

# 프롬프트를 이용한 표 질의응답의 성능향상

박정연<sup>0</sup>, 이동혁, 신형진, 조경빈, 이재성  
충북대학교

{parkjeongyeon, nlp\_dohy, shinhj, whrudqls56, jasonlee}@cbnu.ac.kr

## Improving Table Question Answering Using Prompt

Jeongyeon Park<sup>0</sup>, Donghyeok Lee, Hyeong Jin Shin, Kyungbeen Cho, Jae Sung Lee  
Chungbuk National University

### 요약

표 질의응답이란, 주어진 표에서 질의문에 대한 답변을 자동으로 추출하거나 생성하는 기술을 말한다. 최근 언어모델을 사용한 연구들은 정답을 유도할 수 있는 명령문인 프롬프트를 활용하여 더 높은 성능을 보이고 있다. 본 연구에서는 표 질의응답의 성능을 향상시키기 위해, 프롬프트를 효과적으로 사용할 수 있는 모델을 제안한다. 이와 함께, 다양한 형태의 프롬프트를 사용하여 모델을 평가한다. 실험 결과, 기본 모델에 단순 질의문만 입력으로 사용했을 때의 성능 F1 67.5%에 비해, 다양한 프롬프트를 입력으로 사용한 경우 1.6%p 향상된 F1 69.1%를 보였다. 또한, 다양한 프롬프트와 함께 제안 모델을 사용했을 때에는 기본 모델보다 2.2%p 높은 F1 69.7%를 달성했다.

주제어: 표 질의응답, 프롬프트, BERT, TAPAS

### 1. 서론

표 질의응답(Table Question Answering)이란, 주어진 표에서 질의문에 대한 답변을 자동으로 추출하거나 생성하는 기술이다[1,2]. 과거의 표 질의응답은 데이터베이스를 사용하여, 그 데이터베이스에 맞는 쿼리를 사용하는 것으로 질의응답이 이루어졌었다. 그 후, 트랜스포머 기반의 언어모델인 BERT[3] 등이 개발되어 사용되면서, 자연어로 질의응답을 처리하는 기술이 개발되고 있다 [1]. 이 방법은 주로 사전학습된 언어모델을 사용하여 표 데이터 구조를 표현하기 위한 정보들을 입력으로 함께 사용한다.

최근 질의응답을 포함해 많은 자연어처리 작업에서 사용되는 언어모델은 작업에 맞는 설명이나 예시가 포함된 프롬프트를 입력으로 제공하거나, 언어모델의 학습환경과 비슷한 방법으로 튜닝하는 등 프롬프트를 사용하여 모델을 학습을 했을 때 그 성능이 높아지는 경향을 보이고 있다[4-6].

GPT 등의 디코더 구조의 언어모델 기반의 질의응답 모델에서는, In-context learning 등 프롬프트를 사용하는 많은 방법이 제안되어 왔다[4]. 그러나, BERT 등의 인코더 구조의 언어모델 기반 질의응답에서는 프롬프트를 효과적으로 사용하는 방법이 상대적으로 적다. 또한, 표 질의응답에서 프롬프트를 사용한 경우는 찾기 어렵다.

본 연구에서는 표 질의응답을 위해 프롬프트를 사용하고, 프롬프트를 효과적으로 학습할 수 있는 모델을 제안한다. 실험에서 기본 모델은 F1 67.5%를 보였는데, 기본 모델과 함께 다양한 프롬프트를 사용했을 때에는 기본 모델보다 1.6%p 높은 F1 69.1%를 달성했다. 또한, 다양한 프롬프트와 제안한 모델을 사용했을 때에는 기본 모델보다 2.2%p 높은 F1 69.7%를 달성했다.

