

# PrefixLM에 기반한 한국어 텍스트 요약

이건희<sup>1</sup>, 나승훈<sup>2</sup>, 임준호<sup>3</sup>, 김태형<sup>4</sup>, 최윤수<sup>5</sup>, 장두성<sup>6</sup>

전북대학교<sup>1,2</sup>, 한국전자통신연구원<sup>3</sup>, KT<sup>4,5,6</sup>

# 목차

- Background
- Related works
- Model-PrefixLM
- Experiment setting
- Experiment result
- Conclusion

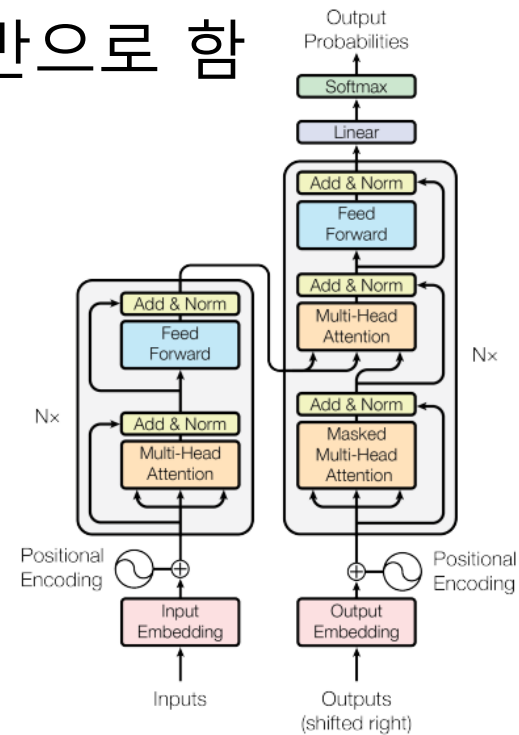
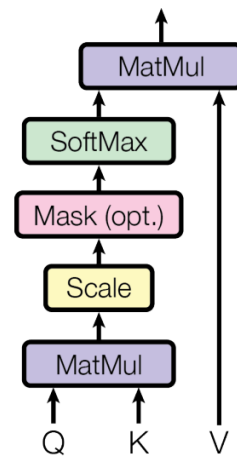
# Background: Summarization Task

- 추출 요약(Extractive Summarization)
  - 요약하고자 하는 문서로부터 구나 절 혹은 문장에 대한 점수를 평가하여 추출해 문서의 요약으로 삼는 방식
  - 이는 추출한 문장들이 조합되면서 가독성이 떨어지고, 원문을 충분히 반영 하지 못하는 등 중요한 정보의 누락이 발생할 수 있음
- 생성 요약(Abstractive Summarization)
  - 원문을 이해하여 그 정보로부터 새로운 문장을 생성하는 방식
  - 생성 방식이기에 원문에 없는 문장을 만들 수 있으며, 잘 학습된 모델은 매끄러운 문장을 생성

# Background: Transformer

- Attention is all you need라는 논문 이름답게 attention만을 통해 획기적인 성능 향상을 이룸
- 이후 대부분의 모델은 Transformer를 기반으로 함

