

제목 : Attention Is All You Need

1. 논문 주제 및 배경:

- 구글 브레인과 구글 리서치 팀에서 기계 번역 분야에서 state-of-the-art 성능을 달성하는 신경망 구조를 제안한 것임.
- Transformer 모델을 소개하며, 기본적으로 encoder-decoder 구조를 가지고 있음.

2. 모델 구조:

- Encoder:

6개의 동일한 레이어로 구성되며, 각 레이어는 multi-head self-attention과 position-wise fully connected feed-forward network로 이루어져 있음.

- Decoder:** Encoder와 유사한 구조이지만, self-attention과 feed-forward network 사이에 multi-head attention이 추가로 들어감.

이건 decoder에서 encoder로의 attention을 나타내는 것.

3. 주요 기술:

- Multi-head Self-Attention:

현재 레이어에 들어오는 시퀀스의 각 위치에서 이전 레이어의 모든 위치에 attention을 주는 구조.