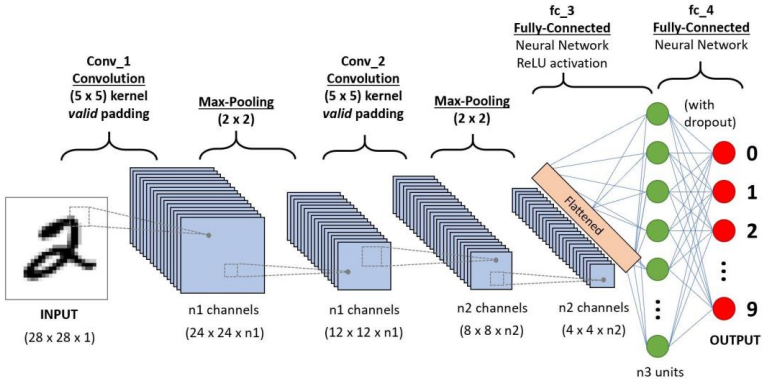
고려대학교 빅데이터 학회

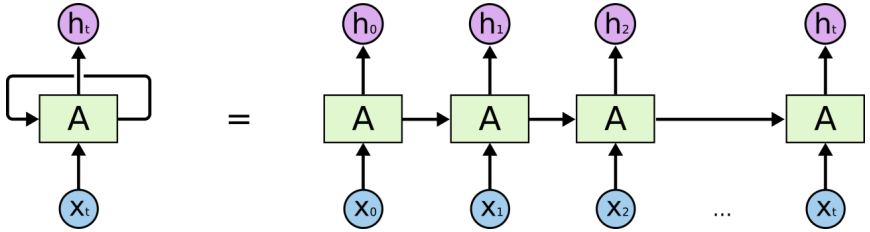
Natural Language Processing

KU-BIG

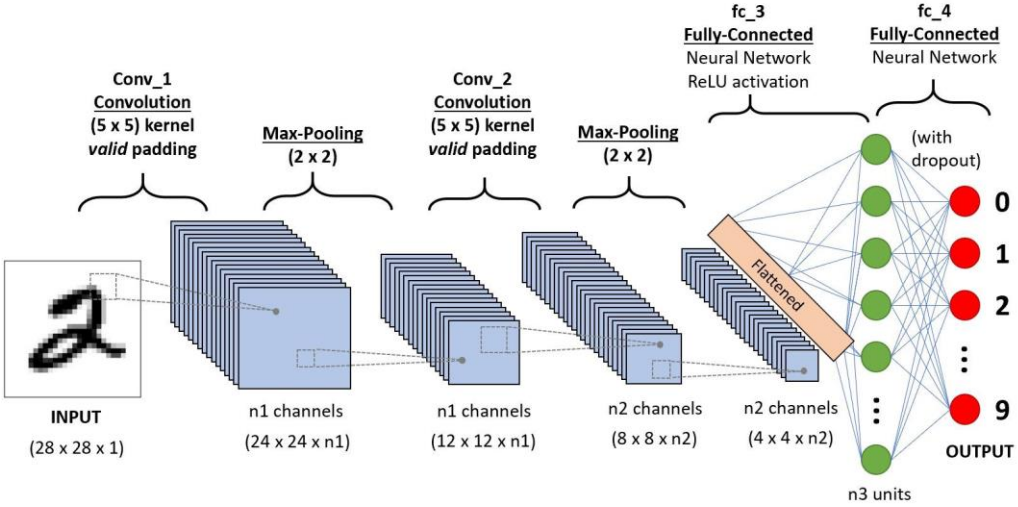
목차

Sequence


Conv_1
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
Conv_2
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
fc_3
Fully-Connected
Neural Network
ReLU activation
fc_4
Fully-Connected
Neural Network
(with
dropout)
Flattened
INPUT
(28 x 28 x 1)
n1 channels
(24 x 24 x n1)
n1 channels
(12 x 12 x n1)
n2 channels
(8 x 8 x n2)
n2 channels
(4 x 4 x n2)
n3 units
0
1
2
9
OUTPUT

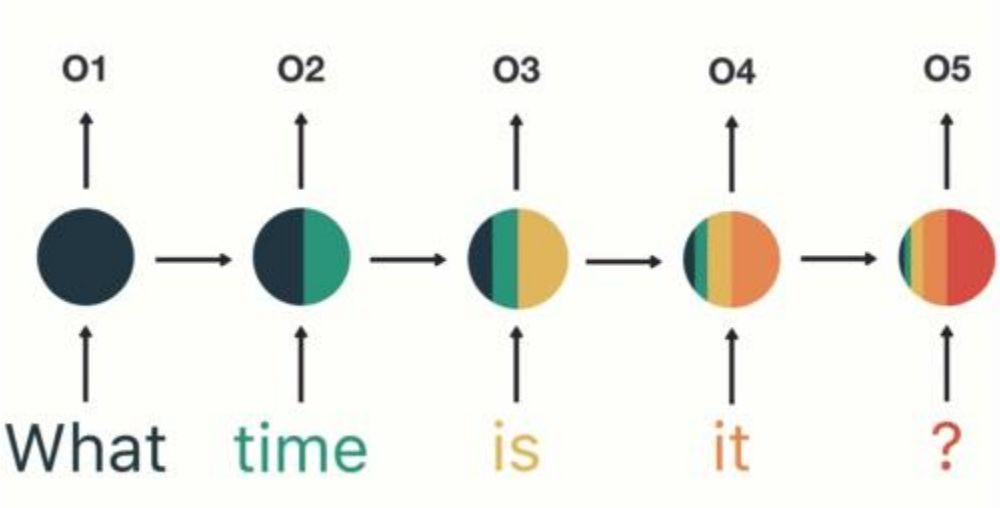

A
=
A
A
A
A
...

