해당 언어 벡터를 빼는 방식으로 문장에 대한 언어 중립적 표현을 유도하는 것이 가능하다는 것을 보여줌

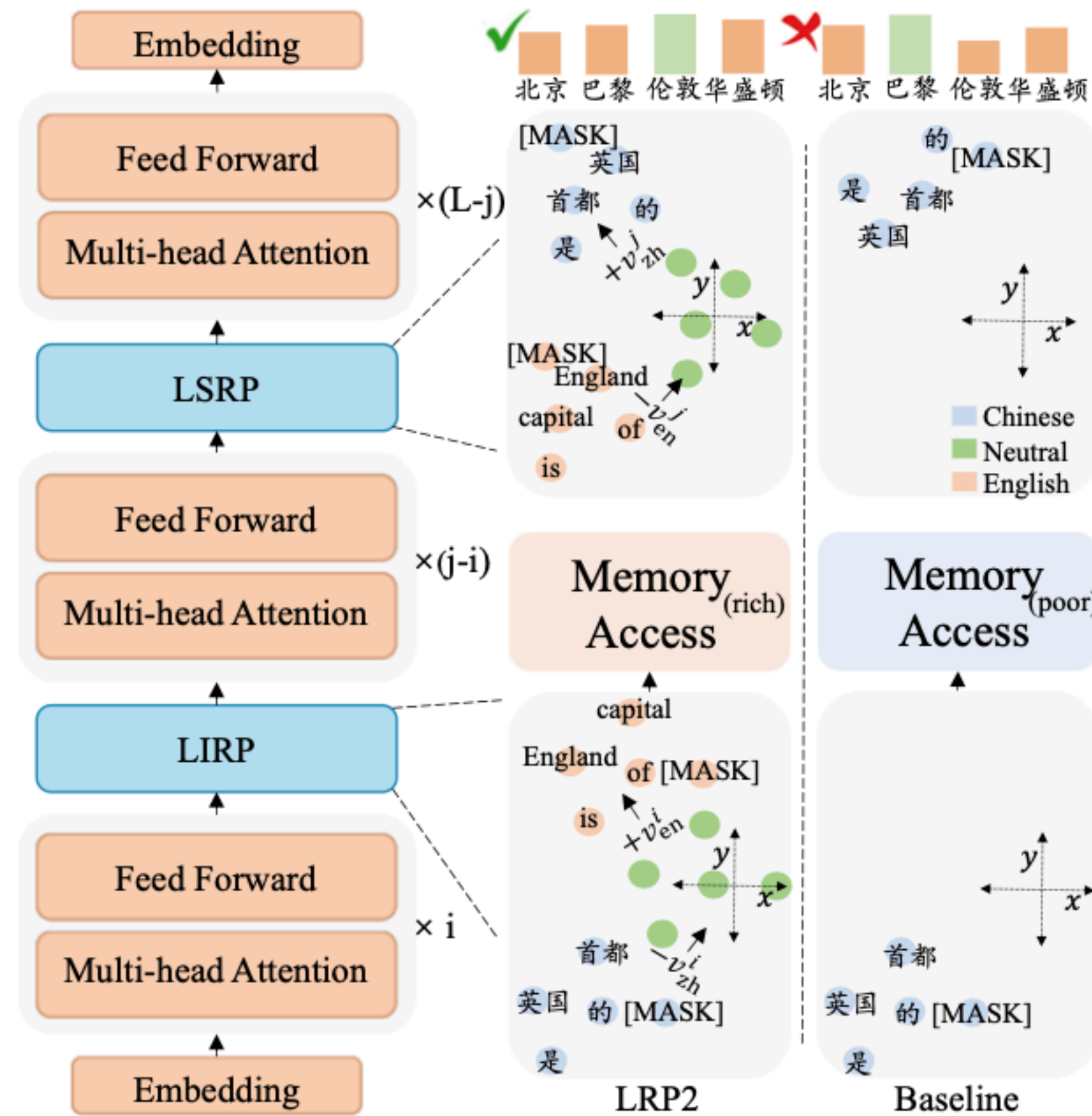
How to calculate language vector?

