**<https://medium.com/@kabbi159/acl-2020-bart-denoising-sequence-to-sequence-pre-training-for-natural-language-generation-7a0ae37109dc>**

**BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation and Comprehension(ACL 2020)**

**Abstract + Introduction**

보완점: AE와 AR이 각각 어디에 강한지 알고 있기

BART는 Bidirectional Transformer와 AR Transformer가 결합된 방식으로  
BERT처럼 denoising autoencoder 방식으로 사전학습이 기반이다.

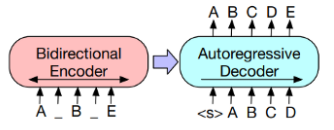
즉 기존 MLM을 변형시킨다.

다양한 noising 접근법으로 실험을 했는데 문장의 순서를 랜덤으로 섞고 최신 in-filling 기법을 사용하는 것이 가장 좋은 성능을 보임.

In-filling: 임의 길이의 text span을 하나의 mask token으로 바꿈.

RoBERTa와 동일한 환경에서 abstractive dialogue, QA, summarization의 Text Generation, Reading Comprehension 류의 작업에서 강했다(SOTA)





**Model**

**Architecture**

Activation Function을 ReLU 대신 GeLUs를 사용했고  
6-6(Base), 12-12(Large)의 인코더-디코더 구조로 이루어진다.

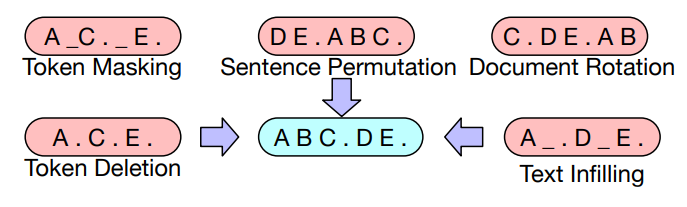
또한 BERT와 달리 디코더에 cross-attention이 있고, BERT는 최종 prediction에 FFNN 사용하지만 BART는 그렇지 않다(인코더에서 이미 있어서 디코더에는 필요 없나??)

같은 수의 레이어일 때, BERT보다 parameters가 10% 많다.

**Pre-Training BART**

기본적으로 원본 문서와 출력 간 cross-entropy는 사용

기존의 denoising autoencoder들은 특정 noising 스키마에서만 한정돼 있지만, BART는 어떠한 문서 corruption이든 가능 (왜 그런지 모르겠다 BERT도 가능한 건 매한가지 아닌가)

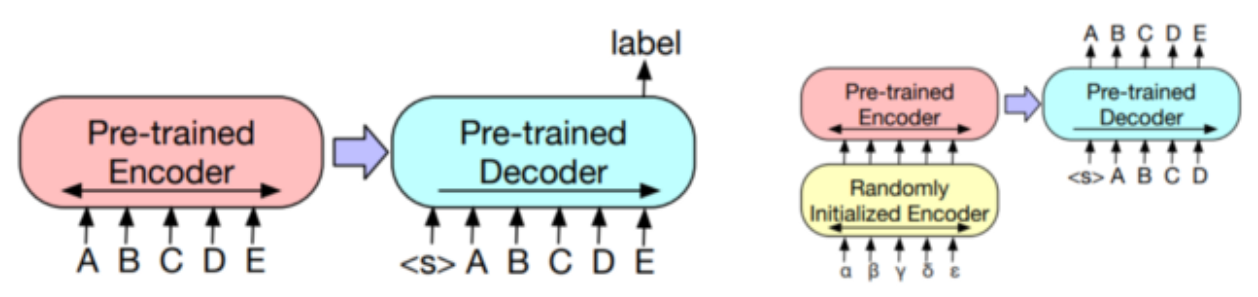


1. Token Masking: BERT 방식, [MASK] 토큰이 원래 뭐였는지 맞춤
2. Token Deletion: 랜덤으로 토큰 삭제, 어느 위치에 어떤 누락된 Token인지 맞춤
3. Text Infilling: 포아송 분포에서 구한 span 길이로 여러 text span 샘플링하고, span(이어지는 tokens) 단위로 [MASK], 원래 어떤 span 일지, 몇 개의 Tokens 였을지 맞춤
4. Sentence Permutation: 문장 단위로 섞고 복구
5. Document Rotation: 토큰 단위로 섞고 복구

이중 Text Infilling이 강했다.

Fine-Tuning BART(Fine-Tuning이라 입력을 인코더 디코더에 둘 다 넣는 경우가 있다, Pre-Training 시에는 인코더에만)

1. Sequence Classification Tasks: 인코더와 디코더에 같은 입력 -> 디코더의 마지막 hidden state를 새로운 Linear Layer에, BERT에서 CLS 토큰을 넣는 것과 유사.  
   이를 위해 스페셜 END 토큰을 추가
2. Token Classification Tasks: 완전한 문서를 인코더와 디코더에 넣음. 디코더 맨 위 hidden state가 각 단어의 representation -> 분류
3. Sequence Generation Tasks: 인코더의 입력으로 sequence, 디코더에서 출력 생성, 기존 AR 방식으로 진행
4. Machine Translation: 임베딩 레이어를 랜덤으로 초기화 -> 외국어를 영어로 de-noise 가능한 입력으로 만든다. (무슨 소리지??) 아래 그림이..



Results는 다시 보기