Sequence

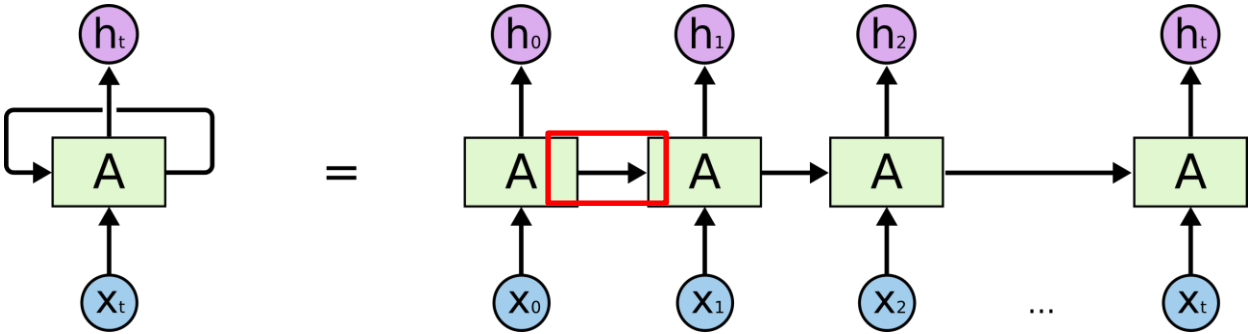


단어의 순서?

RNN

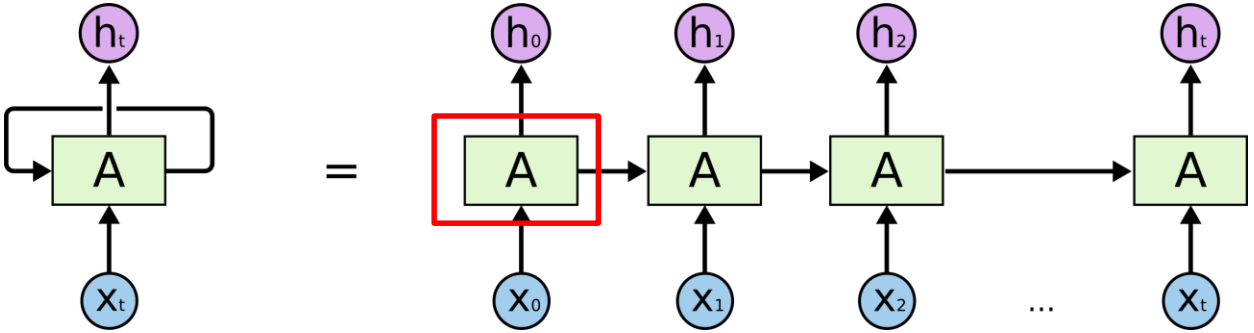


RNN



X = 전처리한 단어 벡터
H = 각 셀 별 연산 결과
(batch_size, sequence_len,
word_embedding_len)

