

제목 : KR-BERT: A Small Scale Korean Specific Language Model

1.연구 배경 및 목적

다중 언어 BERT 모델에서 놓칠 수 있는 한국어의 특수한 언어 현상을 포착하고 이를 해결하기 위해 KR-BERT 모델을 제안했음.

2.주요 기술 및 결과

- 토큰나이저 개선:

BidirectionalWordPiece 토큰나이저를 포함한 여러 토큰나이저를 테스트하여 최적의 토큰나이저를 선정함.

토큰화를 위한 최소 토큰 범위를 하위 문자 수준에서 문자 수준까지 조정하여 모델에 더 나은 어휘를 구축함.

- 성능 향상:

미세 조정을 통해 KR-BERT 모델은 기존 모델보다 성능이 비슷하거나 더 했고, 훨씬 작은 규모의 말뭉치를 사용하여 훈련.