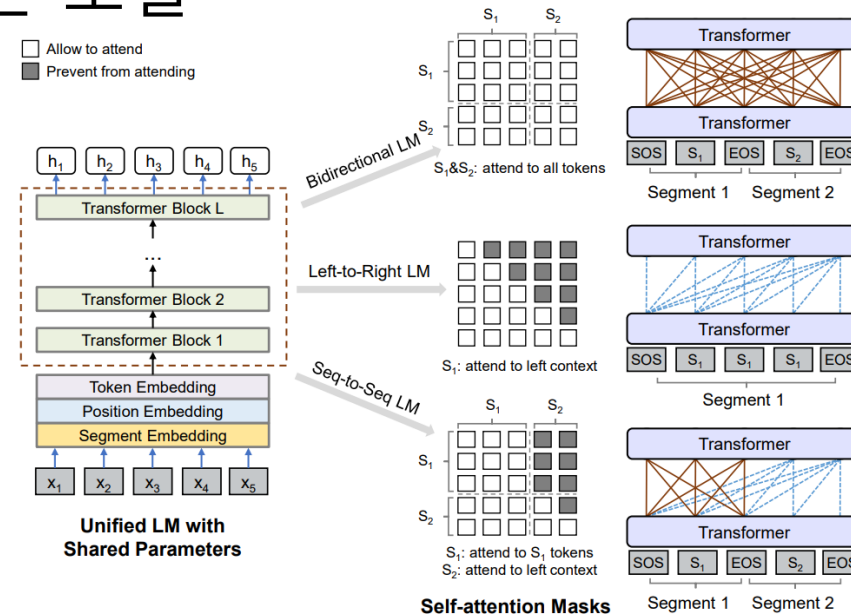
Scaled Dot-Product Attention



A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

# Related works: UniLM

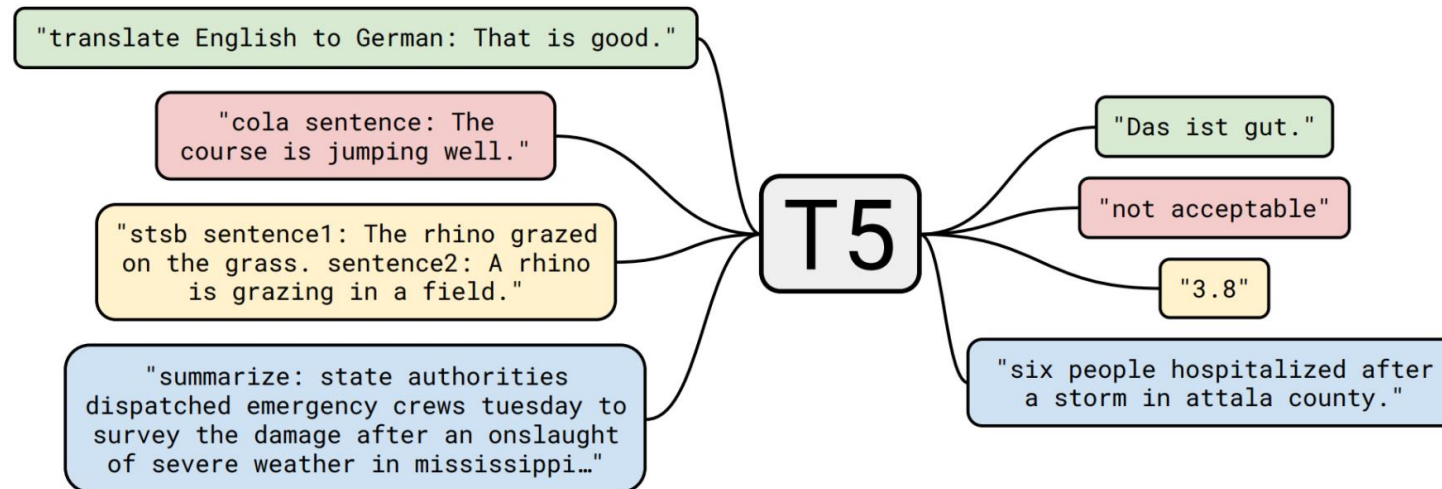
- 단방향 언어 모델(GPT), 양방향 언어 모델(BERT), 인코더-디코더 모델 구조를 일반화한 사전 학습 방법을 제시
- 하나의 사전 학습된 모델로 미세 조정을 통해 언어 이해, 언어 생성 태스크 모두에 사용할 수 있는 모델



L. Dong, N. Yang, W. Wang, F. Wei, X. Liu, Y. Wang, J. Gao, M. Zhou, and H.-W. Hon, "Unified language model pre-training for natural language understanding and generation," 2019.

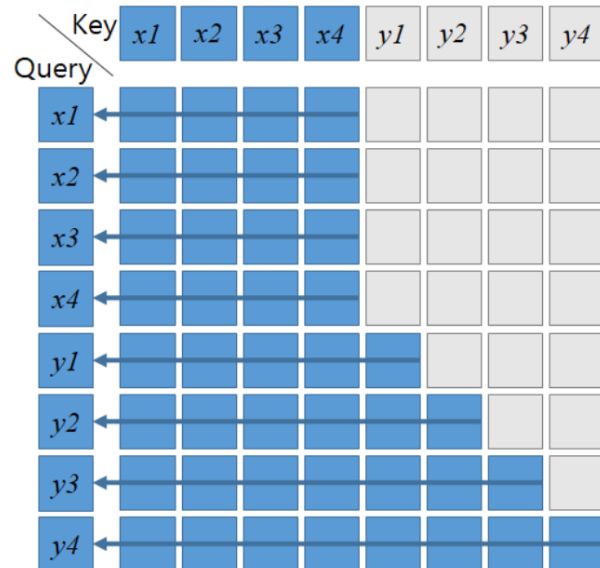
# Related works: T5

- 정답추출, 문서 요약, 등의 여러 태스크들을 인코더-디코더 구조를 사용하는 text-to-text 문제로 치환하는 방법으로 태스크들을 통합하여 언어 이해, 언어 생성 태스크 모두에서 좋은 성능을 달성