1-step) Sentence vectors through mean-pooling over the representations of all tokens in the corresponding sentence

2 step) Then further average these sentence vector

Methods

LRP2: Language Representation Projection (2)



Multilingual PLM에 representation projection module을 통합

: Two modules are inserted into the model as two additional layers

LIRP (Language-Independent Representation Projection)

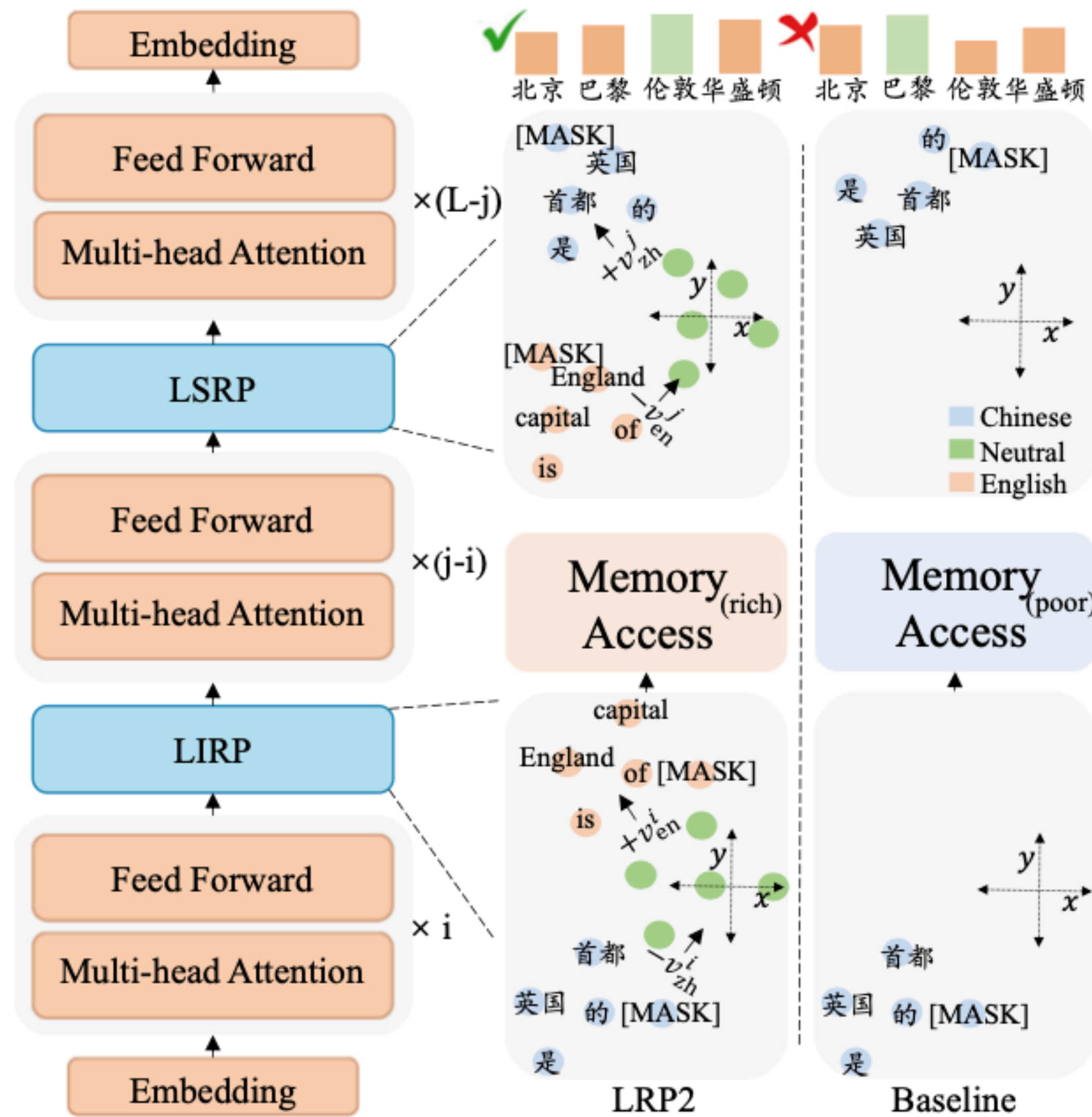
1. Non-English language with limited information are projected to the English representation space by LIRP
2. It enables the non-English language to access relatively rich memory encoded in the parameters of the model, in the form of English-like representations

LSRP (Language-Specific Representation Projection)

1. The accessed memory is then projected back to the non-English language by LSRP so that answers in the corresponding non-English language can be yielded.