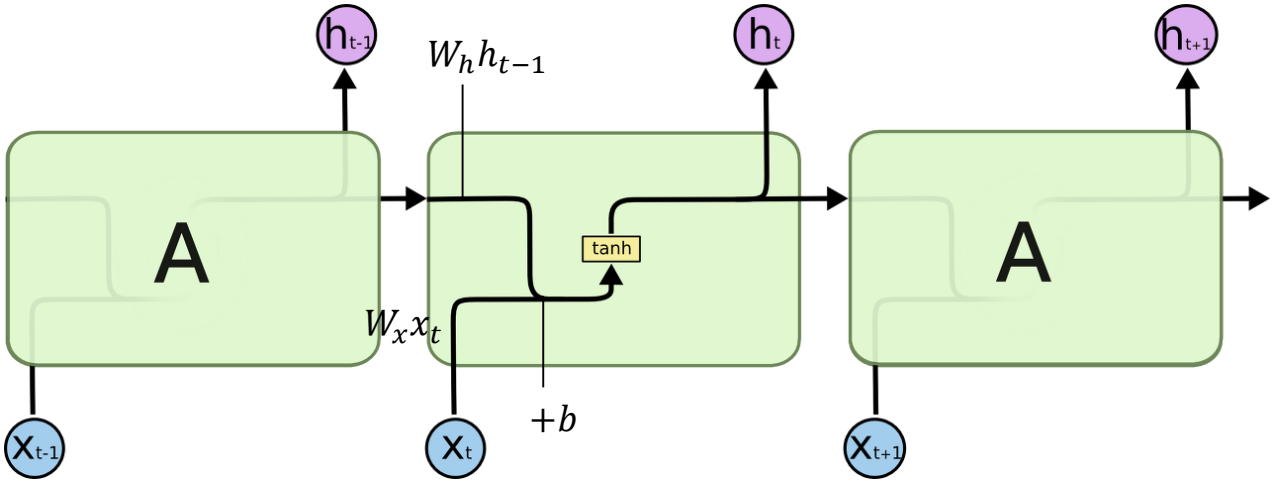
RNN



Vanilla RNN, LSTM, GRU

Vanilla RNN

$$h_t = \tanh(W_h h_{t-1} + W_x x_t + b)$$

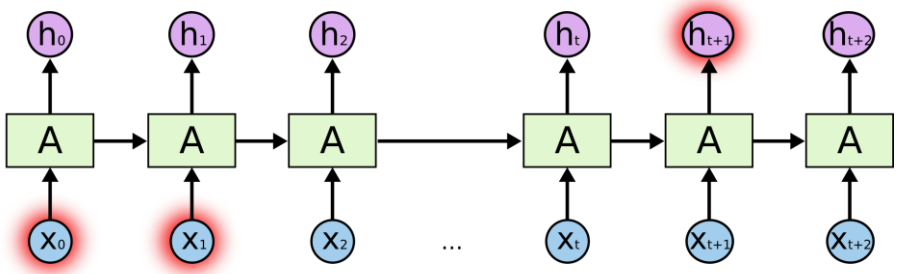


Vanishing gradient

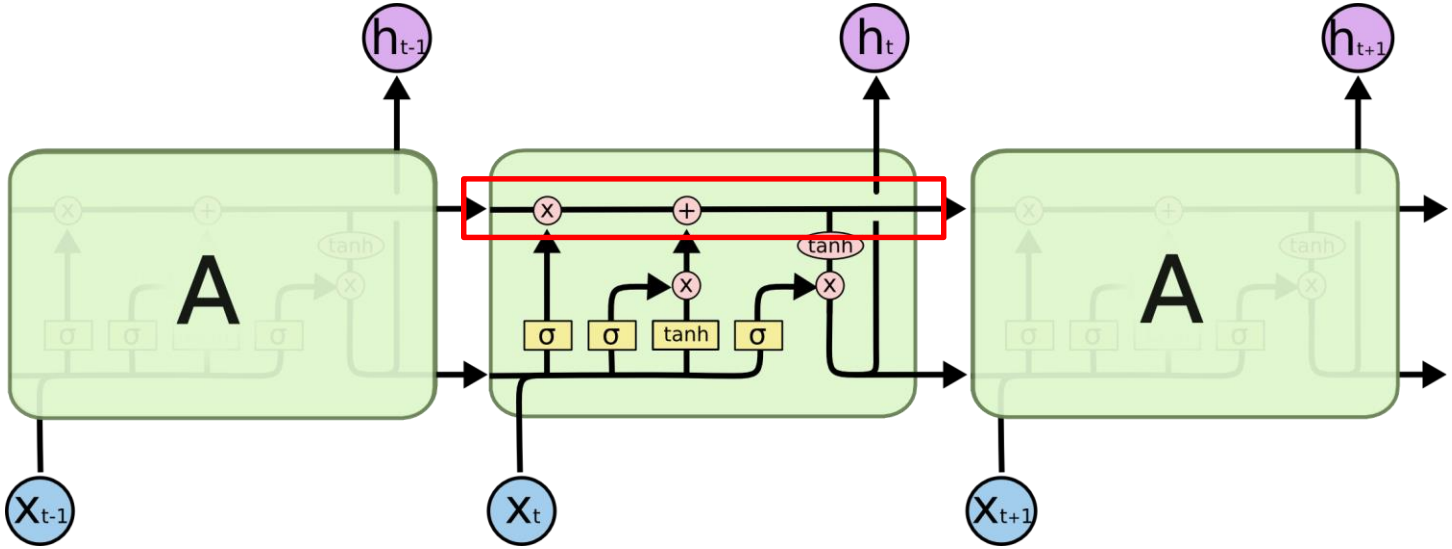
$$h_t = \tanh(W_h h_{t-1} + W_x x_t + b)$$

$$h_t = \tanh(W_h tanh(W_h h_{t-2} + W_x x_{t-1} + b) + W_x x_t + b)$$

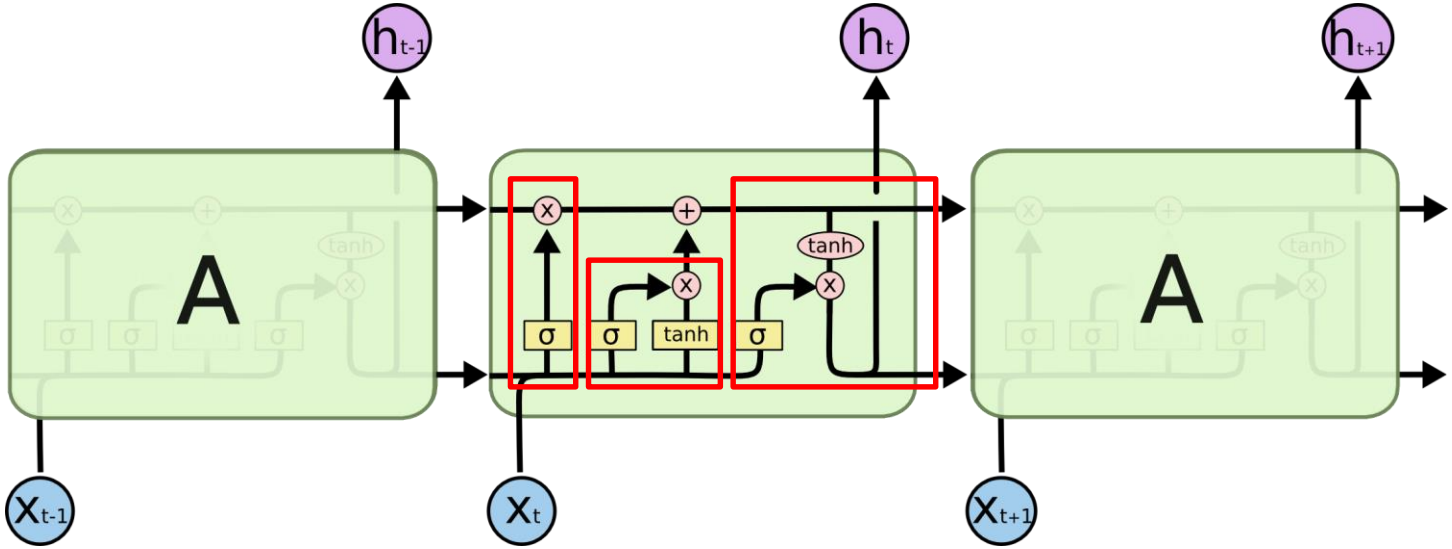
$$h_t = \tanh(W_h tanh(W_h tanh(W_h h_{t-3} + W_x x_{t-2} + b) + W_x x_{t-1} + b) + W_x x_t + b)$$