# Model-PrefixLM

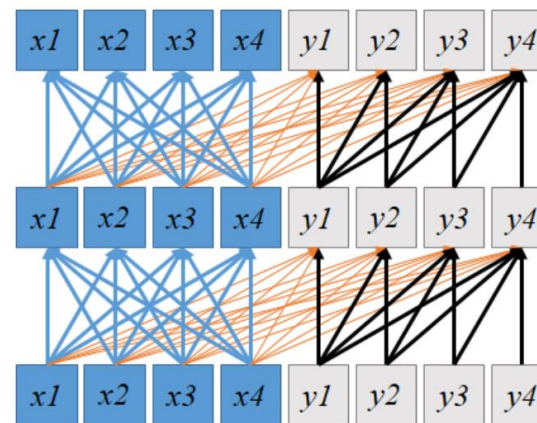
- 기본적으로 BERT와 같은 트랜스포머 인코더 구조
- 인코더 블록의 어텐션 레이어에서 self-attention mask(M)를 변형
  - 어텐션 할 부분은 0으로, 어텐션 하지 않을 부분은  $-\infty$
  - 수식과 같이 어텐션 점수에 M을 더하여 최종 어텐션 점수를 얻어 어텐션 값(A)을 구함



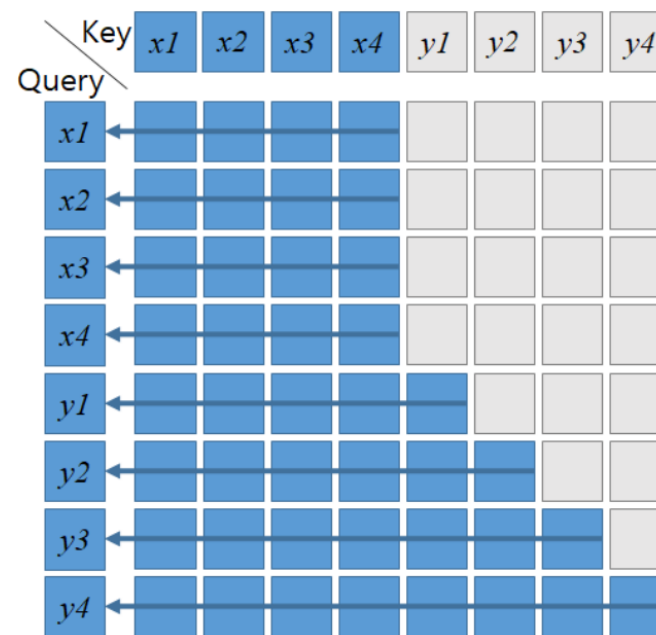
$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V$$

# Model-PrefixLM

- 파란색: 양방향 어텐션, 인코더 역할
- 검은색: 단방향 어텐션, 디코더 역할
- 주황색: 디코더가 인코더의 맥락을 참조



- 이러한 구조를 통해서 인코더-디코더 모델과 같은 역할을 수행할 수 있음





# Experiment setting: Data

- 생성 요약 데이터
  - AIHub 문서 요약 텍스트 - 신문 기사 데이터 사용

집합 타입	데이터 개수	평균 토큰 길이	최대 토큰 길이
학습	260,697	490.58	1,031
평가	10,000	488.92	1,140

```
{  
  "id": "353465974",  
  "media": "충청투데이"  
  "article_original": ["보조금 집행 위법행위·지적사례 늘  
집행·정산 등 운영실태 전반에 대한 자체 감사를 실시할 ;  
사례가 지속적으로 증가함에 따라, 감사를 통해 취약요인  
주무관으로 특별감사반을 편성해 2017년부터 2018년까지  
치를 취할 방침이다.", "시는 지난 15일부터 25일까지 10'  
성 관련 부정수급 유형을 보면 허위·기타 부정한 방법으로  
한 재산에 대해 지자체장의 승인없이 임의 처분 등이 해당  
신고센터(☎850-5031)'를 설치 운영하고 있다.", "지방보조  
해 접수하면 되고, 신고취지와 이유를 기재하고 부정행위  
사를 통해 축제보조금이 제대로 쓰이는지에 대한 반성과  
예산부서 합동으로 컨설팅 위주의 상반기 보조금 특정감,  
  "abstractive": "충주시는 민간보조사업의 증가와 보조  
감사를 진행, 운영실태 전반에 대한 자체 감사를 실시할 ;  
  "extractive": [2, 3, 5]  
}
```

AIHub 문서 요약 텍스트 <https://aihub.or.kr/aidata/8054>

# Experiment setting: Baseline

- 모델: BART, T5
- 사전 학습 데이터: 한국어 위키피디아
- 미세 조정 데이터: AIHub 문서 요약 텍스트 - 신문 기사 - 학습 데이터
- 미세 조정
  - BART: [SOS] doc [EOS]
  - T5: [SOS] doc [EOS] 앞 글을 요약하시오 [EOS]

