@Transformer

각각의 단어의 값을 벡터로 변환하여 self-attention에 넣음

각각의 단어로 나온 어텐션 스코어를 피드포워드에 넣음

모든 것을 다 체크해서 이 단어가 어떤 단어와 가장 강한 어텐션을 가지는 지

스코어를 쿼리, 키, 벨류 세 개의 값으로 어텐션 스코어를 얻음

쿼리와 키를 비교해선 얻은 값이 벨류

벡터에서 내적을 시키면 얼마나 유사한지 유사도가 계산됨 (dot-product)

소프트맥스를 취해서 확률값으로 바꿈

자기 자신과 자신을 내적한 값이 가장 클 것

쿼리 - 비교하는 단어 (타켓 랭귀지)

키 - 비교되는 단어ㅏ (소스 랭귀지)

벨류 - (소스 랭귀지ㅣ)

쿼리와 키를 곱하는 것은 유사도를 구하는 것

K의 길이에 영향을 받지않도록

Concat할 때

**원래는 머신트렌스레이션에서 나온 것**

**입력 불어, 출력 영어의 경우 영어에서 쿼리를 취하고, 불어에서 키, 벨류를 취함**

#positional encoding

* 포지셔널 정보를 넣어주기 위함

자신 이외의 값은 어텐션을 하지 못하게끔

틀리면 weight가 조정됨

8층을 쌓아서 컨캣하는 것이 성능 향상에 무슨 영향이 있는지?

할 때마다 웨이트가 다 다르기 때문

sequential하지 않고, pararell하게 처리할 수 있음

레이어별로 비주얼라이즈를 할 수 있어서 추적이 가능함 (버트 비즈 라는 툴이 있음)