

# 프롬프트 동적 선택을 통한 언어 모델 성능 향상에 대한 연구 :UPRISE 를 기반으로

홍지민, 김선욱, 김민찬, 김남수  
서울대학교 전기정보공학부 뉴미디어통신공동연구소 휴먼인터페이스 연구실  
{jmhong, swkim, mckim}@hi.snu.ac.kr, nkim@snu.ac.kr

## A Study on Enhancing Language Model Performance through Dynamic Prompt Selection: Based on UPRISE

Ji Min Hong, Seonuk Kim, Minchan Kim, Nam Soo Kim

Human Interface Laboratory

Department of Electrical and Computer Engineering and INMC,

Seoul National Univ.

### 요약

본 대형 언어 모델(LLM)의 답변 품질은 사용자 요청의 표현 방식에 따라 크게 영향을 받는다. 본 논문에서는 LLM의 성능을 개선하기 위해, 사용자의 요청과 유사한 질문-답변 쌍을 추출해 프롬프트로 활용하는 UPRISE 모델을 동적으로 개선하는 방식을 제안하였다. 실험 결과, 동적 방식이 기존 방식보다 효과적임을 확인하였다.

### I. 서론

사용자는 대형 언어 모델(LLM)을 통해 필요한 질문이나 요청에 대한 답변을 얻어 유용하게 사용할 수 있다. 그러나 이런 LLM의 경우 사용자가 구체적으로 어떤 표현으로 지시하는지에 따라 그 답변의 품질이 크게 달라진다는 점이 잘 알려져 있다. 이에 따라 요청을 효과적으로 처리할 수 있는 프롬프팅(prompting) 방법이 활발히 연구되고 있다.

그러나 사용자가 항상 최적의 프롬프트를 생성하기란 쉽지 않기 때문에, 수동적인 방식에서 벗어나 자동으로 프롬프트를 생성하여 기존 입력과 결합해주는 프레임워크 등의 연구가 진행되고 있다. 이러한 프레임워크는 외부 입력된 쿼리에 기반해 자동으로 프롬프트를 작성하고 언어 모델의 성능을 향상시키는 데 기여할 수 있다.

본 논문에서는 이러한 모델 중 하나인 UPRISE[1]와 관련해 이를 더 효과적으로 사용할 수 있는 방법을 연구하고자 하였다. UPRISE는 다양한 질문-답변 쌍을 바탕으로 프롬프트 풀(prompt pool) 구축하고 사용자의 요청이 들어왔을 때 비슷한 쌍  $n$  개를 추출하여 이를 프롬프트로 하여 요청에 붙여서 언어모델에 입력하도록 하는 방식이다. 이 논문에서는 추출되는 질문-답변 쌍의 개수  $n$  을 동적으로 결정하도록 개선하였다. 실험을 통해 이 방식이 더 효과적임을 확인하였다.

### II. 본론

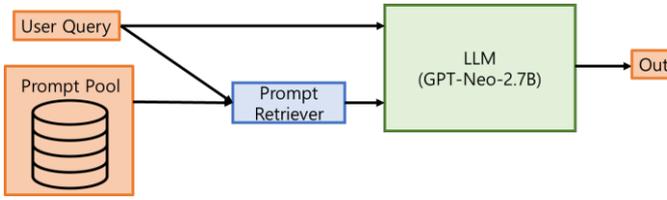
#### i. 프롬프팅(prompting)과 few-shot learning

프롬프팅(Prompting)은 대형 언어 모델(LLM)의 성능을 극대화하기 위해 작업 지침이나 예시를 입력에 포함시키는 방법으로, 다양한 자연어 처리(NLP) 작업에서 중요한 역할을 한다. 효과적인 프롬프팅은 모델이 요청의 맥락을 더 잘 이해하고, 적절한 응답을 생성하도록 돕는다. 어떠한 방법으로 구성할지에 따라 그 성능이 크게 좌우된다.

이 중 Few-shot Learning은 프롬프팅 방법의 대표적인 사례로, 제한된 수의 예시를 제공하여 모델이 새로운 작업에 적응할 수 있도록 한다. Few-shot learning에서는 주어진 작업과 유사한 몇 가지 예시(예: 질문-답변 쌍, 요약 문장 등)를 프롬프트로 포함시켜 모델의 학습과 예측 성능을 향상시킨다. 이 방식은 대량의 추가 학습 데이터 없이도 모델이 높은 적응력을 보이는 점에서 특히 유용하다[2].