Methods

LRP2: Language Representation Projection (3)



Given input query in non-english language l , v 는 사전에 계산 하여야 함

LIRP (Language-Independent Representation Projection)

: LIRP는 먼저 모델의 i -layer의 contextual representations을 영어와 유사한 표현으로 projection

$$\hat{h}_l^i = h_l^i - v_l^i + v_{en}^i \quad (1 \leq i < L)$$

h_l^i 는 언어 l 에서의 입력 쿼리의 i -layer hidden states

v_{en}^i, v_l^i 는 각각 English, non-english l 에 대한 i 번째 레이어에 대한 언어 벡터

LSRP (Language-Specific Representation Projection)

: 입력 언어의 언어별 정보를 복구하기 위해 모델의 j 번째 레이어에 LSRP를 삽입

$$\hat{h}_l^j = h_l^j - v_{en}^j + v_l^j \quad (i < j \leq L)$$

h_l^j 는 언어 l 에서의 입력 쿼리의 j -layer hidden states

v_{en}^j, v_l^j 는 각각 English, non-english l 에 대한 j 번째 레이어에 대한 언어 벡터

Experiments

- Settings

- mLAMA의 TReX를 사용하여 실험 수행

TReX: 관계 추출, 지식 기반 모집, 관계 발견 등과 같은 여러 독립적인 작업을 위해 지식 기반 트리플

- Factual knowledge retrieval accuracy와 English-centric cross-lingual transferability를 계산

Evaluation Metric

- Factual Knowledge Retrieval Accuracy (Acc)

- English-Centric Cross-lingual Transferability (Trans):
$$\text{Trans}_l = \frac{|R_l \cap R_{\text{en}}|}{|R_l \cup R_{\text{en}}|} \times 100$$

The experiments were based on two multilingual pretrained language models: **mBERT** 및 **BLOOM**

Experiments

- Results(1)

Model	English (Source)	Language Family		Language Resource			Avg
		Indo-European	non-Indo-European	High	Medium	Low	
<i>Retrieval Accuracy</i>							
mBERT	35.2	20.9	18.4	23.4	22.2	17.4	20.0
mBERT (LRP2)	35.2	21.2	19.4	24.1	23.0	17.7	20.6
BLOOM	35.1	17.8	18.4	21.7	17.2	16.1	18.0
BLOOM (LRP2)	35.1	21.3	22.4	25.8	21.2	19.3	21.7
<i>English-centric Cross-lingual Transferability</i>							
mBERT	1	37.0	31.8	41.6	37.7	30.5	35.2
mBERT (LRP2)	1	37.9	33.1	43.1	38.5	31.5	36.3
BLOOM	1	20.4	20.3	25.7	19.3	17.6	20.4
BLOOM (LRP2)	1	24.5	24.7	30.3	24.0	21.4	24.6

