**schuBERT: Optimizing Elements of BERT(ACL 2020)**

거대한 파라미터 -> 서비스에 바로 적용 무리 -> 경량화 방법 많음(Knowledge Distillation, Pruning, Quantization) 여기서는 Pruning 이용

**Main Idea**

트랜스포머 인코더 이용 = BERT

아직 BERT가 하이퍼 파라미터 최적화는 덜 됐다고 생각 => 하이퍼 파라미터 더 세분화하여 모델 구조 수 늘림 & 프루닝 기반 기법 => 제한된 파라미터 크기 내에서 구조 자체를 최적화

결과적으로 동일한 파라미터 수에서 BERT-base보다 우수