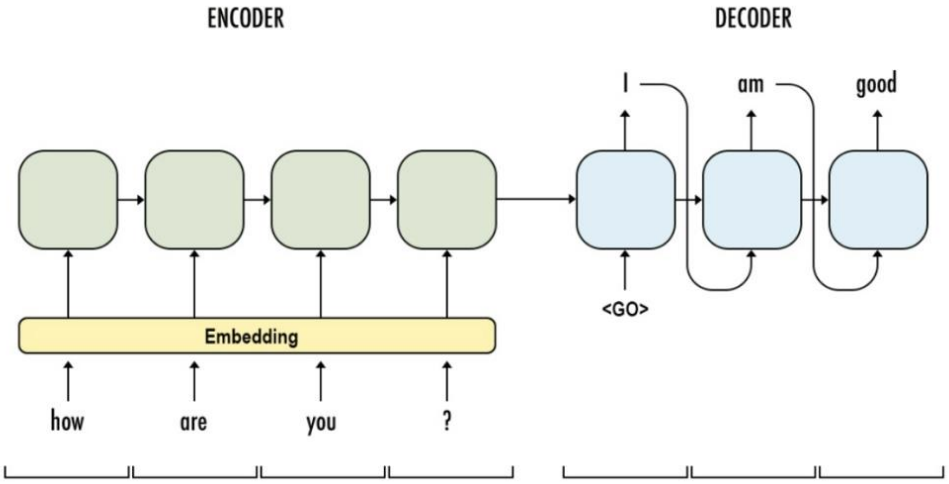
LSTM(Long short-term memory)



LSTM(Long short-term memory)

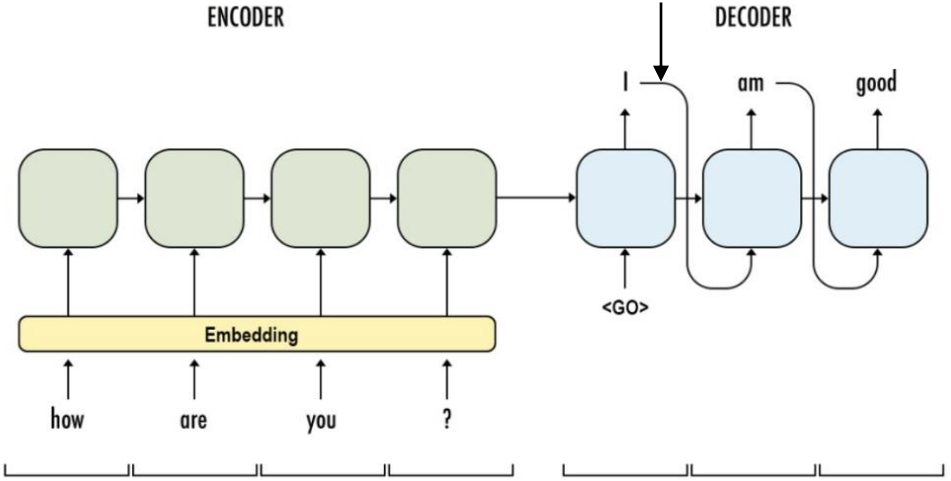


Seq2Seq?



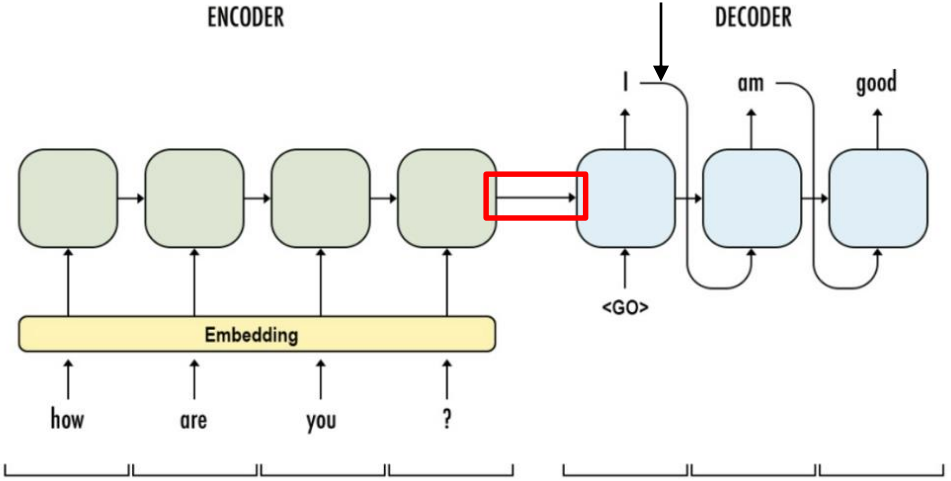
Seq2Seq

학습 중에는 실제 label값을 입력(교사강요)



