

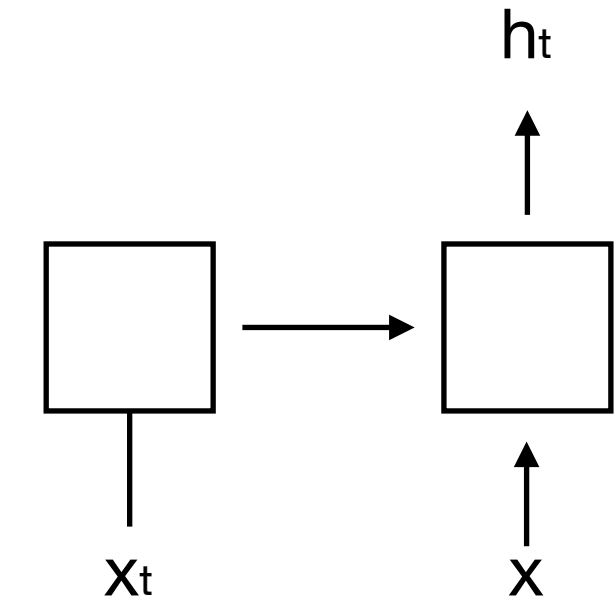
순차 데이터(자연어) 처리

: RNN, LSTM, Attention, Transformer

RNN(Recurrent Neural Network)

개요

RNN은 순환신경망 이라는 이름 그대로 과거의 정보를 기억하고 그 정보를 다음 계산에 전달하여 처리하는 구조.
시계열 데이터나 문장처럼 순서가 중요한 데이터를 처리할 때 적합하다.



RNN의 구조

- 입력 : 단어 시퀀스 x_1, x_2, \dots, x_t
- 은닉 : $h_t = \tanh(Wx_t + Uh_{t-1} + b)$
- 출력 : 필요하면 softmax등으로 변환

중요한 특징

- h_t 는 이전 상태 h_{t-1} 를 기반으로 업데이트 되기 때문에 문맥 정보를 유지할 수 있음.
- 문장의 앞쪽 정보가 뒷쪽 단어 해석에 영향을 줄 수 있음.

RNN의 한계

- 장기 의존성 문제 : 긴 문장에서는 앞쪽 정보가 뒤로 갈수록 소실되어 잘 전달 되지 않음.
- 기울기 소실 : 역전파시 tanh나 sigmoid때문에 gradient가 0또는 무한대로 발산 할 수 있음.
- 병렬 처리의 어려움 : 시퀀스를 순차적으로 처리해야 하므로 GPU의 효율이 떨어짐.

RNN 사용처

- 간단한 텍스트 분류
- 시계열 예측(ex: 주식, 날씨)
- 음성 데이터 처리
- 자연어 처리의 기초 학습 및 구조 이해

LSTM(Long Short-Term Memory)