### 2. 관련 연구

#### 2.1 표 질의응답

대표적인 표 질의응답 모델은 BERT 모델을 기반으로 하는 TAPAS, TABLEFORMER 등이 있다[1,2]. 이 연구들은 표 데이터 구조에 대한 정보인 행렬의 위치값, 질의문과 표 데이터 사이의 13가지 관계(질의문과 표 데이터 내용, 질의문과 표 데이터 열 제목 등)를 추가 입력으로 사용한다. 그런데, BERT는 비정형 텍스트로부터 학습된 모델이기 때문에, 정형화된 구조를 가진 표 데이터로부터 답변을 추출하기 위해서는 그 구조를 학습할 필요가 있다. 따라서, [1,2] 연구에서는 1차적으로 대규모 표 데이터와 MLM(Masked Language Modeling) 방법을 추가 입력과 함께 사용해 표 데이터 구조를 BERT에 학습시킨 뒤, 2차로 질의응답 데이터셋을 사용해 표 데이터 구조가 학습된 BERT를 미세조정(Fine-tuning)하여 표 질의응답 모델을 학습했다. 여기서 MLM 학습 방법이란, 입력 문장의 일부 텍스트를 [MASK] 토큰으로 대체한 뒤, 이 [MASK] 토큰이 본래 어떤 텍스트였는지를 예측하도록 학습하는 것이다. 본 연구에서도 [1,2] 연구와 마찬가지로, BERT에 대규모 표 데이터를 사용하여 표 데이터 구조를 학습한 뒤, 표 질의응답 모델 학습에 사용했다.

#### 2.2 프롬프트

프롬프트(Prompt)란, 특정 작업을 원활하게 수행하기 위해 사용되는 명령 텍스트를 말한다[4]. 최근 사전학습된 언어모델을 사용할 때, 입력에 프롬프트를 함께 사용하면 더 높은 성능을 보일 수 있다는 연구 결과들이 있다[4-6]. 이 연구들에서는 모델에 처리하고자 하는 작업에 대한 자세한 설명이나 예시가 포함된 프롬프트를 입력에 함께 사용하여 정답을 유도했다.

정답을 유도하기 위한 방법은 사용되는 모델에 따라

다를 수 있다. 왜냐하면, 모델을 학습한 방법과 추론 방법 등이 달라지면 그 성능이 떨어질 수 있기 때문이다. BERT는 사전학습 단계에서 문맥의 빈칸([MASK] 토큰)에 들어갈 단어를 예측하도록 하였기 때문에, 프롬프트에 기반하여 모델을 학습시킬 때에는 [MASK] 토큰을 포함시킨 뒤, 그 토큰에 들어갈 텍스트가 정답이 될 수 있도록 프롬프트 템플릿을 만들어 사용하곤 한다[4]. 본 연구에서는 표 질의응답의 성능을 시키기 위해 프롬프트를 효과적으로 사용하여 학습할 수 있는 모델을 제안한다.

### 3. 모델

#### 3.1 기본 모델

[1,2] 연구에 따르면, 비정형 텍스트로 학습된 언어모델을 그대로 사용하면 표 데이터에 기반한 문제를 해결하기 어렵다. 따라서, 본 연구에서도 [1,2] 연구 방법과 마찬가지로 사전학습된 언어모델인 BERT에 대규모 표 데이터와 MLM 방법으로 표 데이터 구조를 학습한 뒤, 표 구조가 학습된 BERT를 표 질의응답 모델 학습에 사용했다. 이 때, 표 데이터를 표현하기 위해 TAPAS[1]와 마찬가지로 표 데이터의 행렬 위치값 등을 입력으로 사용했다(그림 1). [표 1]은 [그림 1]의 질의문 예시에 사용된 표 데이터다.