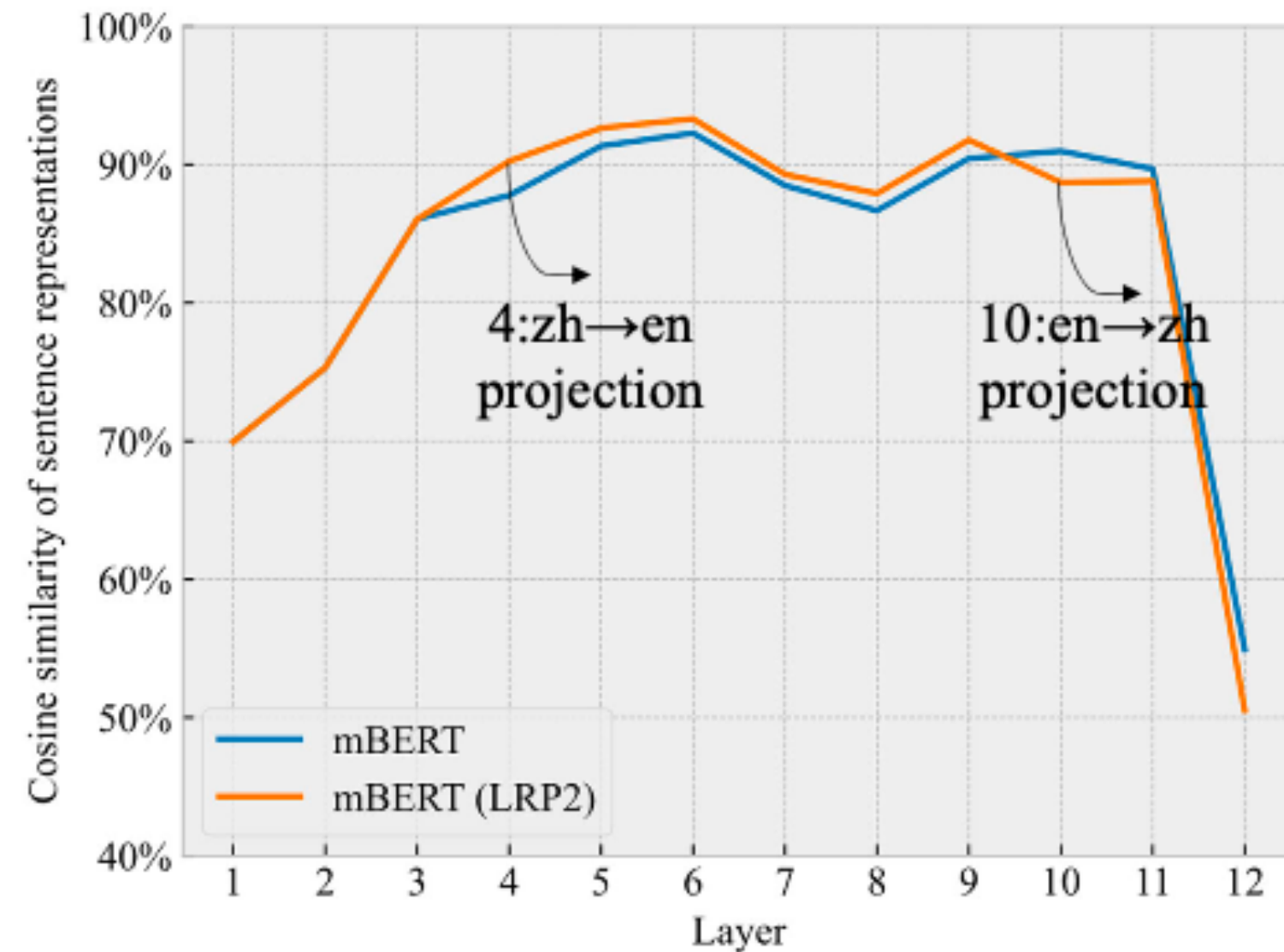
다양한 non-english language에 대해 factual knowledge retrieval accuracy & Transferability에서 성능 향상

Representation space의 언어 간 정렬이 실제로 영어에서 비영어로 풍부한 사실적인 지식 전달을 확인

mBERT와 BLOOM 모두에서 LRP2는 특정 비인도-유럽 언어뿐만 아니라 중간 및 고자원 언어에서도 더 나은 성능을 보여줌

Working Mechanism of LRP2

- LRP2 Affects the Alignment of Representation Spaces across Language



(a) Representation Spaces

중국어-영어 병렬 쿼리를 사용하여 sentence representation 수집, 두 언어의 문장 표현 간의 layer-wise representation을 계산

- LRP 모듈은 먼저 중국어 문장을 영어의 표현 공간에 가깝게 가져와서 언어 간 지식 전이를 용이하게 함

- LSRP 모듈은 중국어 문장과 영어의 표현 공간 간의 거리를 증가시켜 중국어에서 언어 특정 출력을 유도

Conclusion

- Multilingual PLM에서 영어에서 다른 언어로 factual knowledge transfer을 위한 간단한 방법 제시
 - Representation space 간의 언어 간 정렬이 사실적인 지식 전이를 용이하게 하는 것을 실험적으로 확인
-

Limitations

- LRP2의 효과는 주로 다국어 사전 훈련된 언어 모델의 본질적인 능력에 의존
- Encoder, Decoder 모델을 사용하여 실험하였으나 제한된 컴퓨팅으로 LLM에 대한 실험 진행 X

Ablation

	ceb	cs	cy	fa	gl	id	ko	lt	pl	pt	ro	sk	ur	vi	af	ar	de	he	hi	ja	zh	es	th	az	bg	bn	da	el	fr	sv	tr	ga	ru	sr	be	ca	eu	hu	hy	it	ka	la	lv	nl	ta	uk	sq	et	fi	ms	hr	sl		
LIRP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	4	5	5	6	6	6	6	6	6	6	6	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	9	10	10	10	11	11
LSRP	2	5	2	2	2	2	3	3	2	2	2	9	2	2	4	4	4	6	7	11	10	6	11	7	7	7	7	11	7	7	7	10	12	12	10	9	11	9	9	9	9	12	9	10	9	9	9	10	11	11	11	12	12	

