

트랜스포머 재욱리 강의

rnn : 줄거리 벡터를 만들지 않고 끝

seq2seq : 줄거리 벡터를 만들다

- 에이포 용지 한장 하나를 보다
- 데이터 많거나 긴 문장의 경우, 병목현상 발생
 - 디테일한 걸 잡지 못하더라

attention

- 떡밥을 잘 캐치하기 위해 등장
- 내적 dot product
 - $a \text{ 벡터} * b \text{ 벡터} = a \text{ 크기} * b \text{ 크기} * \cos \text{ 세타}$
 - 막대기가 같은 곳을 바라보다 == 1
 - 중요한 정보일수록 1에 가까운 좁은 사잇각이 나오다
- 그때 그때 인코더의 h를 다 보다(내적을 하다)
- 모든화를 똑같은 중요도로 보지않고
 - **현재 데이터를 토대로 내적을 했더니 더 중요한 정보들에 집중을 하겠다**

transformer - positioning encoder

- 1화, 2화, 각자 보라고 시키는데 1화인지 2화인지 순서 파악이 안된다
- 넘겨줄때부터 1화야. 2화야. 하고 적어준다
- 그리고 decoder에서 attention하면 되더라

transformer

- key query value
 - query : 3화 예측해줘
 - key : 1화의 값, 2화의 값
 - value : 1화의 내용, 2화의 내용

softmax

- 확률로 바뀌주는 활성화 함수
- 내적을하면 $-1 \sim 1$ 사이가 나옴
 - 지수함수를 써서 (0~양수)
 - 0~1 확률로 바뀌주기

self attention

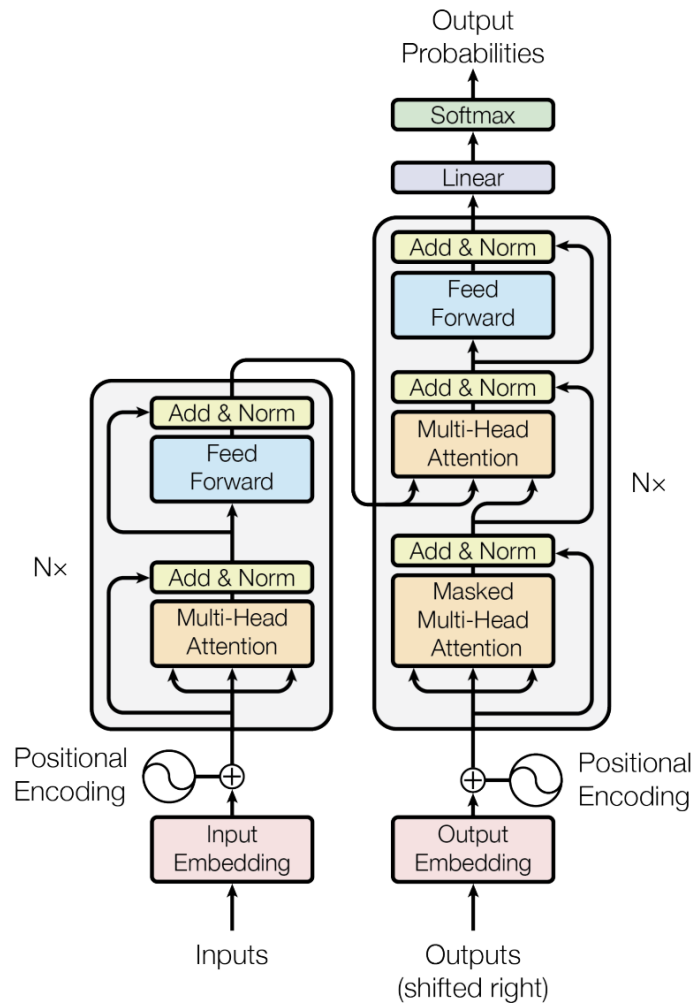


Figure 1: The Transformer - model architecture.

positional encoding

- 위치 정보 저장 (순서를 알리는)

multi head attention

- 1화랑 2화랑,
 - 2화랑 4화랑
 - 3화랑 5화랑

- **여러 관계를 다 보겠다! N^2**

add & norm

- norm : 벡터의 크기
 - 벡터 n개 있으면 다 제곱해서 루트 씌어서 더하는 것

feed forward

- RNN
 - 친구 등에 그림그려서 전달(레크레이션 시간!)
 - 혜준이가 그휘에게, 근휘가 한나에게
- FEED FORWARD
 - 근데 근휘에게 재욱이 달려가서 다시 그려주고, 원본 데이터도 같이 전달(혜준이가 근휘에게, 근휘가 혜준정보와 함께 한나에게)
 - 모든 관계성을 다 본다
 - $f(x) = f(x_{n-1}) + x_{n-1}$ (시그마 포함)

bert

- **encoder만 여러 붙인 것!**

gpt

- decoder만 여러개 붙은 것!

weighted sum

- n개의 벡터 중요도를 1로 만들다
 - encoding을 하면 차원 갯수가 어마하게 늘어남
- 닭갈비 레시피 비율
- attention에서 weight값 확률을 다 더해 1로 만드는 행위 자체를 말함
- softmax 돌리면 그게 weighted sum

self attention ☆

- 문맥 파악 위해 등장
- 내적을 하면 스칼라(숫자 하나)가 나오는데
- **자기가 자기자신 내적**
- **문맥을 가지고 있는 벡터가 나옴!**
 - I에 대한 벡터(문맥까지 포함하고 있는)
 - I와 모든 단어들 관계(문맥)을 가지고 있다
 - am에 대한 벡터(문맥까지 포함하고 있는)
 - 문맥 벡터를 보니 다른 방향을 가르키는 다른 단어더라 하고 알 수 있는 것
- 난 밥을 먹는다는 self attention 안하고 번역하면
 - i eat food라고 안나오고 i food eat라고 나옴

역전파

- 톱니바퀴가 알아서 뚝뚝뚝뚝