<https://vanche.github.io/spanbert_roberta/>

새로운 loss function을 제안하며 BERT의 성능을 크게 향상한다.

SpanBERT: Improving Pre-Training by Representing and Predicting Spans(TACL 2020)  
TACL은 저널인 듯 싶다

Span selection과 관련 있는 tasks(QA, coreference resolution task) 뿐만 아니라 GLUE에서도 task 성능 향상을 보임

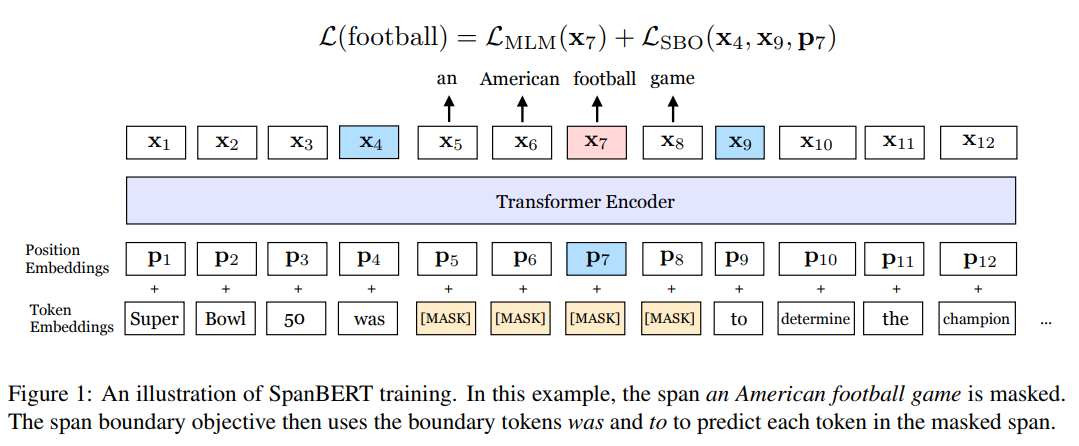
많은 NLP Tasks들이 두 개 이상의 span 사이의 관계에 대한 추론을 포함  
ex) “Which NFL team won Super Bowl50?” -> “Broncos” 라는 단어를 안 뒤 “Denver”를 유추, “Denver Broncos”를 바로 유추하기는 어렵다.(??)

SpanBERT는 BERT에 random continuous span masking & masking된 span 전체를 그 경계의 토큰에서 유추하는 SBO(Span Boundary Objectives) 도입(??)

SBO는 span-level의 정보를 boundary tokens에 저장(??)

Single segments를 pretraining & NSP를 제외 -> 성능 향상

**Model**



위와 같이 token 단위가 아닌 span 단위로 masking한다.(random continuous span masking)

또한, span boundary tokens인 was와 to를 span tokens 예측에 사용한다.