개요

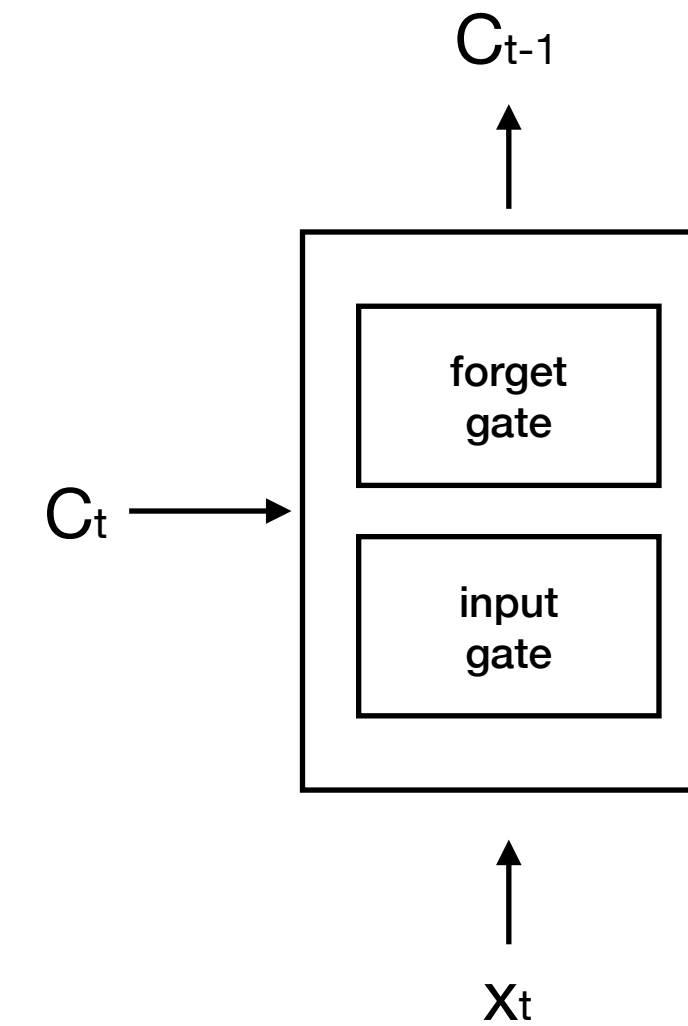
RNN의 단점을 보완하기 위해 등장.

기억 셀(cell state)을 도입해 장기 기억 유지

- 장점 : 장기 의존성 문제 해결 (Cell State 사용)
- 단점 : 구조가 복잡하고 학습 시간 증가

구조 요약(Gate 3개 : 입력, 망각, 출력 게이트)

- 기억 할지 말지를 Gate가 결정
- 장기 기억(cell state) 과 단기 기억(hidden state)로 구성



RNN과 LSTM의 비교

항목	RNN	LSTM
구조	가장 단순	복잡(3개 Gate)
성능	낮음	높음
학습시간	빠름	느림
장기 기억	불가능	가능(Cell State)
실전 활용도	낮음	높음(기본 선택자)

주요 사용 목적

- RNN : 교육 목적 또는 아주 간단한 시퀀스 데이터
- LSTM : 문장 생성, 번역 등 긴 문맥이 중요한 작업.

Attention

개요

Attention은 입력 시퀀스에서 중요한 부분에 가중치를 두어 집중한다.

문장이 길어질수록 모든 단어를 동일하게 처리하지 않고, 문맥상 중요한 단어에 집중 한다.

