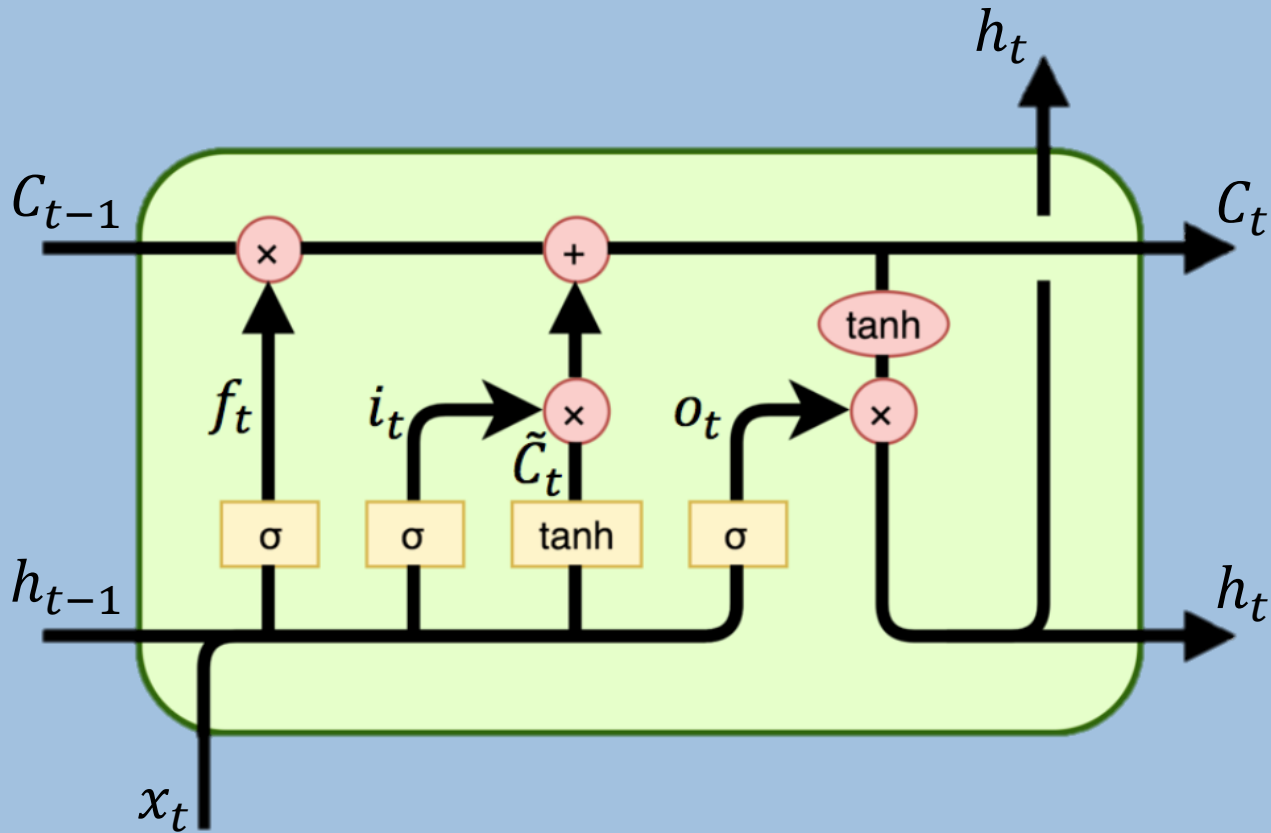


2022-1 겨울방학 딥러닝 분반

분반장: 구은아, 김혜림

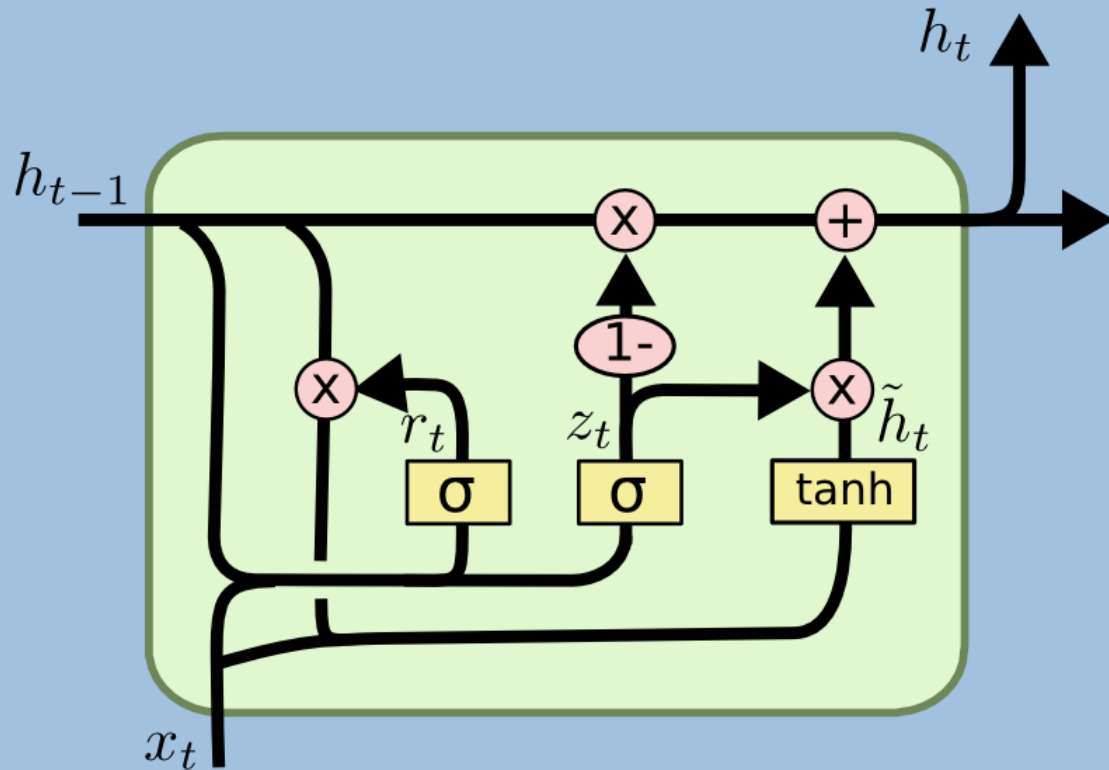
RNN Architecture

LSTM



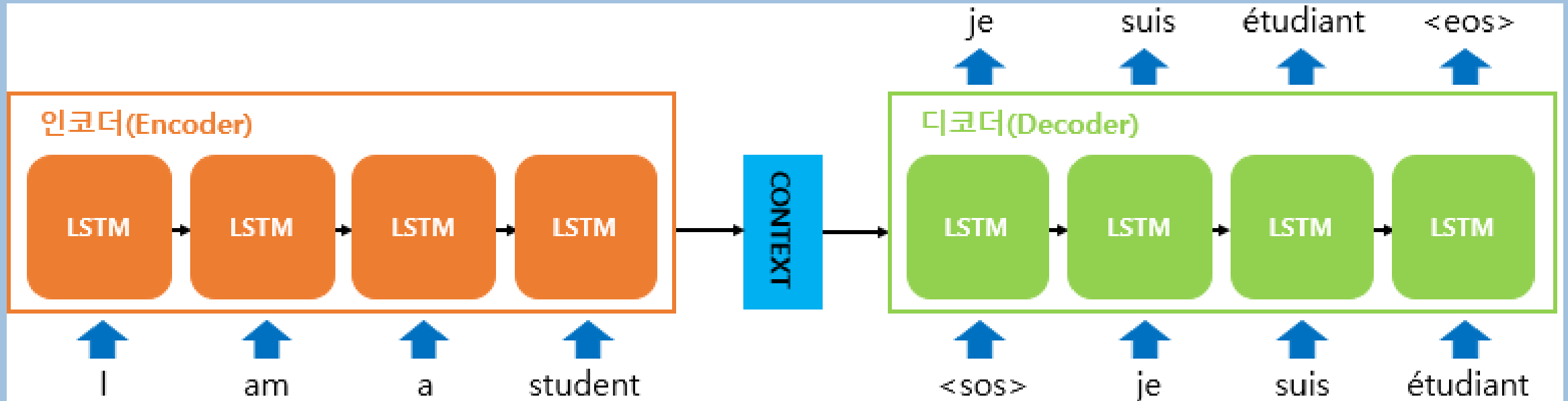
- 바닐라 RNN의 long term dependency 문제를 해결하기 위해 등장한 구조.
- 시그모이드 함수를 사용하는 input gate, forget gate, output gate가 추가되었다.
- 추가된 게이트를 사용하여 이전 cell state에서 잊을 내용, 현재 input에서 선택할 내용을 조절한다.