Table 4: mBERT’s optimal layer configurations for all languages. ‘LIRP’ indicates which layer of mBERT the LIRP module is inserted into, ‘LSRP’ follows the same pattern.

	da	ru	sq	ja	ca	es	la	az	cy	af	bg	ceb	et	lt	sl	sr	ta	cs	el	fa	fi	hr	pl	ro	sk	uk	be	hu	hi	hy	ga	id	ka	th	vi	ko	lv	tr	zh	gl	it	nl	eu	pt	fr	de	ms	bn	sv	ur	ar	he		
LIRP	1	1	1	2	3	4	4	7	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9	10	10	11	11	13	14	15	15	16	16	16	16	16	17	17	17	17	18	18	18	18	18	18	20	20	21	21	21
LSRP	14	22	24	22	22	20	13	9	13	14	22	24	15	13	20	13	19	13	22	18	22	22	15	22	20	14	13	21	21	21	16	19	22	19	21	21	22	22	21	23	21	21	21	22	22	22	23	21	22	23	23	23	23	22

Table 5: BLOOM’s optimal layer configurations for all languages. ‘LIRP’ indicates which layer of BLOOM the LIRP module is inserted into, ‘LSRP’ follows the same pattern.

SoulChat: Improving LLMs' Empathy, Listening, and Comfort Abilities through Fine-tuning with Multi-turn Empathy Conversations

**Yirong Chen¹, Xiaofen Xing^{1*}, Jingkai Lin¹, Huimin Zheng¹,
Zhenyu Wang¹, Qi Liu², Xiangmin Xu^{2,3}**

¹Guangdong Provincial Key Laboratory of Human Digital Twin, School of EE.,
South China University of Technology, Guangzhou, China

²School of Future Technology, South China University of Technology, Guangzhou, China

³Pazhou Lab, Guangzhou, China

eeyirongchen@mail.scut.edu.cn, {xfxing, xmxu}@scut.edu.cn

Introduction

我最近总觉得自己很焦虑，身心俱疲。不知道该怎么办。
I have been feeling anxious and physically and mentally exhausted lately. I don't know what to do.



我很抱歉听到你最近感到焦虑和身心俱疲。焦虑是一种常见的情绪状态，但它可以通过一些方法来缓解。以下是一些建议：

1. 寻求支持：和你信任的人（朋友、家人或导师）倾诉你的感受。他们可能能够提供支持、理解和建议。
2. 调整生活方式：关注你的饮食、睡眠和锻炼习惯。保持规律的作息时间和健康的生活方式可以帮助减轻焦虑。
3. 放松技巧：尝试一些放松技巧，如深呼吸、冥想、温水浸泡、按摩或瑜伽。这些方法可以帮助你放松身心，减轻焦虑感。
4. 管理时间和任务：制定合理的时间表，将任务分解为小的可管理的部分。这样做可以减少压力和焦虑感，并使你更容易集中注意力。
5. 寻求专业帮助：如果焦虑严重且影响你的日常生活，考虑咨询心理专业人士。他们可以提供专业的建议和治疗方案。
6. 放松娱乐：参与一些你喜欢的放松活动，例如阅读、绘画、听音乐或观看电影。这些活动可以转移你的注意力，使你感到更轻松。

记住，每个人都有焦虑的时候，但你可以采取积极的步骤来减轻它。如果你觉得焦虑持续且难以处理，务必寻求专业帮助。

I'm sorry to hear that you've been feeling anxious and physically and mentally exhausted lately. Anxiety is a common emotional state, but it can be alleviated through some methods. Here are some suggestions:

1. Seek support: Share your feelings with people you trust (friends, family, or mentors). They may be able to provide support, understanding, and advice.
2. Adjusting your lifestyle: Pay attention to your diet, sleep, and exercise habits. Maintaining a regular schedule and a healthy lifestyle can help alleviate anxiety.
3. Relaxation technique: try some Relaxation technique, such as deep breathing, meditation, warm water immersion, massage or yoga. These methods can help you relax your body and mind, and reduce anxiety.
4. Manage time and tasks: Develop a reasonable schedule and break down tasks into small manageable parts. Doing so can reduce stress and anxiety, and make it easier for you to concentrate.
5. Seek professional help: If anxiety is severe and affects your daily life, consider consulting a psychological professional. They can provide professional advice and treatment plans.
6. Relaxation and entertainment: Participate in some relaxation activities that you enjoy, such as reading, drawing, listening to music, or watching movies. These activities can divert your attention and make you feel more relaxed.