Seq2Seq

학습 중에는 실제 label값을 입력(교사강요)

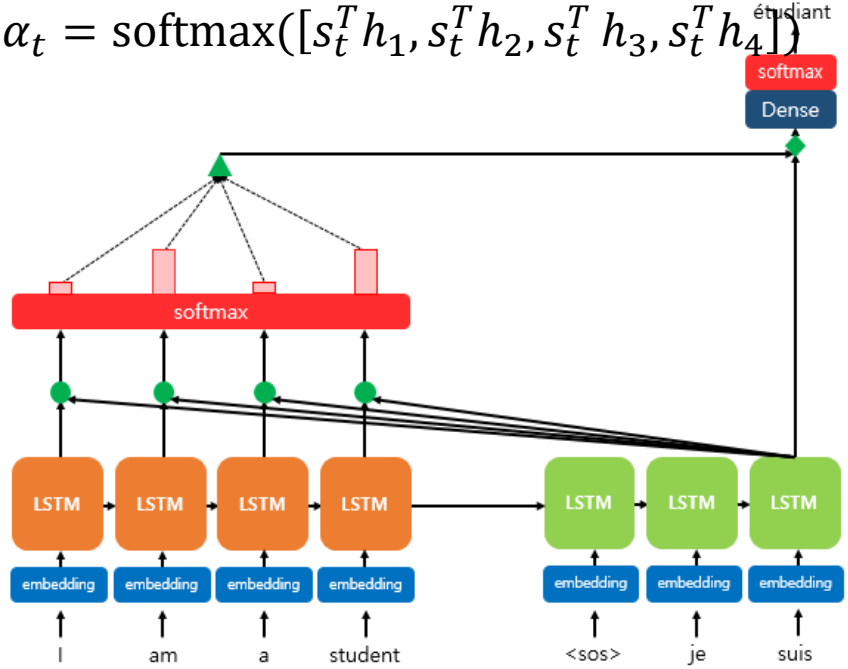


1. 한정된 context 벡터 크기
2. Sequence가 길어질 때

Attention

$$\alpha_t = \text{softmax}([s_t^T h_1, s_t^T h_2, s_t^T h_3, s_t^T h_4])$$

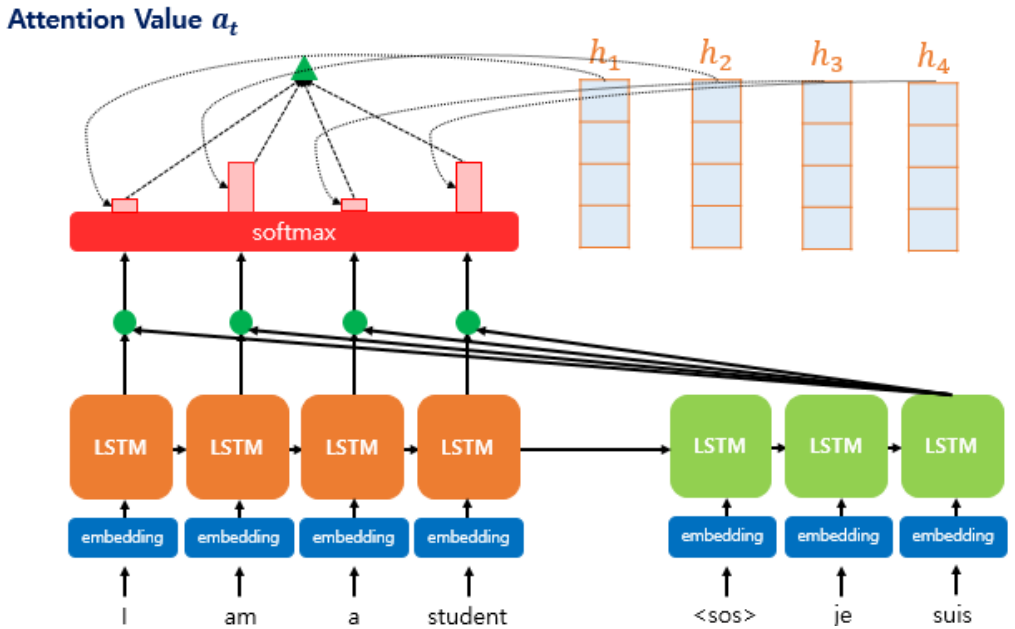
$\alpha_t =$



$h_i =$ 인코더의 출력 $s_t =$ 디코더의 출력

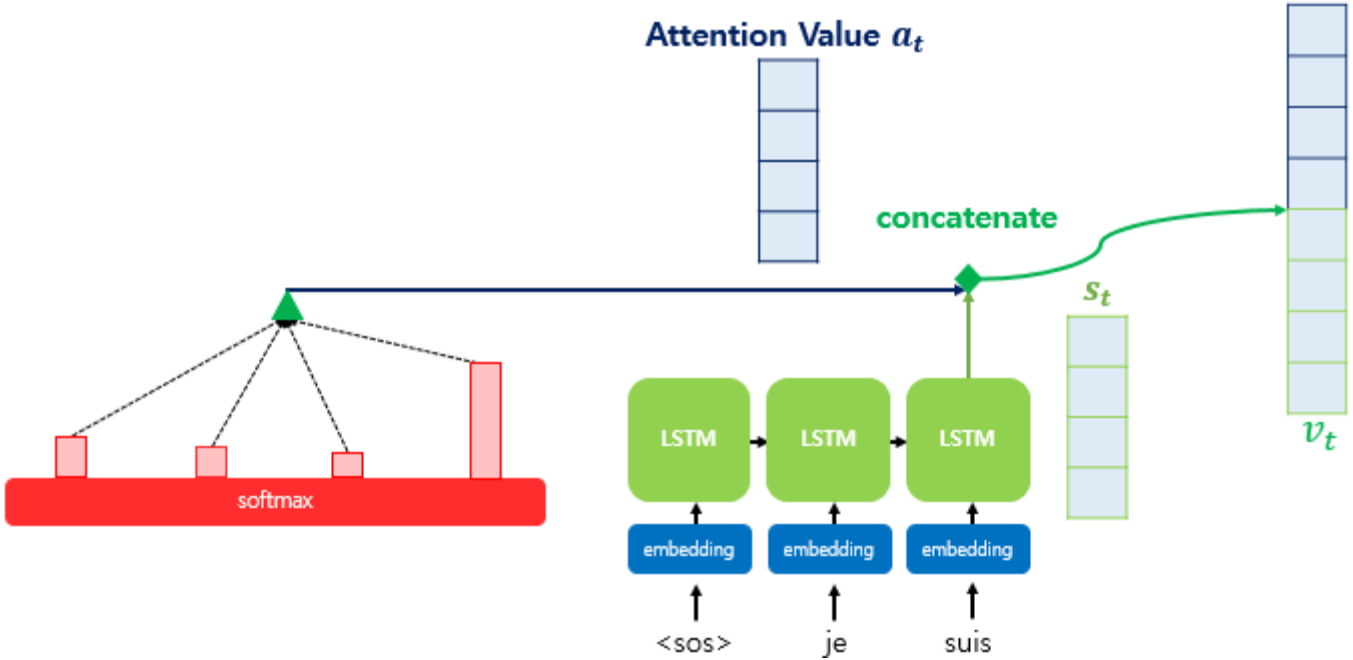
Attention

$\alpha_t =$



