



Attention is all you need

논문 리뷰



이희종



Introduction

- 기존의 RNN과 CNN을 방식을 과감히 포기한 모델이다.
- RNN의 원론적인 문제는 순차적인 처리과정에 있다. 그렇기 때문에 데이터를 처리하는데 있어 시간이 오래걸렸고 병렬처리가 불가능했다.
- 이를 해결한 것이 바로 Attention이다. Attention 기법은 input 과 output의 거리에 상관없이 dependency를 학습하는 것이다.
- 이로 인해 병렬성을 증가시켜 적은 시간을 들여도 최신 모델의 성능을 내게 하는 결과를 가져온다.

구성 요소

최신 모델들이 가지고 있는 구조와 비슷하게 **Attention** 또한 인코더와 디코더를 가지고 있다.

- 인코더, 디코더 둘다 **FNN**과 **Skip Connection** 기반의 모듈을 **N**개 쌓아서 **LSTM**이나 **GRU** 같은 **RNN** 구조 없이도 두 벡터 사이의 연관성을 기반으로 결과를 도출해내는 새로운 방식을 채택했다.

기본적인 모델 구조

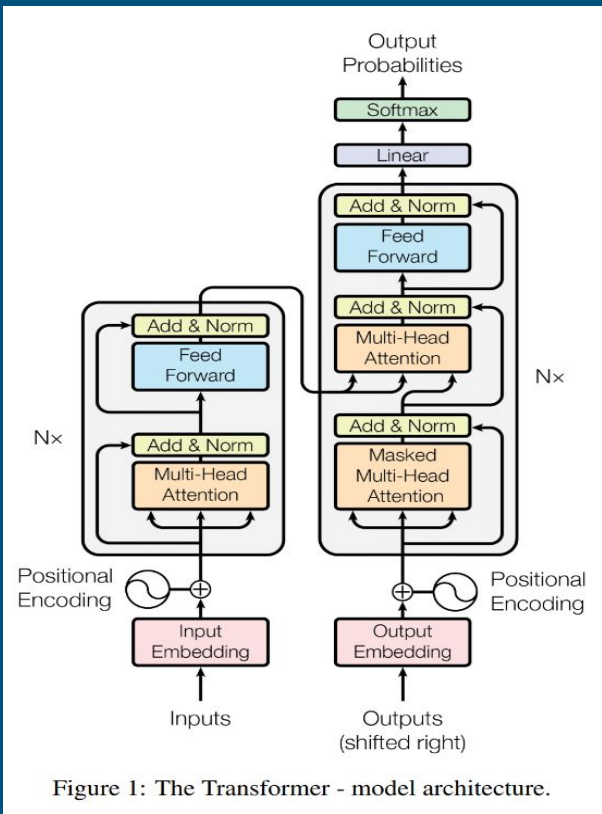


Figure 1: The Transformer - model architecture.

- Transformer를 사용하는 구조를 가지고 있다.
- 자연어 처리를 하는데 적합한 구조를 지닌다.
- Attention은 Key-value의 쌍과 query의 매핑이라 할 수 있다. 즉 최종 출력을 value들의 가중치합의 결과로 구해진다.