

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit,

Llion Jones, Aidan N. Gomez, Lukasz Kaiser

December 6, 2017

01. Introduction

RNN, LSTM, GRU 와 같은 언어 모델링들이 좋은 성능을 내왔다. 하지만 이러한 Recurrent 기반의 모델들은 이전 결과를 입력으로 받는다는 점 때문에 병렬처리가 불가능하다. 최근에는 factorization tricks 와 conditional computation 을 통해 연산 효율의 대폭적인 향상을 달성했지만 여전히 recurrent 기반의 순차적 계산으로 이루어진다.

이 연구에서는 recurrent 의 특성들을 피하고 입력과 출력 사이의 전역 의존성을 이끌어내기 위해 transformer 을 제안한다.

02. Model Architecture

Attention 기법만으로 encoder-decoder 구조를 설계했다.

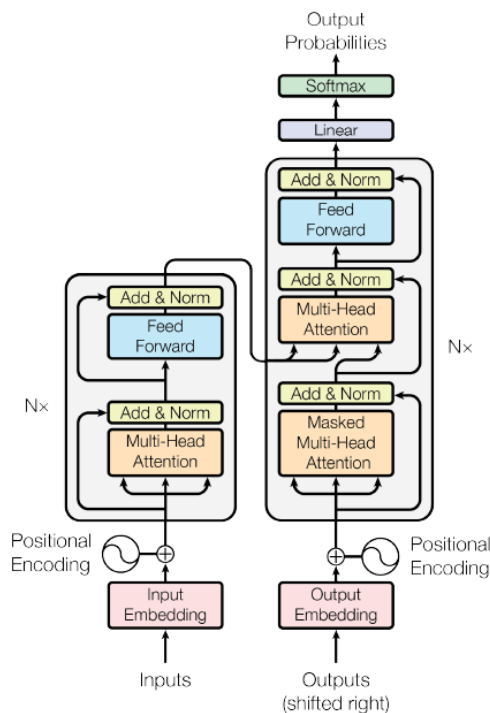


Figure 1: The Transformer - model architecture.

Encoder 은 6 개의 동일한 층으로 구성되어 있고 각 층에는 2 개의 sub-layer 이 있다. 이 sub layer 주의에 residual connection 을 채택하고 있다. Decoder 은 encoder 과 동일하게 6 개의 층과 각 층에는 2 개의 sub layer 으로 구성이 되어 있다.

Self-Attention 을 사용하면 다음과 같은 장점들이 있다. 먼저 레이어당 계산량이 줄어들고 병렬처리가 가능한 계산이 늘어난다. 그리고 장거리 학습이 가능해진다. 이러한 점들은 아래의 표에서 확인이 가능하다.

Layer Type	Complexity per Layer	Sequential Operations	Maximum Path Length
Self-Attention	$O(n^2 \cdot d)$	$O(1)$	$O(1)$
Recurrent	$O(n \cdot d^2)$	$O(n)$	$O(n)$
Convolutional	$O(k \cdot n \cdot d^2)$	$O(1)$	$O(\log_k(n))$
Self-Attention (restricted)	$O(r \cdot n \cdot d)$	$O(1)$	$O(n/r)$

03. Training

WMT 2014 English-German Dataset 을 이용했으며 GPU 8 대로 기본 모델은 12 시간 동안을 big 모델은 3.5 일 동안 학습시켰다.

Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [18]	23.75			
Deep-Att + PosUnk [39]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [38]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	41.29	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$3.3 \cdot 10^{18}$	
Transformer (big)	28.4	41.8	$2.3 \cdot 10^{19}$	

성능에 비해 빠른 속도로 우수한 BLEU scores 을 획득했음을 알 수 있다.

Reference

<https://arxiv.org/pdf/1706.03762.pdf>

<https://hipgyung.tistory.com/entry/ATTENTION-IS-ALL-YOU-NEED-%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%B0>