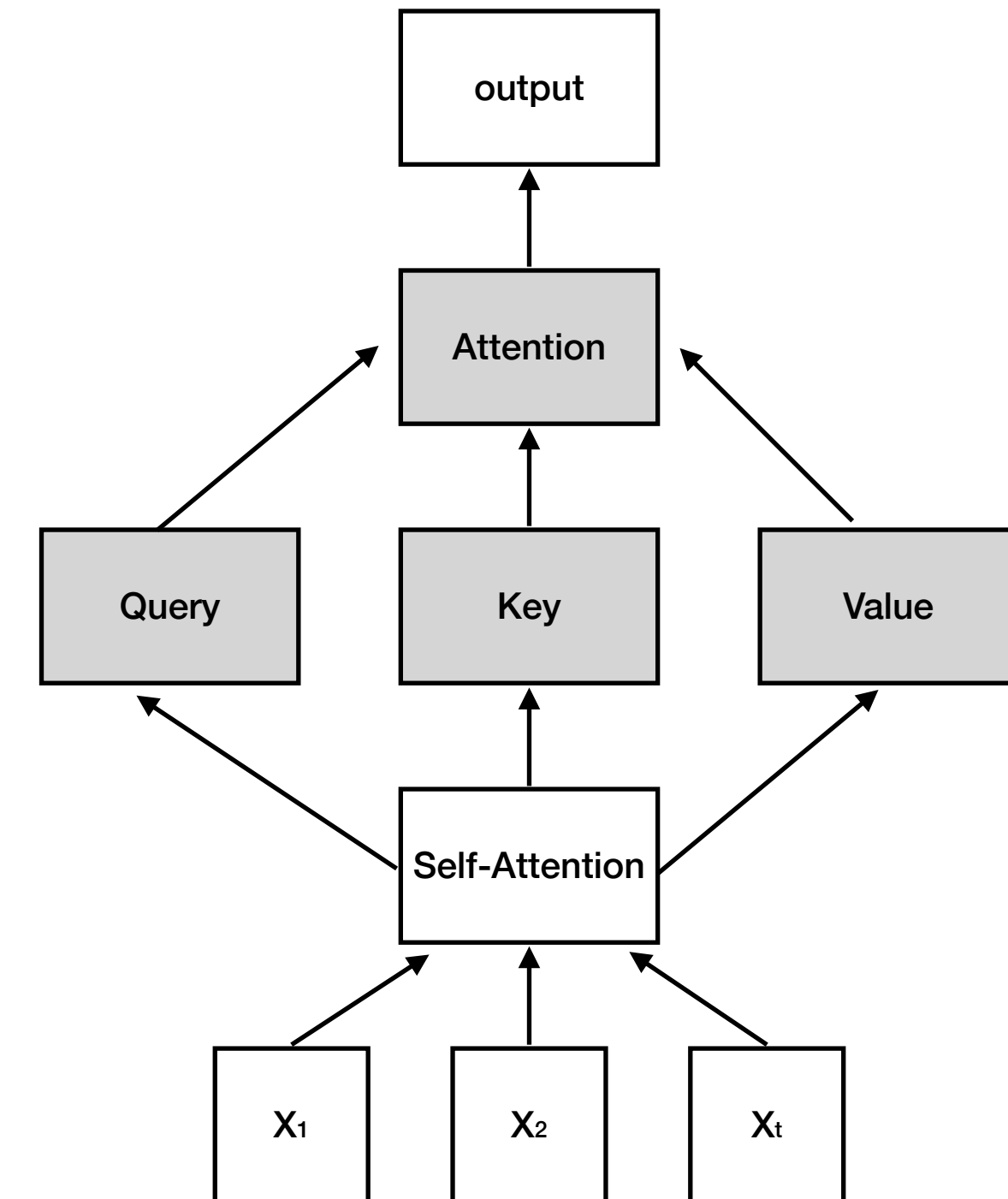
Self-Attention

자기 자신에 대해 Attention을 계산하는 방법으로 Transfomer에서의 핵심 역할.

Example

"The animal didn't cross the street because it was too tired."