GRU



- LSTM을 간소화한 버전.
- cell state가 존재하지 않고, 시그모이드 함수를 사용하는 reset gate와 update gate로 이루어져 있다.
- 2개의 게이트를 사용하여 이전 hidden state와 현재 input의 비율을 조정한다.

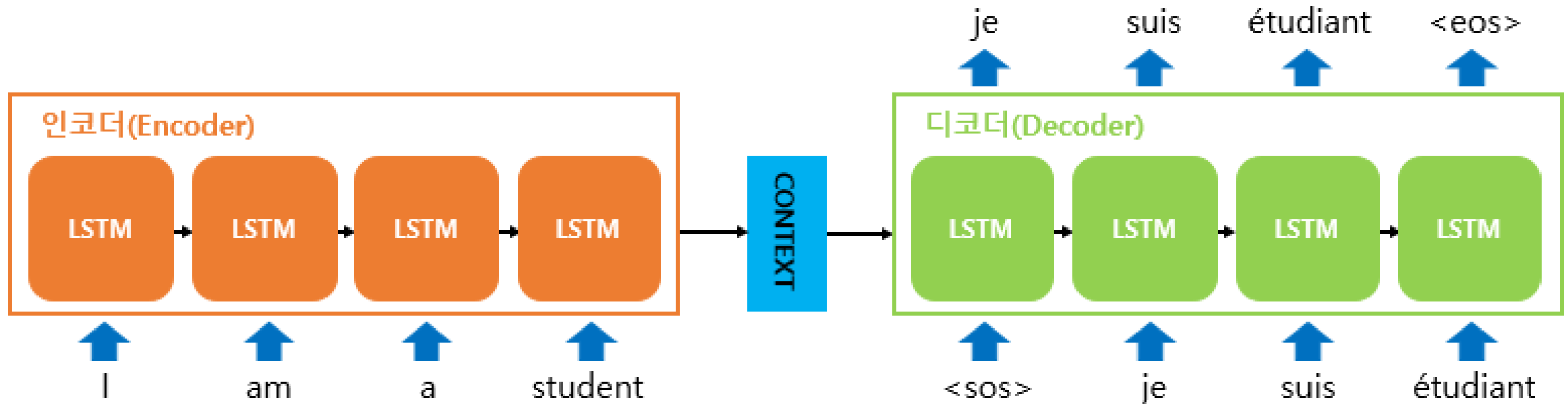
seq2seq



- 문장을 문장으로 변환하는 기본 모델로, 인코더와 디코더가 존재한다.
- 인코더 마지막 RNN 셀의 hidden state가 context vector가 되어 디코더의 첫번째 RNN 셀로 이어짐으로써 입력 문장의 정보가 디코더로 전달된다.

