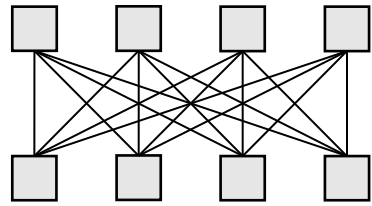


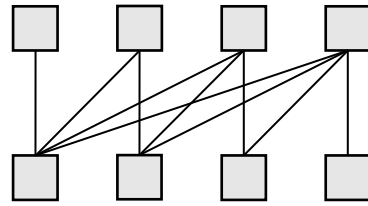
# 점진적 Fully-Visible 어텐션에 기반한 PrefixLM 확장 및 텍스트 생성 응용

이건희, 나승훈, 임준호, 김태형, 최윤수, 장두성  
전북대학교, 한국전자통신연구원, KT

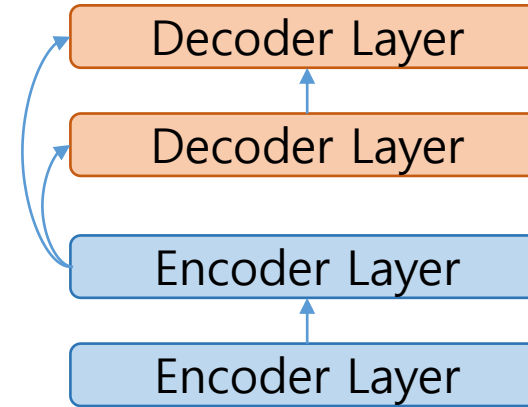
# 서론



양방향 어텐션



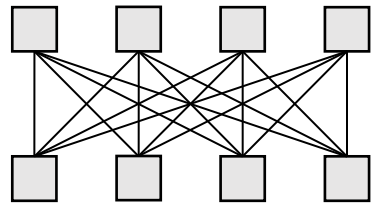
단방향 어텐션



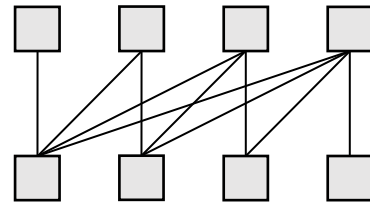
- Transformer

	역할	어텐션	학습	추론
Encoder	입력을 인코딩	양방향	동시입력	동시입력
Decoder	입력을 토대로 다음 토큰 추론	단방향	동시입력	순차입력