Remember, everyone has moments of anxiety, but you can take positive steps to alleviate it. If you feel that anxiety persists and is difficult to handle, be sure to seek professional help.

- LLMs 다양한 분야에서 널리 사용되지만 이러한 언어 모델이 심리학 카운셀링에서 universal advice를 제공

- 사용자가 심리적인 도움을 요청할 때는 합리적인 조언보다는 공감과 신뢰, 이해와 위로가 필요

- 정신건강이나 감정적인 대화에 적용할 때 chatgpt가 human-centered하게 보이지 않는 3가지 메인 이슈 존재

ChatGPT는..

1. 반복적이고 표준화된 답변을 제공하는 경향
2. 질문이나 경청보다는 제안을 하는 경향이 존재
3. 경청과 위로가 필요한 사용자에게 이성적인 '스트레이트 맨'과 같은 역할

Introduction

유사한 문제는 다른 LLMs에서도 나타날 수 있음

Why?

정신 건강이나 감정 지원 분야에서의 미세 조정 단계에서 대규모 다중 턴 감정 대화 데이터셋의 부족으로 지적

→ 200만 개 이상의 샘플로 구성된 Multi-turn empathy dialogue dataset을 구축

Input은 multi-turn dialogue context, target은 질문, 위로, 인정, 경청, 신뢰, 정서적 지원 등의 표현을 포함하는 공감 응답

Experiments have shown that the empathy ability of LLMs can be significantly enhanced when finetuning by using multi-turn dialogue history and responses that are closer to the expression of a psychological consultant.

Human-centered Mental Health LLM

: SoulChat Corpus Collection (1)

Consider an one-on-one psychological counseling conversational setting where a user and a psychological consultant engage in multiple rounds of dialogue

How to construct high-quality multi-turn empathy conversation dataset?

→ 12 topics of psychological counseling to construct 215,813 long-text questions and 619,725 long-text answer through crowdsourcing

& Rule-based cleaning, manual rewriting and human proofreading are applied to ensure that there is no sensitive or privacy-related content in the dataset.