표 1. 한국 지역별 기후 예시

	서울	인천
평균기온(°C)	12.8	12.5
최고기온(°C)	17.4	16.6
최저기온(°C)	8.9	9.1

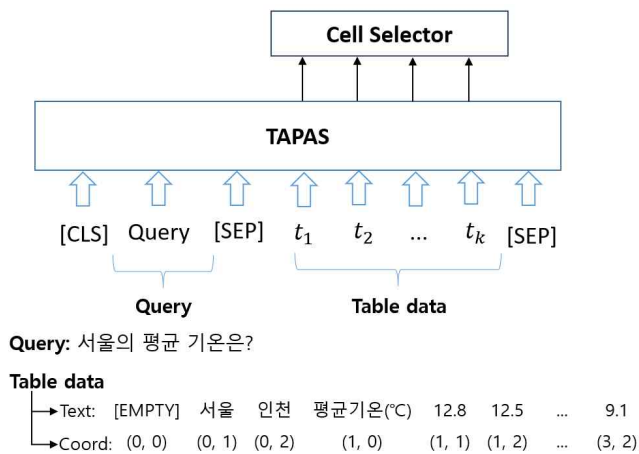


그림 1. 기본 모델(TAPAS)

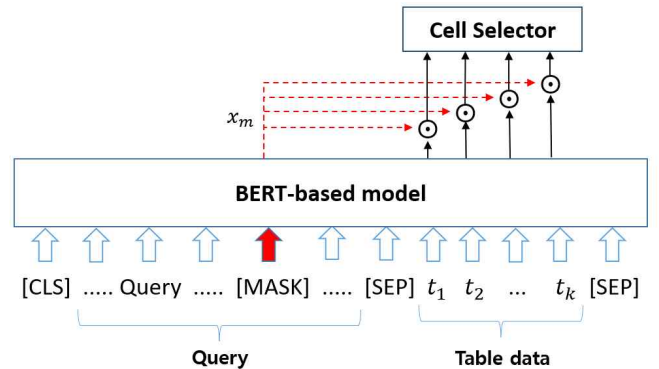
기본 모델에서는 질의문과 함께 1차원 구조로 바꾼 표 데이터를 동시에 입력한 뒤, 표 데이터에서 정답이 될 수 있는 영역을 모두 추출하도록 아래 식 (1)을 사용했다. 즉, 분류기에 정답이 될 수 있는 표 데이터 토큰인

$t_i$ 를 임베딩하고, 그 임베딩 값을 사용하여 정답영역을 분류하는  $s$ 를 학습했다.

$$P(Q, T) = \frac{\exp((t_i)^T s)}{\sum \exp((t_j)^T s)} \quad (1)$$

#### 3.2 제안 모델

기존 BERT 모델에 프롬프트를 사용한 연구들은 대부분 정답 예측을 위해 입력에 포함된 [MASK] 토큰이 임베딩된 결과를 사용했다. 그런데, 이 방법을 질의응답 모델에 그대로 적용한다면, 정답 예측을 위해 모든 표 데이터의 텍스트를 [MASK] 토큰으로 바꾼 뒤 입력으로 사용해야 하는 문제가 있다. 이 경우, 정확한 정보를 모델에 전달하기 어렵다. 이를 해결하기 위해, 식 (1)을 변형한 [그림 2]와 같은 제안 모델을 사용했다.



Query: 서울의 평균 기온은? 정답은 [MASK] 이다.

Table data

- Text: [EMPTY] 서울 인천 평균기온(°C) 12.8 12.5 ... 9.1
- Coord: (0, 0) (0, 1) (0, 2) (1, 0) (1, 1) (1, 2) ... (3, 2)

그림 2. 제안 모델

제안 모델에서 사용한 방법은 다음과 같다. 먼저, 질의문  $Q$ 와 정답이 포함된 표 데이터  $T$ 가 주어졌을 때, BERT의 사전학습 단계인 MLM 학습에서 사용되었던 [MASK] 토큰인  $x_m$ 을 사용한다. 이 때,  $x_m$ 은 정답을 유도할 수 있도록 하는 텍스트 안에 1개만 포함시킨 후, 질의문과 함께 입력으로 사용한다. 그런 뒤, [MASK] 토큰이 임베딩된 값과, 표 데이터의 토큰  $t_i$ 가 임베딩된 값을 아다마르 곱(Hadamard product)한 결과를 사용하여,  $t_i$ 가 정답 영역인 확률을 분류하는  $s$ 를 학습시킨다. 제안 모델에 사용된 수식은 아래 식 (2)와 같다.