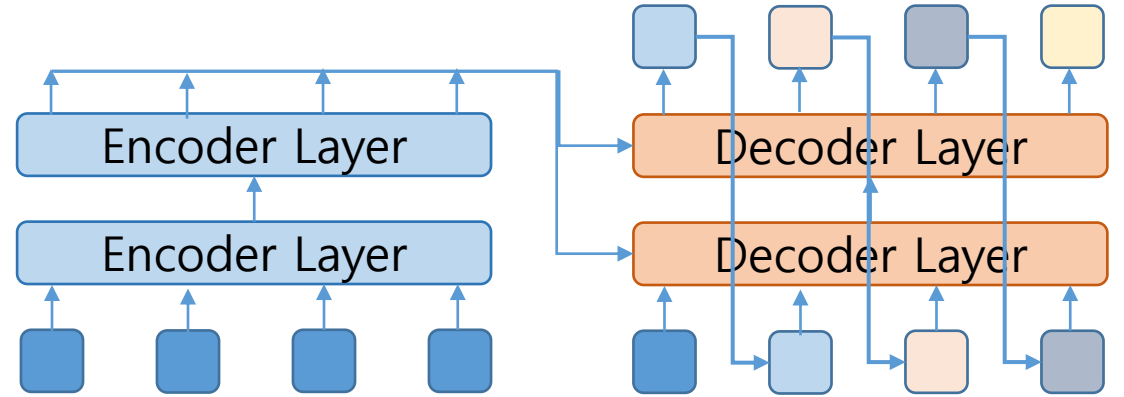
# 서론



양방향 어텐션

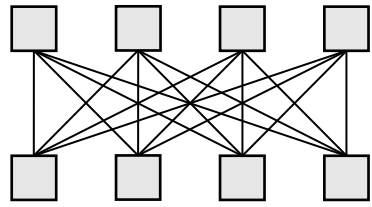


단방향 어텐션

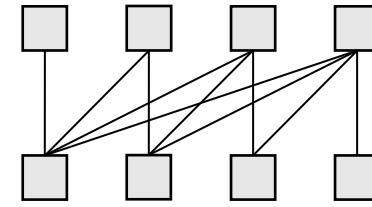
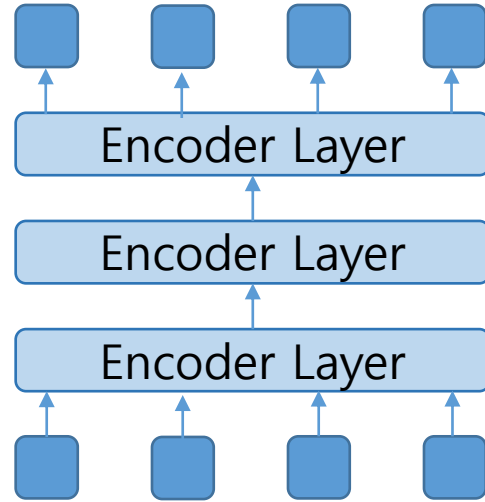