#### ii. UPRISE[1]

본 논문에서 베이스라인으로 잡은 UPRISE[1]는 prompt pool에서 검색기를 통해 적합한  $n$ (원 논문 실험에서는 3)개를 추출하여 원래 요청에 붙이게 되는데, 해당 pool의 데이터는 FLAN[3] 형식으로 제공되었으며, 이는 NLI, QA, 요약 등 다양한 task로 구성되었다. 학습과 평가를 위한 언어모델은 EleutherAI의 GPT-Neo-2.7B[4]가 사용되었으며, 프롬프트 검색을 위한 retriever는 BERT<sub>BASE</sub>[5]를 기반으로 학습되었다.



### iii. 연구 방법론

원래의 UPRISE[1]는 retriever 를 통해 검색되어 쿼리에 붙여지는 프롬프트의 개수를  $n$  개로 고정하여 실험하며,  $n=3$  환경에서 전체 평가를 수행하였다. 다시 말해, 프롬프트와 쿼리의 유사도 점수의 절댓값과는 관계없이 상위 3 개의 프롬프트를 붙이는 방식으로 작동한다. 본 논문에서는 유사도 점수의 절대적인 크기가 의미를 가질 것이라는 가정에 따라, 그 점수를 바탕으로 원래 쿼리의 붙이는 프롬프트의 개수를 동적으로 변경하고자 했다. 점수가 높은 경우 입력된 쿼리와 검색된 프롬프트 간의 유사도가 높기에 적은 수도도 효과적인 성능을 낼 수 있을 것이며, 점수가 낮은 경우 조금 더 많은 예시를 주어 언어 모델의 답변 생성에 도움을 줄 것으로 생각된다.

따라서 임계값을 설정하여 해당 임계값 보다 검색된 프롬프트의 유사도의 합(개수에 따라 스케일링 된 값)보다 높을 때까지 추가적으로 프롬프트를 더하는 방식으로 설정하였으며, 점수가 낮은 경우 무분별한 검색을 방지하기 위해 프롬프트 개수의 상한(본 실험에서는 3)을 설정하였다.

실험 결과는 다음과 같으며, 해당 실험은 원 UPRISE[1]의 실험과 같은 task 와 조건에서 진행되었다.

task	RTE (NLI task)	MRPC (Paraphrase Detection)	ARC-e (Closed QA)
0-shot	33.6%	46.6%	45.7%
UPRISE	34.7%	<b>67.9%</b>	55.6%
Proposed	<b>38.2%</b>	66.7%	<b>56.2%</b>

Proposed Method 가 더 우수하거나 유사한 성능을 보이는 것을 실험적으로 확인하였다. 또한, UPRISE 가 항상 일정한 개수의 프롬프트를 원래 쿼리에 더하는 것과 달리, 제안된 방법의 경우, 유사도 점수가 높은 경우 더 적은 수의 프롬프트를 붙이기 때문에, 한정된 자원을 효과적으로 활용할 수 있다는 점이 이점이다. 따라서, 본 논문에서 제시한 동적 프롬프트 개수 할당 방식의 효과가 입증되었다.

### III. 결론

본 논문에서는 대형 언어 모델의 성능을 개선하기 위해 프롬프트 개수를 동적으로 조정하는 방식을

제안하였다. 실험 결과, 제안된 방법은 기존 UPRISE 더 우수하거나 유사한 성능을 보임을 확인하였다. 또한, 프롬프트를 필요한 만큼만 사용함으로써 입력이 비용이 되는 언어 모델의 효과적인 사용을 가능하게 했다. 이를 통해, 동적 프롬프트 할당 방식이 대형 언어 모델의 성능 향상에 효과적인 접근법임을 입증하였다.

### ACKNOWLEDGMENT

이 논문은 2024 년도 BK21 FOUR 정보기술 미래인재 교육연구단에 의해 지원되었음.

### 참 고 문 헌

- [1] D. Cheng et al., 'UPRISE: Universal Prompt Retrieval for Improving Zero-Shot Evaluation', in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 12318-12337.
- [2] Tom B. Browne et al., 'Language models are few-shot learners', in NeurIPS, 2020
- [3] Jason Wei et al., 'Finetuned language models are zero-shot learners', in ICLR, 2022
- [4] Sid Black et al., 'GPT-Neo: Large Scale Autoregressive Language Modeling with MeshTensorflow', 2021
- [5] Jacob Devlin et al., 'BERT: pre-training of deep bidirectional transformers for language understanding', in NAACL-HLT, 2019, pp. 4171-4186.