$$P(Q, T) = \frac{\exp((t_i \odot x_m)^T s)}{\sum \exp((t_j \odot x_m)^T s)} \quad (2)$$

### 4. 실험 및 결과

#### 4.1 실험 환경

본 연구에서 사용한 표 데이터 사전학습용 데이터는

표 2. 실험 결과

	프롬프트 템플릿	F1(%)		
		A	B	성능 차이 (B-A)
		기본 모델 (중감률)	제안 모델 (중감률)	
1	[Q] [SEP] [T]	67.5 (0.0)	67.5 (0.0)	0.0
2	[Q] 정답은 [MASK] 이다. [SEP] [T]	<b>69.1</b> (+1.6)	69.5 (+2.0)	+0.4
3	[Q] 다음 표에서, 정답은 [MASK] 이다. [SEP] [T]	66.8 (-0.7)	66.7 (-0.8)	-0.1
4	[Q] 다음 테이블에서, 정답은 [MASK] 이다. [SEP] [T]	68.7 (+1.2)	<b>69.7</b> (+2.2)	<b>+1.0</b>

KorWikiTabular 데이터셋[7]을 전처리하여 사용했다. 전처리를 통해 한국어, 영어, 한자, 일본어, 숫자, 일부 특수문자 외에 다른 문자열이 들어간 문장은 모두 제거하고 약 74만개의 표 데이터를 표 데이터 구조 학습에 사용했다. 표 데이터 구조 사전학습에 사용한 언어모델은 BERT-multilingual이다. 이 과정에서, 표의 빈칸 위치에는 “[EMPTY]” 토큰을 모델의 입력으로 사용했다.

표 질의응답 모델의 학습은 KorQuAD 2.0 데이터셋[8]에서 표 데이터에 정답이 있는 질의문-표 데이터 13,063 쌍을 추출하여 사용했다. 이 때, 학습데이터, 개발데이터, 평가데이터는 각각 10,420개, 1,614개, 1,029개의 질의문-표 데이터 쌍이다.

#### 4.2 실험 결과

본 연구에서는 질의문 앞뒤에 다양한 명령어를 추가한 프롬프트를 입력으로 사용하는 기본 모델과 제안 모델을 각각 학습하고 평가했다. 모델 학습에 사용한 프롬프트 템플릿과 그 실험 결과는 [표 2]와 같다. [표 2]에서 [Q]는 질의문을, [T]는 표 데이터를 나타낸다. 이 때, 질의문을 포함한 프롬프트와 표 데이터는 [SEP] 토큰을 사용하여 구분했다.

실험 결과, 2번, 4번 프롬프트 템플릿을 사용했을 때에는 기본 모델에 단순 질의문만 사용했을 때의 성능인 F1 67.5%보다 1.2%p 향상된 F1 69.1%의 성능을 보였다. 이를 통해, 정답을 추출할 수 있도록 유도하는 프롬프트를 입력으로 사용하는 것이 효과적임을 알 수 있다. 또한, 제안 모델과 함께 프롬프트를 사용했을 때에는 단순 질의문만 사용했을 때의 성능 F1 67.5%보다 2.2%p 높은 성능인 F1 69.7%를 달성했다. 즉, 기본 모델에 다양한 프롬프트를 사용했을 때보다 제안 모델에 다양한 프롬프트를 사용하는 것이 더 높은 성능을 보인다. 뿐만 아니라, 같은 프롬프트 템플릿을 사용했을 때에도 기본 모델을 사용했을 때보다 제안 모델을 사용했을 때 0.4%p에서 1.0%p 더 높은 성능을 보여, 제안 모델이 효과적임을 보인다.