- Encoder-Decoder
- Encoder로 입력을 인코딩하여 Decoder의 추론에 참조

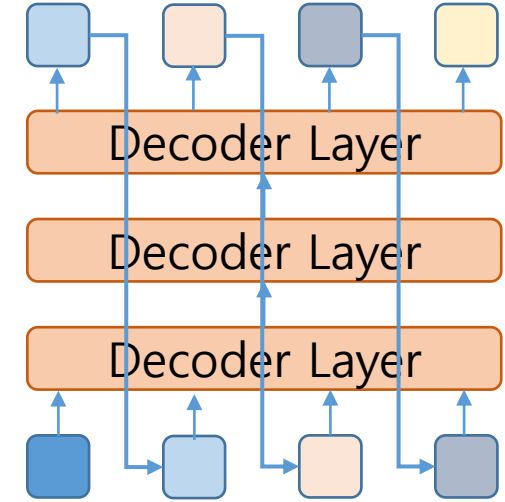
# 서론



양방향 어텐션



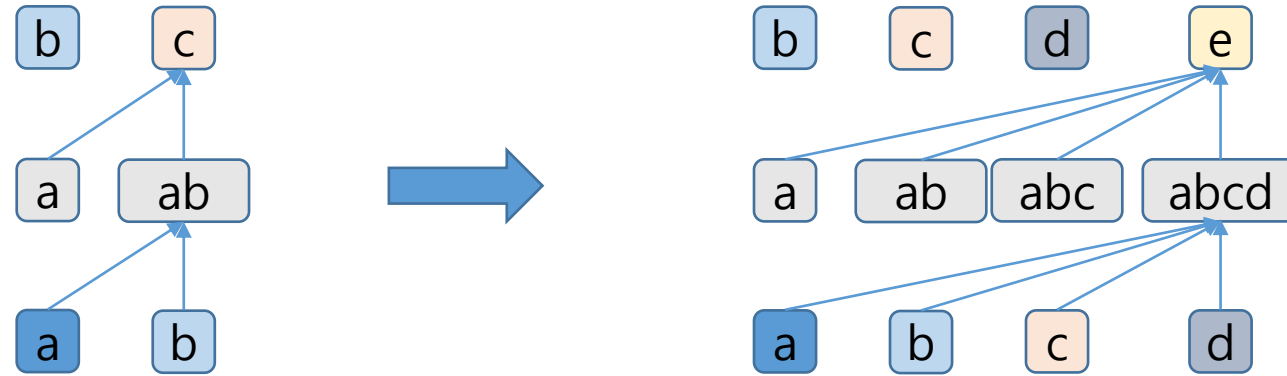
단방향 어텐션



- Encoder
- BERT, RoBERTa, SpanBERT...
- 언어 이해 태스크에 강점

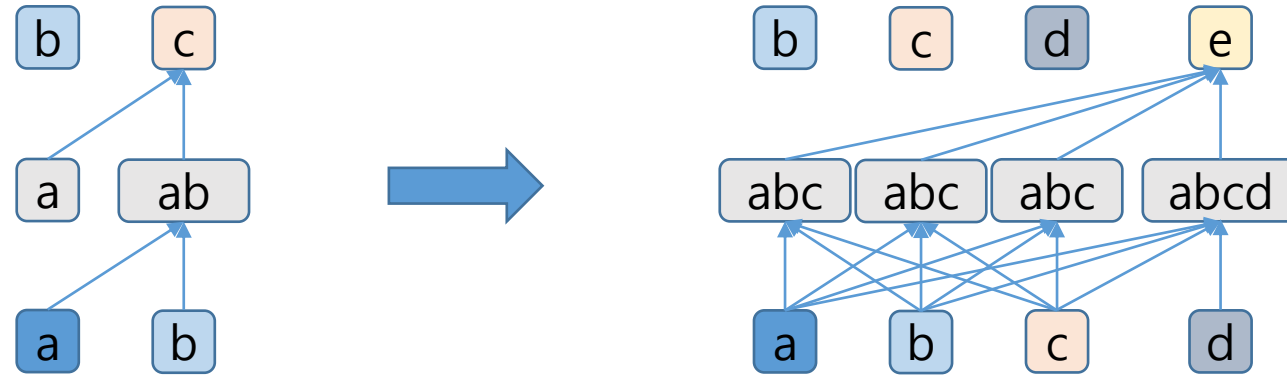
- Decoder
- GPT1,2,3...
- 언어 생성 태스크에 강점

# 제안



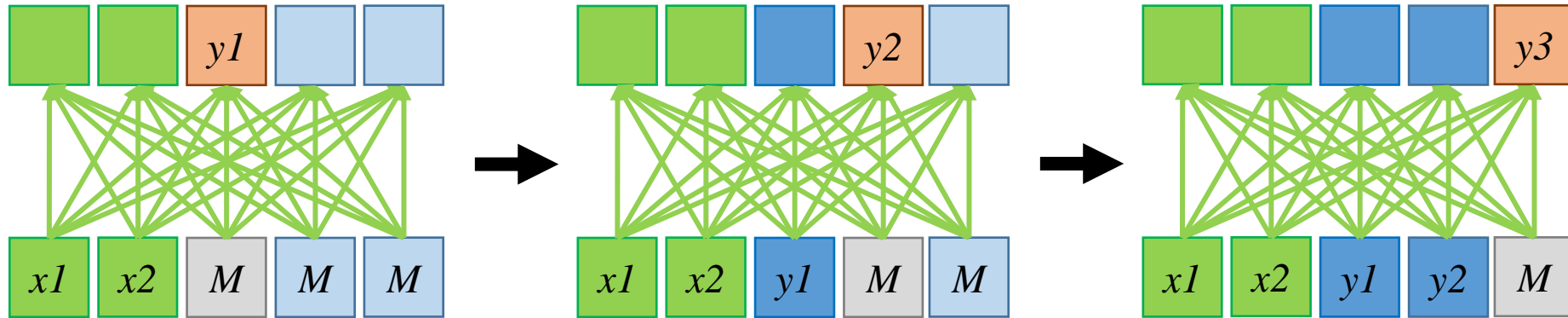
- e를 추론하는 과정에서 이미 생성된 b c d의 은닉 상태는 각 토큰의 이후 토큰은 반영하지 않음