# Experiment setting: PrefixLM pre-training

- 모델: PrefixLM
  - 초기화 방식, 사전 학습 여부 조합 3종
- 사전 학습 데이터: 한국어 위키피디아

	초기화	사전 학습
모델1	RoBERTa	700MB
모델2	random	700MB
모델3	RoBERTa	X

# Experiment setting: PrefixLM fine-tuning

- 미세 조정 데이터: AIHub 문서 요약 텍스트 - 신문 기사 - 학습 데이터
- 미세 조정
  - [SOS] doc [EOS]
  - [SOS] doc [EOS] 앞 글을 요약하시오 [EOS]
  - [SOS] 이 글을 요약하시오 [EOS] doc [EOS]

# Experiment result: Abstractive Summarization

	초기화	사전 학습
모델1	RoBERTa	700MB
모델2	random	700MB
모델3	RoBERTa	X

	마스크 확률	Prefix	Rouge F1		
			1	2	L
BART	-	X	0.5003	0.3364	0.4191
T5	-	Front	0.3945	0.2340	0.3545
모델1-1	1.0	X	0.4040	0.1113	0.3258
모델1-2	0.7	X	0.5335	0.3335	0.4409
모델1-3	0.7	Front	0.5383	0.3398	0.4446
모델1-4	0.7	Back	0.5387	0.3406	0.4458
모델2-1	0.7	Front	0.5325	0.3335	0.4404
모델2-2	0.7	Back	0.5323	0.3336	0.4412
모델3-1	0.7	Front	0.5375	0.3375	0.4420

## Baseline 비교

같은 데이터로 사전 학습된 모델2가 Baseline(BART)보다 0.02 높은 성능을 보임

# Experiment result: Abstractive Summarization

	초기화	사전 학습
모델1	RoBERTa	700MB
모델2	random	700MB
모델3	RoBERTa	X

	마스크 확률	Prefix	Rouge F1		
			1	2	L
BART	-	X	0.5003	0.3364	0.4191
T5	-	Front	0.3945	0.2340	0.3545
모델1-1	1.0	X	0.4040	0.1113	0.3258
모델1-2	0.7	X	0.5335	0.3335	0.4409
모델1-3	0.7	Front	0.5383	0.3398	0.4446
모델1-4	0.7	Back	0.5387	0.3406	0.4458
모델2-1	0.7	Front	0.5325	0.3335	0.4404
모델2-2	0.7	Back	0.5323	0.3336	0.4412
모델3-1	0.7	Front	0.5375	0.3375	0.4420

초기화 비교

RoBERTa로 초기화하면 random 초기화보다 0.004 증가

# Experiment result: Abstractive Summarization

	초기화	사전 학습
모델1	RoBERTa	700MB
모델2	random	700MB
모델3	RoBERTa	X

	마스크 확률	Prefix	Rouge F1		
			1	2	L
BART	-	X	0.5003	0.3364	0.4191
T5	-	Front	0.3945	0.2340	0.3545
모델1-1	1.0	X	0.4040	0.1113	0.3258
모델1-2	0.7	X	0.5335	0.3335	0.4409
모델1-3	0.7	Front	0.5383	0.3398	0.4446
모델1-4	0.7	Back	0.5387	0.3406	0.4458
모델2-1	0.7	Front	0.5325	0.3335	0.4404
모델2-2	0.7	Back	0.5323	0.3336	0.4412
모델3-1	0.7	Front	0.5375	0.3375	0.4420

사전 학습 여부 비교

RoBERTa 초기화 후 사전 학습을 한 경우 0.0026 증가

# Experiment result: Abstractive Summarization

	초기화	사전 학습
모델1	RoBERTa	700MB
모델2	random	700MB
모델3	RoBERTa	X

	마스크 확률	Prefix	Rouge F1		
			1	2	L
BART	-	X	0.5003	0.3364	0.4191
T5	-	Front	0.3945	0.2340	0.3545
모델1-1	1.0	X	0.4040	0.1113	0.3258
모델1-2	0.7	X	0.5335	0.3335	0.4409
모델1-3	0.7	Front	0.5383	0.3398	0.4446
모델1-4	0.7	Back	0.5387	0.3406	0.4458
모델2-1	0.7	Front	0.5325	0.3335	0.4404
모델2-2	0.7	Back	0.5323	0.3336	0.4412
모델3-1	0.7	Front	0.5375	0.3375	0.4420

프롬프트 사용 여부 비교

프롬프트를 사용한 경우 0.005 가량 증가



# Conclusion

- **결론**

인코더, 디코더 단일 구조인 PrefixLM 구조는 한국어 생성 요약에서 BART, T5와 비교하여 성능 향상을 보였으며, 프롬프트 사용을 통한 성능 향상을 확인하여 프롬프트 사용이 유의미함을 확인함

- **향후 연구**

PrefixLM 구조를 확장하여 생성 요약뿐 아니라 여러 태스크에 적용해 볼 예정임