$h_i =$ 인코더의 출력 $s_t =$ 디코더의 출력

Attention

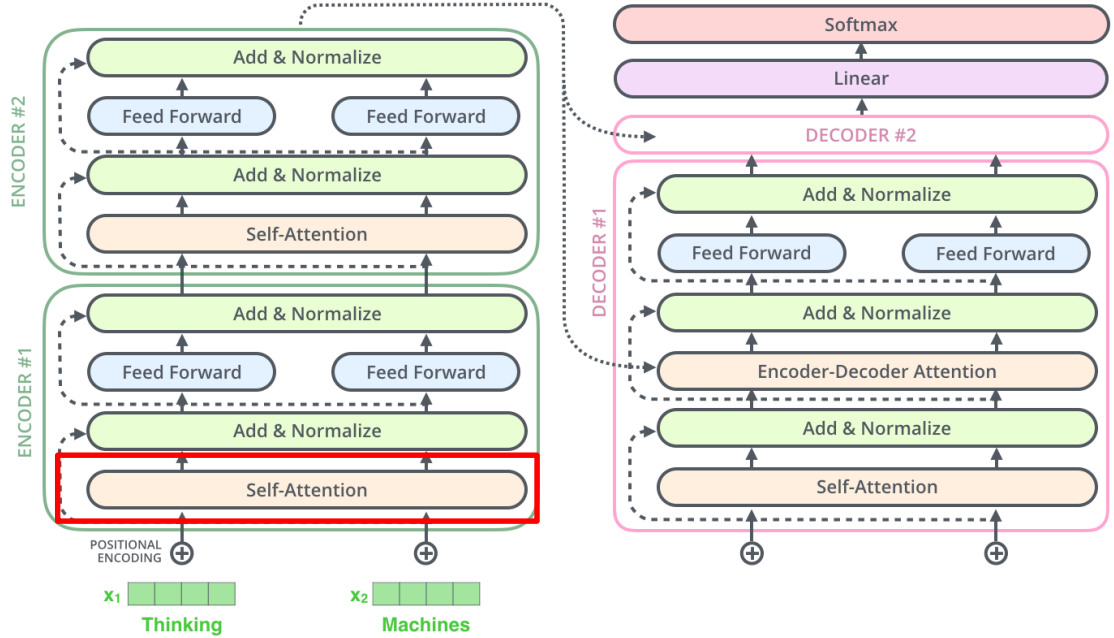

Attention Value a_t
concatenate
s_t
v_t
softmax
LSTM
LSTM
LSTM
embedding
embedding
embedding
<sos>
je
suis

Transformer

RNN을 쓰지 않고
보정용이 아니라 순수 Attention만 써서?

Attention Is All You Need(2017)

Transformer

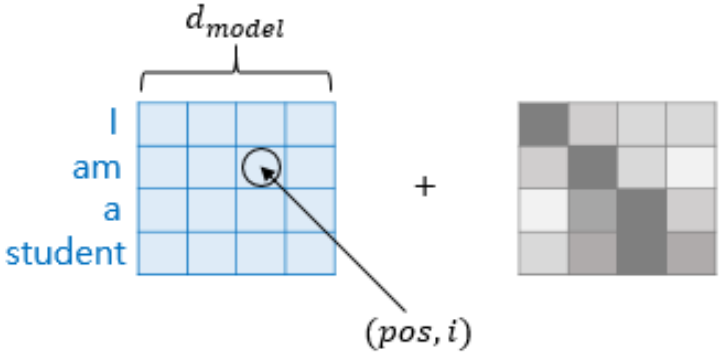


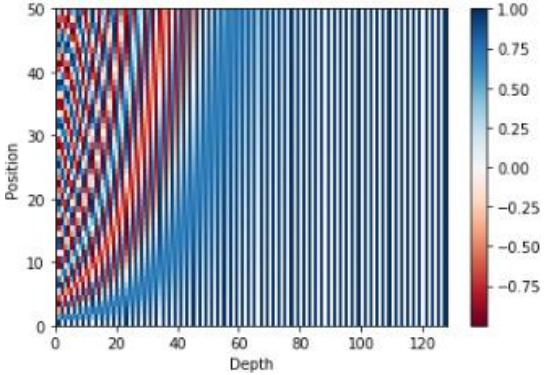
Transformer

단어의 순서?