# 제안



- 다음 토큰을 추론할 때 좀 더 풍부한 표현을 얻기 위해 이전 토큰들의 은닉상태를 업데이트하는 실험

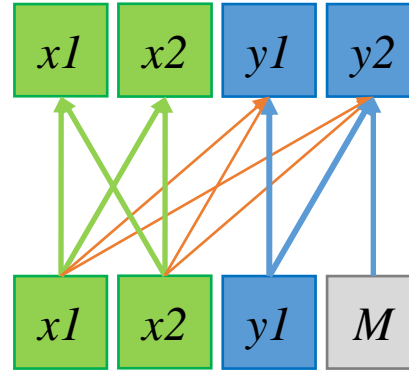
# 관련연구



- PMLM

- BERT를 변형하여 생성 모델로서 작동하도록 하는 연구
- 기본적으로 마스크로 채운 뒤 특정 순서로 하나씩 추론하여 치환

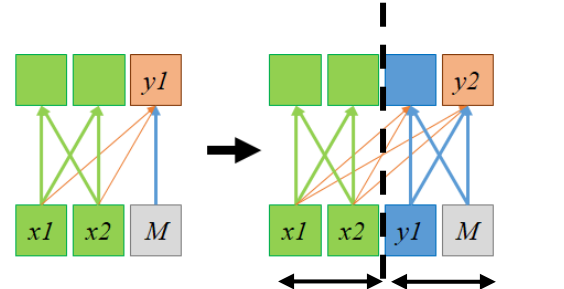
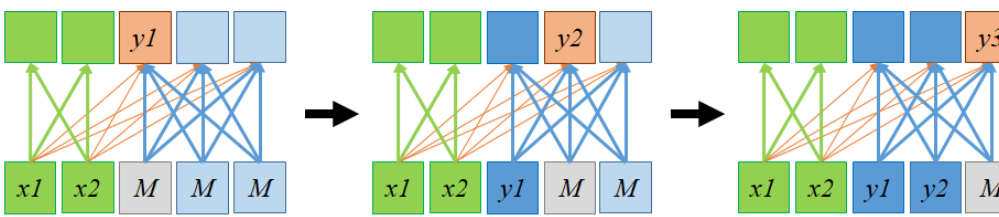
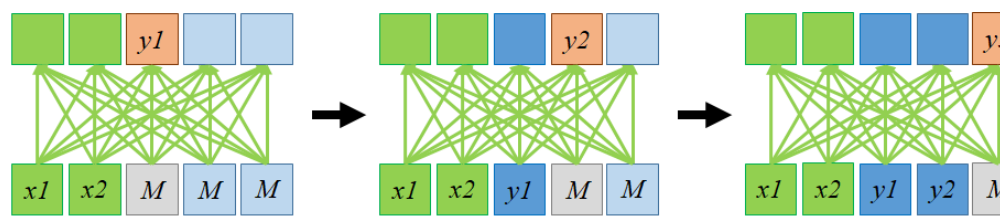
# PrefixLM



- Encoder와 Decoder 역할을 동시에 수행하는 모델
- 앞부분은 양방향 어텐션
- 뒷부분은 단방향 어텐션



# 모델 & 디코딩 방식

디코딩 방식	Sep-bi-dir	Full-bi-dir
Incremental 생성할 마스크까지만	 <p style="text-align: center;">인코더 부분 : 디코더 부분 4 : 1</p>	
PMLM 전체 마스크 (고정길이)		

Sep 구조는 디코딩 시 인코더 부분(80%)은 한번만 계산하므로 Full 방식에 비해 속도가 빠름  
 1만 샘플 512 토큰 기준 55m vs 3h40m