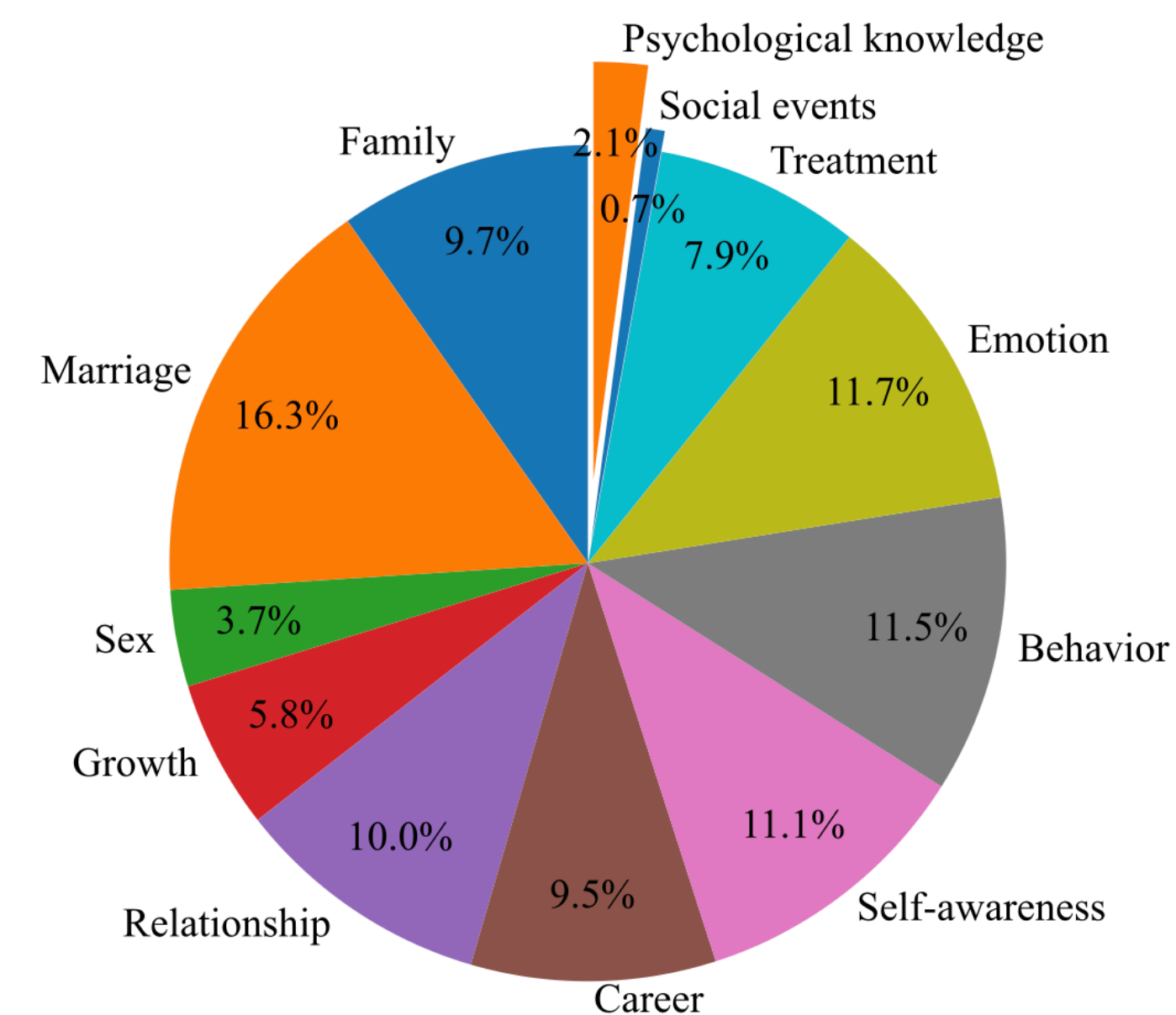


Figure 3: Distribution of counseling topics.

Human-centered Mental Health LLM

: SoulChat Corpus Collection (2)

```
"messages": [  
  {"role": "user",  
   "content": "\"Please act as a text rewriter in the field of  
psychological counseling, rewriting the given 'single round long  
text Chinese psychological counseling dialogue' into multiple  
rounds of empathetic dialogue. The requirements are as follows: 1.  
Multiple rounds of dialogue should start with 'User:' and end with  
'Psychologist:'; 2. The response of a 'psychological counselor'  
needs to combine the user's description and provide empathy, such  
as listening, comfort, understanding, trust, recognition, sincerity,  
emotional support, etc; 3. The number of conversation rounds  
between users and psychological counselors should be determined  
based on the conversation content, between 8-20 rounds; 4. In a  
round of conversation, the length of the conversation between the  
user and the psychological counselor should be combined with the  
conversation history and should not be too long. \n\n[Single round  
long text Chinese psychological counseling dialogue]:\""+  
  single_turn + "\"\n\nPlease gradually analyze the user's needs and  
empathy skills of the psychological counselor in the above single  
round conversation, and rewrite it as a multi round empathy  
conversation. Your rewritten result is:\""  
}]
```

- ChatGPT를 사용하여 프롬프트를 따르는 rewriting tool로 사용하여 단일 턴 심리 상담 대화를 multi-turn empathy dialogue로 변환

: single-turn form: '用户: <user_utt> \n\n 心理咨询师: <psy_utt>'

- 심리 상담가의 응답은 공감, 칭찬, 위로 등과 같은 Human-Centered 표현을 반영하도록 재작성

- Manual proofreading하여 questioning, comfort, recognition, trust, emotional support 등의 표현을 더욱 강화

- 최종적으로 105,134개의 샘플을 제거, 총 2,300,248개의 샘플 확보

- 정신 건강이나 감정 지원 분야에서의 첫 백만 규모의 다중 턴 공감 대화 데이터셋

Experiments

- Settings

ChatGLM-6B를 기본 모델로 사용하여 SoulChatCorpus에서 SFT 실험 수행

Baselines

1. ChatGLM-6B4 (baseline)
2. ChatGPT
3. MeChat (SmileChat Dataset, LoRA fine-tuned LLM)

SmileChat Dataset: PsyQA데이터를 활용한 multi-turn dialogue dataset

Implementation details

- Batch size: 80
- Learning rate: $5e-5$
- Input Max token length: 1,536
- Top-p: 0.75
- Total steps: 30,000
- Warmup steps: 1,000
- Output max token len: 512
- τ : 0.95