Positional Encoding!

Transformer

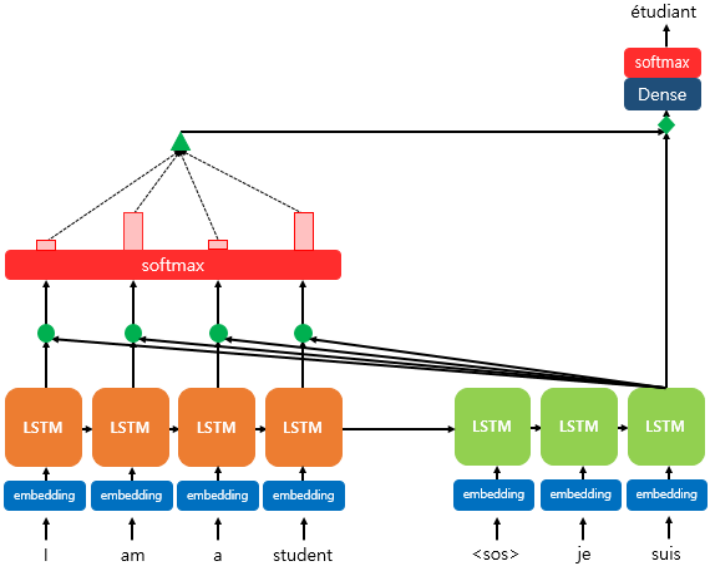




$$PE_{(pos,\,2i)} = sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,\,2i+1)} = cos(pos/10000^{2i/d_{model}})$$

Transformer



Attention에 대한 이해

Transformer

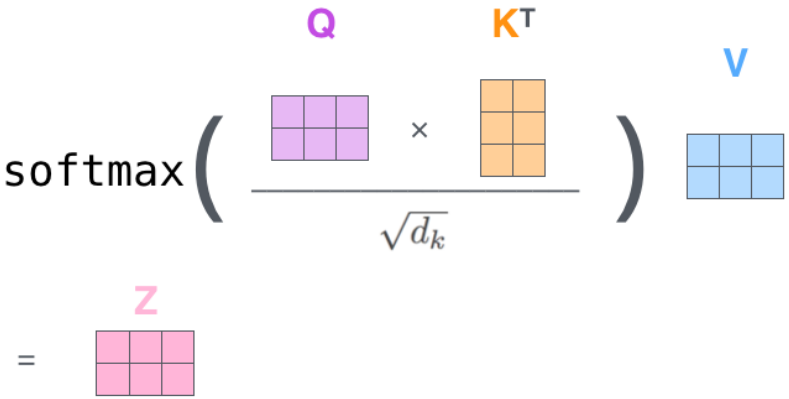
Query, Key, Value

Q와 유사한 것을 찾기 위해 Q와 K로 특정한 연산을 한 뒤, 유사도를 계산

높은 유사도를 갖는 애의 Value를 더 많이 참고하여 어텐션 벡터를 구한다.

