ELECTRA는 ICIR 2020에서 구글 리서치 팀이 새로운 pre-training 기법을 적용한 language model 입니다. 기존 BERT를 비롯한 많은 language model들은 Input을 mask token으로 변환하고 이를 원본 token으로 복원하는 MLM(masked language modeling) 태스크를 통해 pre-training을 합니다.

하지만 이런 모델은은 학습 시 상당히 많은 계산량을 필요로 하기에 충분한 컴퓨팅 파워가 없으면 모델 연구에 어려움이 존재하고, 이를 해결하고자 ELECTRA 모델은 모델의 정확도와 함께 학습의 효율성에 집중한 모델입니다.

이를 위해 Replaced Token Detection(RTD) 라는 새로운 pre-training 태스크를 제안하였습니다. 결과적으로 기존 BERT 모델과 동일한 조건에서 성능을 능가하고, Large 모델의 경우 RoBERTa나 XLNet 대비 1/4의 계산량만으로 비슷한 성능에 도달함을 보였습니다.

RTD 태스크에 간략하게 설명하자면 Generator를 통해 실제 Input의 일부 token을 그럴싸한 가짜 token으로 바꾼 후, 각 token이 실제 Input에 존재하는 original token인지 replaced token인지 이진 분류로 판별하는 방법으로 학습하는 방법입니다.

이는 기존 MLM 태스크가 가진 문제점인 전체 token 중 15%에 대해서만 loss가 발생하고 이로인해 학습에 많은 비용이 발생한다는 점과 또한 학습 때는 [mask] 토큰을 모델이 참고하여 예측하지만 실제 inference에는 [mask] 토큰이 존재하지 않는 문제가 있습니다. 이와 같은 문제를 해결하고자 RTD라는 새로운 pre-training 태스크를 제안하여 기존 BERT보다 훨씬 빠르게 학습할 수 있으며 downstream task에서의 성능이 높게 나타나 해당 사전학습 모델을 채택하였습니다.