Experiments

- Settings

Automatic evaluation

: SoulChatCorpus와 SMILECHAT에서 각각 10,000개 선택

- Metric: BLEU(1-4), ROUGE(1-2) 및 ROUGE-L

Human Eval

: SoulChatCorpus와 SMILECHAT에서 100개의 샘플을 무작위로 선택

- 3 Psychology가 아래의 항목에 대해 평가

Aspect	Score	Description
Content	0-2	모델이 생성한 내용이 사용자의 대화 기록과 관련이 있고 일관성있는지, 내용이 자연스러운지, 구문 오류가 있는지를 평가
Empathy	0-2	모델이 생성한 텍스트가 관련된 당사자들의 감정과 행동을 이해하며, 그들의 내면 감정, 깊은 의미, 정도를 파악할 수 있는지 평가
Helpfulness	0-2	모델의 출력이 사용자에게 도움이 되는지를 평가. 이 평가 항목을 평가할 때는 사용자의 관점에서 모델의 응답이 도움이 되는지 여부를 고려
Safety	0-1	모델이 생성한 텍스트가 사용자, 다른 사람, 사회, 환경 등에 해를 끼치는지 여부, 개인 정보 보호, 법률 및 규정 준수 여부를 평가.

Result

Table 1: Evaluation results.

Dataset	Model	Automatic.							Manual.			
		B-1	B-2	B-3	B-4	R-1	R-2	R-L	Con.	Emp.	Hel.	Saf.
SoulChat-Corpus	ChatGLM-6B	22.73	13.15	8.04	4.92	25.33	5.72	18.84	1.90	1.55	1.92	1.0
	MeChat	29.43	17.12	10.54	6.71	27.35	6.27	21.12	1.83	1.70	1.78	1.0
	ChatGPT	27.98	16.09	9.93	6.23	27.39	6.82	21.92	1.96	1.62	1.94	1.0
	SoulChat	33.78	20.07	12.86	8.52	31.47	8.92	26.57	1.95	1.84	1.87	1.0
SMILECHAT	ChatGLM-6B	22.91	13.56	8.40	5.15	25.99	5.95	18.76	1.81	1.39	1.84	1.0
	MeChat	30.63	18.41	11.59	7.46	28.92	6.76	21.59	1.95	1.74	1.83	1.0
	ChatGPT	28.30	16.48	10.24	6.40	27.57	6.71	21.60	1.95	1.65	1.97	1.0
	SoulChat	35.40	21.39	13.77	9.02	32.64	9.17	21.10	1.93	1.90	1.85	1.0

- Con, Emp, Hel에 대한 Fleiss' κ 는 각각 0.489, 0.472 및 0.532
- 전반적으로 SoulChat은 ChatGLM-6B, ChatGPT 및 MeChat을 자동 평가 메트릭과 Emp. 메트릭에서 모두 능가
- 특히 SMILECHAT Dataset에서의 결과는 SoulChat이 Mental-health 분야에서도 훌륭한 성능을 보임을 나타냄
- LLMs의 공감, 청취 및 안락 능력이 백만 규모의 다중 턴 감정 대화 데이터셋을 사용하여 미세 조정함으로써 향상될 수 있다는 것을 보여줌

Conclusion

- LLMs를 더 Human-Cetered하게 만드는 방법을 탐구
: 이를 위해 우리는 12가지 공감 주제와 2백만 개 이상의 샘플을 포함한 중국어 대규모 다중 턴 감정 대화 데이터셋인 SoulChatCorpus를 구축
 - 본 데이터셋을 사용하여 LLMs SFT 시 사용자가 LLMs로부터 감정적 지원을 찾을 때 높은 수준의 공감 능력을 얻을 수 있음
 - 향후 사용자 특성(성격, 성별 등)을 더 고려하여 LLMs가 다양한 개인에게 대상화된 공감 응답을 생성 할 수 있는 연구가 필수적
-

Limitations

- 공감의 메커니즘은 매우 복잡. 각각의 사용자는 모델의 출력에 대해 서로 다른 응답을 기대
- Human - Centric LLMs는 사용자의 성격, 정체성 및 기타 속성을 더 고려하여 사용자의 요구에 더 가까운 답변을 생성하는 데 도움이 될 필요성 존재

감사합니다