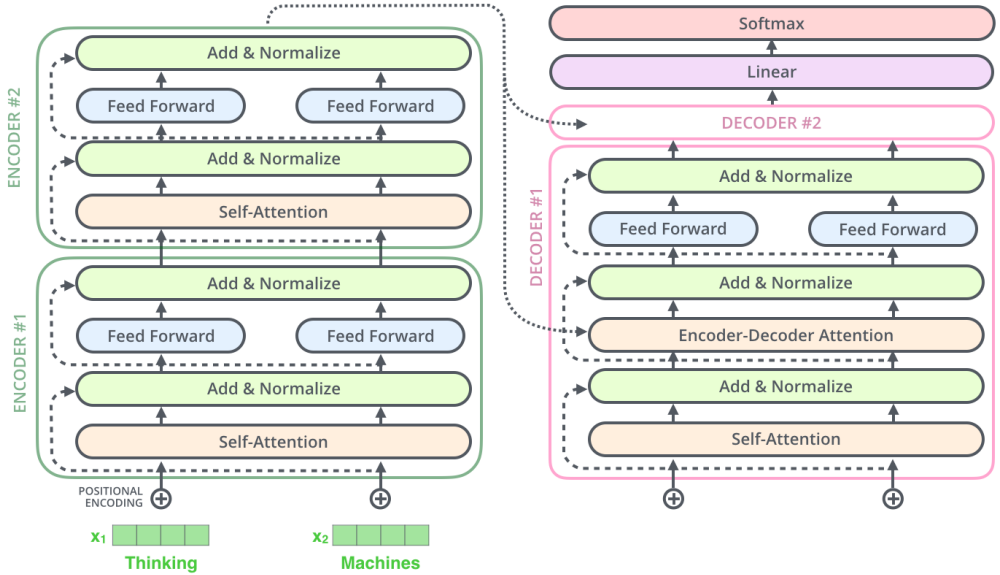
Transformer

Q와 K와 V사이
연산으로 추상화 가능



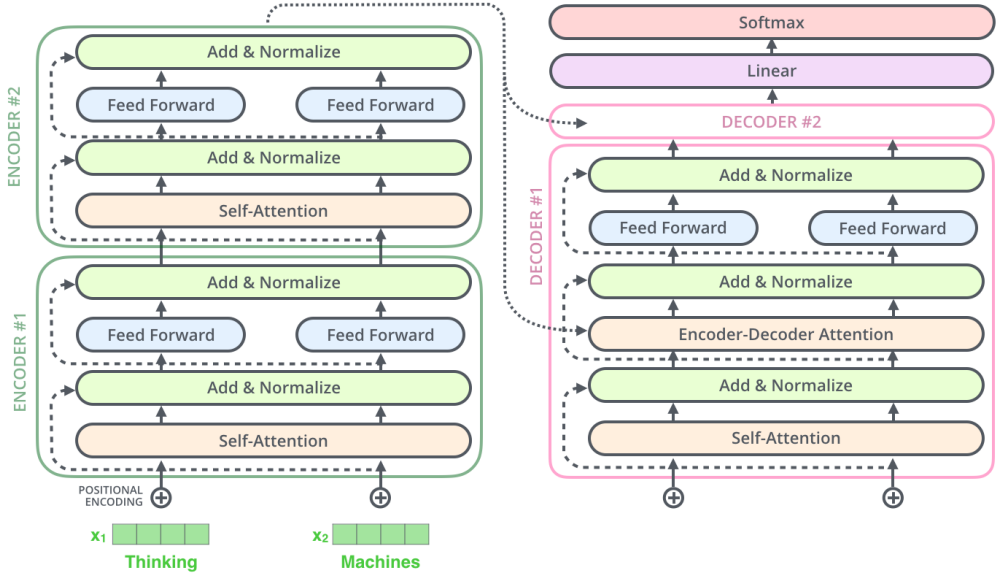
Transformer

Self Attention
Encoder – Decoder Attention



Transformer

$$FFNN(x) = MAX(0, xW_1 + b_1)W_2 + b_2$$



BERT

Pre-training of Deep Bidirectional Transformers for Language Understanding

(Bidirectional Encoder Representations form Transformer)

BERT

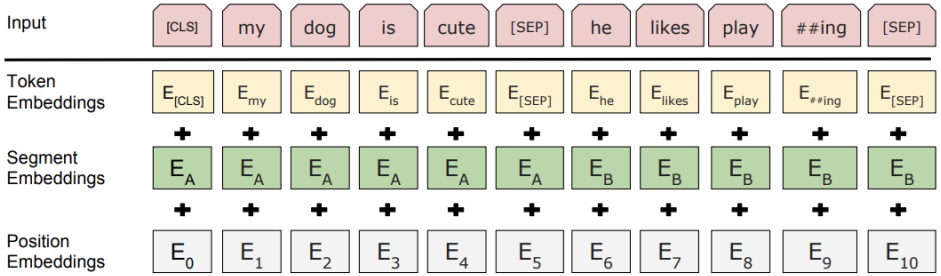
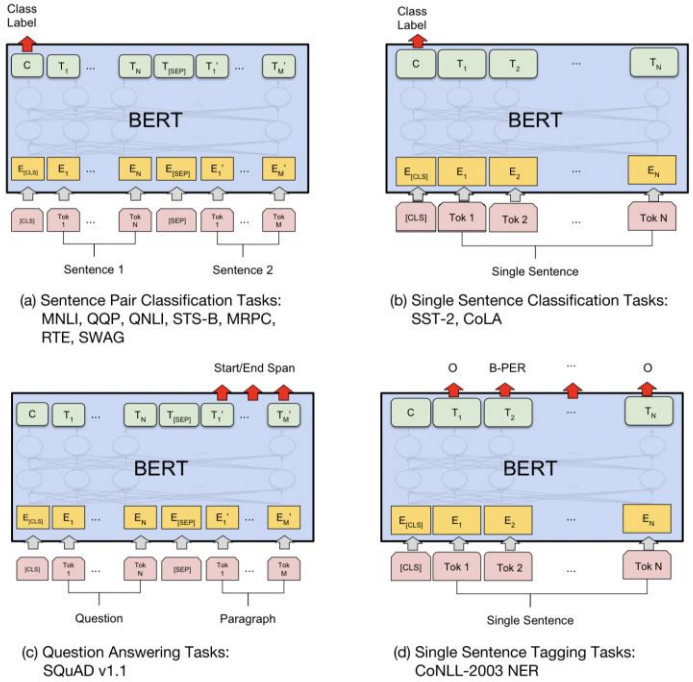


Figure 2: BERT input representation. The input embeddings is the sum of the token embeddings, the segmentation embeddings and the position embeddings.

BERT

Finetuning에 매우 유용함

BERT


Class Label
BERT
Sentence 1
Sentence 2
(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG
Class Label
BERT
Single Sentence
(b) Single Sentence Classification Tasks: SST-2, CoLA
Start/End Span
BERT
Question
Paragraph
(c) Question Answering Tasks: SQuAD v1.1
O
B-PER
O
BERT
Single Sentence
(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

BERT

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, Il milione (or, The Million, known in English as the Travels of Marco Polo), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge through contact with Persian traders since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: through contact with Persian traders

런던 놀러가자 - 장소 예약
강남가는 버스가 있어? - 장소 교통수단 물음

오늘 택배 온대
헉 말만 들어도 설렌다 - Positive Sample
지금 시간 돼?
항상 응원 할게! - Negative Sample

Thank you