이와 달리, 3번 프롬프트 템플릿을 사용했을 때에는 기본 모델과 제안 모델 모두 질의문 [Q]만 사용하는 1번 프롬프트 템플릿을 사용했을 때보다 낮은 성능을 보인다. 게다가, 제안 모델을 사용했을 때가 기본 모델을 사

용했을 때에 비해 낮은 성능을 보이고 있다. 다만, 그 성능 차이는 0.1%p로 작다. 3번 템플릿을 사용했을 때 단순 질의문만 사용했을 때보다 낮은 성능을 보이는 이유는 3번 템플릿과 4번 템플릿을 사용했을 때의 결과 비교를 통해 추측할 수 있다. 3번, 4번 템플릿의 차이는 “표” 또는 “테이블” 단어를 사용한 것이다. 따라서, 모델이 “표” 라는 단어를 정확하게 인식하지 못한다고 판단할 수 있다. 이것은 BERT-multilingual 모델을 사용했기 때문으로 보이는데, BERT-multilingual에서 사용한 학습 데이터에 “표” 보다 “테이블” 이 많이 사용된 것으로 판단된다. 이에 따라, 언어모델의 사전학습 단계에서 자주 사용된 단어를 활용하는 것이 프롬프트를 사용한 모델 학습 결과에 큰 영향을 미친다고 볼 수 있다.

## 5. 결론

본 연구에서는 표 질의응답을 위해 TAPAS의 학습방법으로 BERT 모델을 사전학습한 뒤, 입력으로 다양한 프롬프트를 사용했으며, 이를 위한 모델을 제안했다. 제안한 모델과 다양한 프롬프트를 사용한 결과, 프롬프트만 기본 모델에 추가하고 사용했을 때에는 F1 67.5%에서 F1 69.1%으로 향상되었다. 또한, 프롬프트와 함께 제안한 모델을 사용했을 때에는 F1 69.7%를 달성했는데, 이 성능은 프롬프트만 추가로 사용했을 때보다 높은 성능이다. 따라서, 다양한 프롬프트를 사용하는 것과, 본 연구의 제안 모델을 사용하는 것 모두 표 질의응답에서 효과적이라고 할 수 있다. 향후 연구에서는 좀 더 다양한 프롬프트 템플릿을 사용하고, [9] 연구와 같이 입력 텍스트를 형태소 분석한 뒤 사용해보고자 한다.

### 감사의 글

이 (성과물)은 중소벤처기업부 ‘산업전문인력역량강화사업’의 재원으로 한국산학연협회(AURI)의 지원을 받아 수행된 연구임. (2023년 기업연계형연구개발인력양성사업, 과제번호 : RS-2023-00258172)

### 참고문헌

- [1] Herzig, J., Nowak, P. K., Müller, T., Piccinno, F., and Eisenschlos, J. M. (2020). TaPas: Weakly supervised table parsing via pre-training. arXiv preprint
- [2] Yang, J., Gupta, A., Upadhyay, S., He, L., Goel, R., and Paul, S. (2022). TableFormer: Robust Transformer Modeling for Table-Text Encoding. arXiv preprint arXiv:2203.00274.
- [3] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [4] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1-35.
- [5] Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2021. GPT understands, too. arXiv:2103.10385.
- [6] T, Schick and H, Schutze, "Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference", 2021.
- [7] Jun, C., Choi, J., Sim, M., Kim, H., Jang, H., and Min, K. (2022). Korean-Specific Dataset for Table Question Answering. arXiv preprint arXiv:2201.06223.
- [8] 김영민, 임승영, 이현정, 박소윤, 김명지. (2020). KorQuAD 2.0: 웹문서 기계독해를 위한 한국어 질의 응답 데이터셋. 정보과학회논문지, 47(6), 577-586.
- [9] 박정연, 이동혁, 조경빈, 신형진, 이재성. (2023). 형태소를 이용한 한국어 표 데이터 기계독해. 한국 컴퓨터교육학회 학술발표대회논문집, 27(1), 241-244.