: it이 가리키는 것은 "animal" <- 이 연결을 해내는게 attention의 강점



Transformer

개요

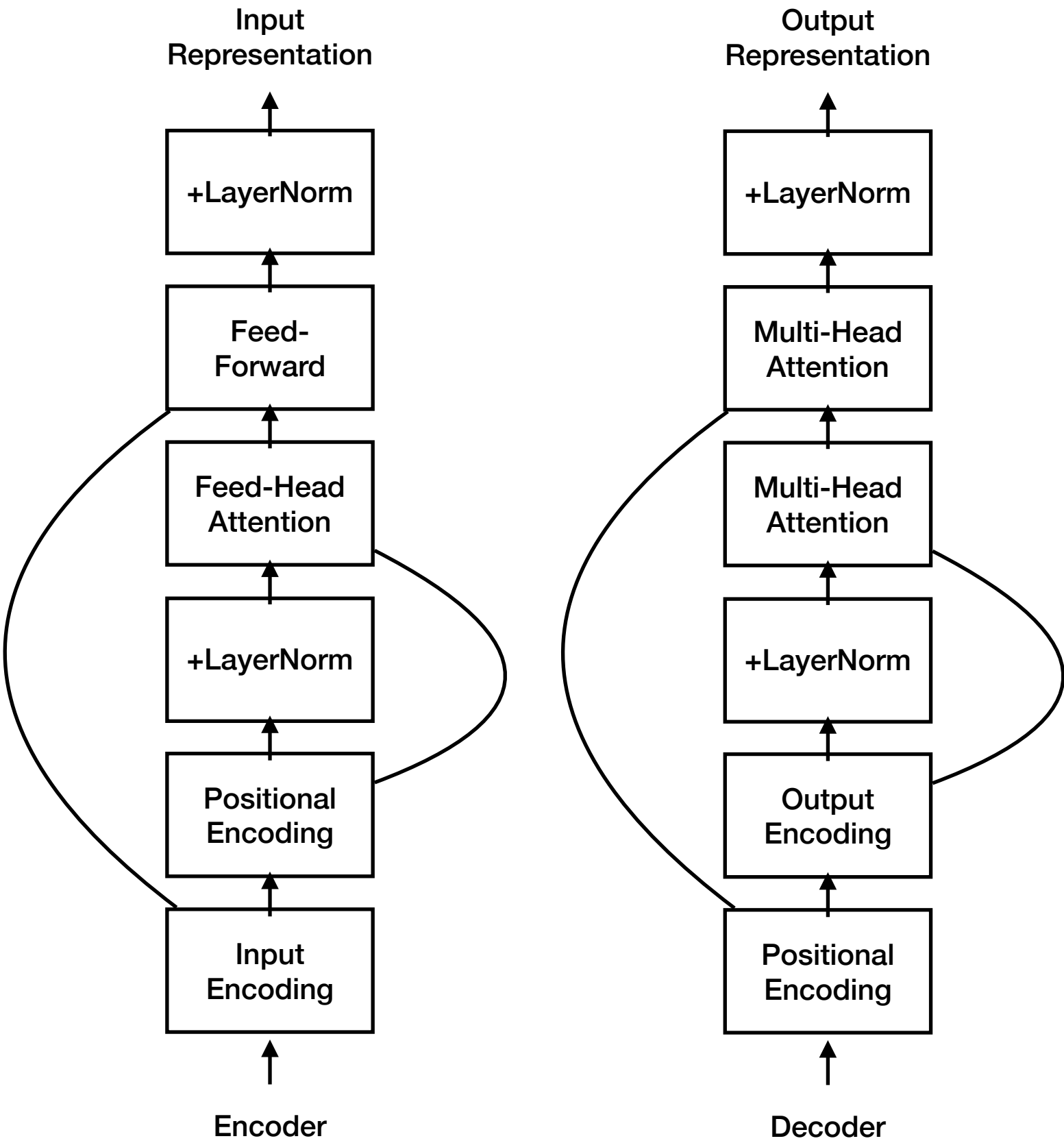
- 2017년 "Attention is All You Need" 논문에서 제안
- RNN없이 순차 처리 가능(병렬 처리 가능 해서 빠르다)
- Self-Attention + Positional Encoding + Feed Forward로 구성

구조 구성 요소

구성요소	내용
Input Encoding	단어를 벡터로 변환
Positional Encoding	순서를 나타내는 벡터 추가
Multi-Head Attention	다양한 관점에서 Attention 수행
Feed-Forward Network	각 토큰을 독립적으로 처리
Residual + Layer Norm	안정적인 학습을 위한 보조

장점

- 장기 의존성 문제 해결 : RNN이나 LSTM보다 더 멀리 떨어진 단어 관계도 잘 처리함.
- 병렬 처리 가능 : 학습 속도 우수
- 모듈화 용이 : 다양한 모델이 쉽게 적용 가능(ex: BERT, GPT 등)



RNN, LSTM, Attention, Transfomer의 비교

항목	RNN / LSTM	Attention / Transformer
순차처리	필수(느림)	병렬 가능(빠름)
장기 의존성 처리	어려움	매우 우수
구조 복잡도	상대적으로 간단	다소 복잡
학습 속도	느림	빠름
대표 모델	LSTM, GRU	BERT, GPT, T5, ViT 등