# 생성 요약 데이터

표 1: 생성 요약 학습 및 평가 데이터 집합 구성

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

- AIHub 문서 요약 텍스트 내 신문 기사 생성 요약 데이터

```
"article_original": ["그룹 위키미키(Weki Meki)가 드디어 오늘(14일) 두 번째 싱글 앨범을 발표하며",  
"article_original": ["대구 남구청이 지역민의 매로사항은 효율적으로 관리·해결하고자 내년부터 '고충민",  
"article_original": ["세종시청 전경.", "세종시 연동면장과 도담동장으로 황미라 사무관과 양진복 사무",  
"article_original": ["군포시는 지난 25일 열린 '2019 시정 인식조사 보고회'에서 시민 10명 중 8명 이",  
"article_original": ["박효상 소규모 공사 127건 신속 처리 청도군은 올해 소규모 주민숙원사업에 대한",  
", "article_original": ["향수를 만드는 지휘자 파비오 루이지가 KBS교향악단의 12월 특별연주회 지휘를",  
"article_original": ["서울시 주요 공공시설물 노후화가 심각한 수준인 것으로 나타났다.", "하수도 관",  
"article_original": ["사드 핑계로...", "포항 기업유치 빨간불사드 핑계로...", "포항시가 야심차게",  
"article_original": ["지난 21일 세종시에서 개최된 전국 농어촌지역 군수협의회 회장단 8개 지역 군수",  
"article_original": ["[이데일리 성주원 기자] 한국교통안전공단은 한국어린이안전재단과 함께 7세 이하",  
"article_original": ["'백통신원리조트' 등 대규모 개발 사업에 대한 연장·변경계획안이 개발 사업 심의",  
"article_original": [""우석대학교(총장 장영달)가 교육부와 한국대학교육협의회와 주관하는 '2019 고교",  
", "article_original": ["코스닥 상장사 지엠피가 자회사와 합병을 계기로 새로운 기업으로 도약할 것으로",  
1" "article_original": ["2021 한국오라세계수업선생수권대회 조직위원회 관계자들이 15일 오전, 대전, 주
```

```
{'media': '당진시대',  
'id': '252106905',  
'article_original': ['여기구 국회의원 정책간담',  
'여기구 국회의원이 천연가스의 안정적 수급을',  
'여 의원은 지난 17일 '천연가스 장기수급 전망',  
대해 논의했다.',  
'여기구 의원은 "석탄화력발전은 미세먼지, 온실',  
'이며 "석탄화력발전과 원전에 의존하는 현재 정',  
'이날 간담회에서 한국가스공사 김치만 수급관리',  
발전 증가로 수요증가폭이 둔화됐다"며 "특히 현행',  
'이며 발표에 나선 에너지경제연구원 이호무 연',  
판단된다"고 연구결과를 발표했다.',  
'이에 대해 여기구 의원은 "제12차 장기 천연가',  
대로 차질 없이 진행되어야 한다"고 당부했다.'],  
'abstractive': '여기구 국회의원이 천연가스',  
년 착공 예정된 액화천연가스 제5기지 건설을 차질',  
'extractive': [1, 2, 7]}
```

# 성능 비교

표 3: 생성요약에 대한 성능평가

	Rouge F1			고정 길이
	1	2	L	
BART	50.03	33.64	41.91	-
T5	49.58	33.30	41.34	-
PrefixLM	53.22	33.36	44.12	-
Full-Parallel	52.60	32.15	41.54	102
Sep-Parallel	52.80	32.08	42.14	102
Sep-Parallel	53.62	33.24	43.62	75
RoPrefixLM	53.75	33.75	44.20	-
Full-Incr	53.70	33.73	44.09	-

- 평가 방법

- Rouge-score[13] rouge-1, rouge-2, rouge-L F1 점수
- 최장일치를 평가하는 rouge-L을 기준

# 성능 비교

표 3: 생성요약에 대한 성능평가

	Rouge F1			고정 길이
	1	2	L	
BART	50.03	33.64	41.91	-
T5	49.58	33.30	41.34	-
PrefixLM	53.22	33.36	44.12	-
Full-Parallel	52.60	32.15	41.54	102
Sep-Parallel	52.80	32.08	42.14	102
Sep-Parallel	53.62	33.24	43.62	75
RoPrefixLM	53.75	33.75	44.20	-
Full-Incr	53.70	33.73	44.09	-

- Full-Parallel 모델과 Sep-Parallel 모델은 미리 고정 길이를 정해 주어야 함
- 정답 요약의 평균 토큰 수는 약 70개. 비슷한 길이인 75 토큰으로 평가 시 1.5의 성능향상

# 성능 비교

표 3: 생성요약에 대한 성능평가

	Rouge F1			고정 길이
	1	2	L	
BART	50.03	33.64	41.91	-
T5	49.58	33.30	41.34	-
PrefixLM	53.22	33.36	44.12	-
Full-Parallel	52.60	32.15	41.54	102
Sep-Parallel	52.80	32.08	42.14	102
Sep-Parallel	53.62	33.24	43.62	75
RoPrefixLM	53.75	33.75	44.20	-
Full-Incr	53.70	33.73	44.09	-

- Incremental 한 디코딩을 한 Full-Incr 모델은 비교 대상인 RoPrefixLM과 0.11의 큰 차이 없는 결과를 보임

# 결론

- 인코더와 디코더가 통합된 PrefixLM 구조에서 양방향 어텐션을 사용한 디코딩 성능을 생성 요약 태스크에서 비교 하였다.
- 같은 적은 데이터로 학습된 BART, T5보다는 좋은 성능을 보여 충분히 유효한 방식임을 보였지만 기반 구조인 PrefixLM과는 비슷한 성능을 보였다.

감사합니다