어텐션

seq2seq 모델의 문제점

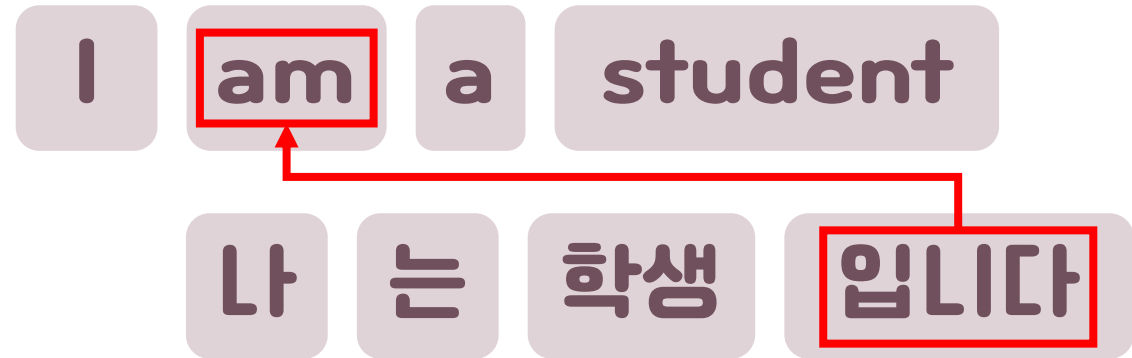


인코더의 마지막 hidden state만 디코더가 공유한다. 즉, 하나의 context vector에 입력 문장의 모든 정보가 압축된다. 이로 인해 입력 문장의 정보가 손실 된다.

정보 손실 문제를 해결하기 위해 어텐션 등장

어텐션 매커니즘

- 어텐션 매커니즘은 디코더의 매 time step마다 input을 다시 확인하되, 어느 부분에 집중할지 알려주는 역할을 한다.
- 매 time step마다 입력 문장을 다시 확인하므로 정보 손실 문제가 줄어든다.

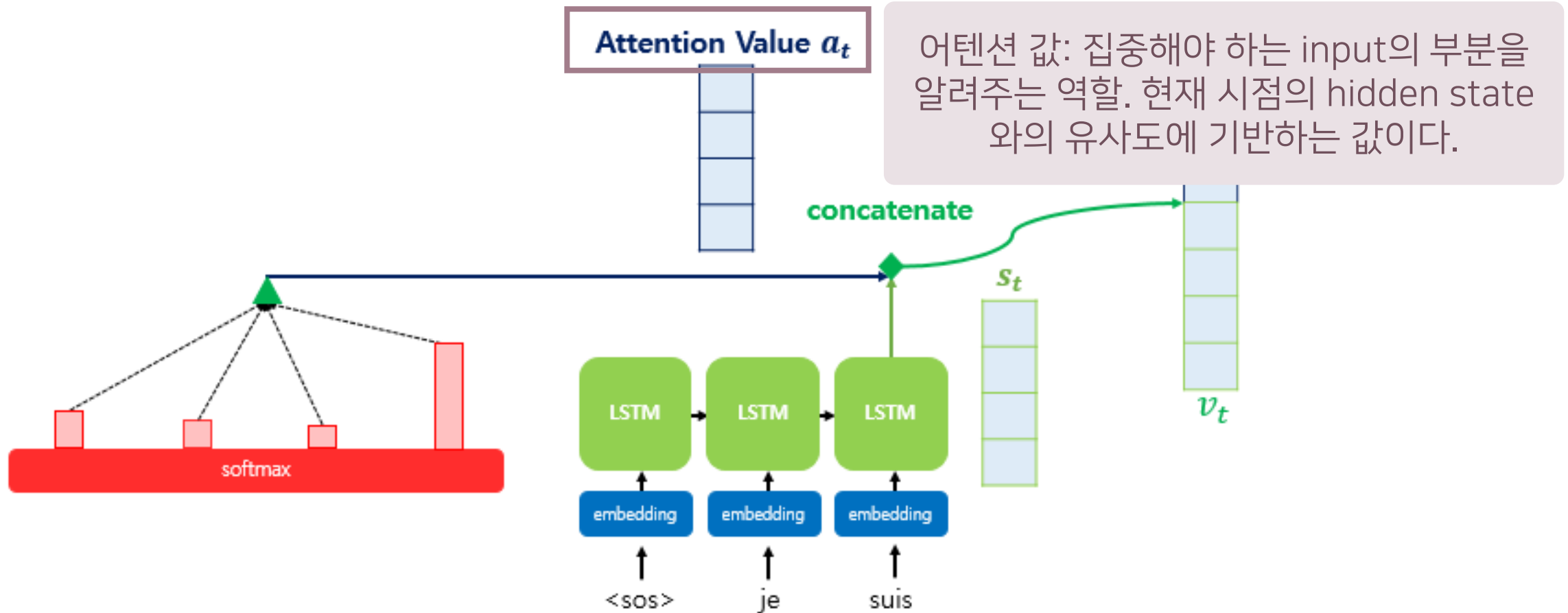


이 사람은 왜 칼을 들고 있나요?



케이크

어텐션 값 구하기



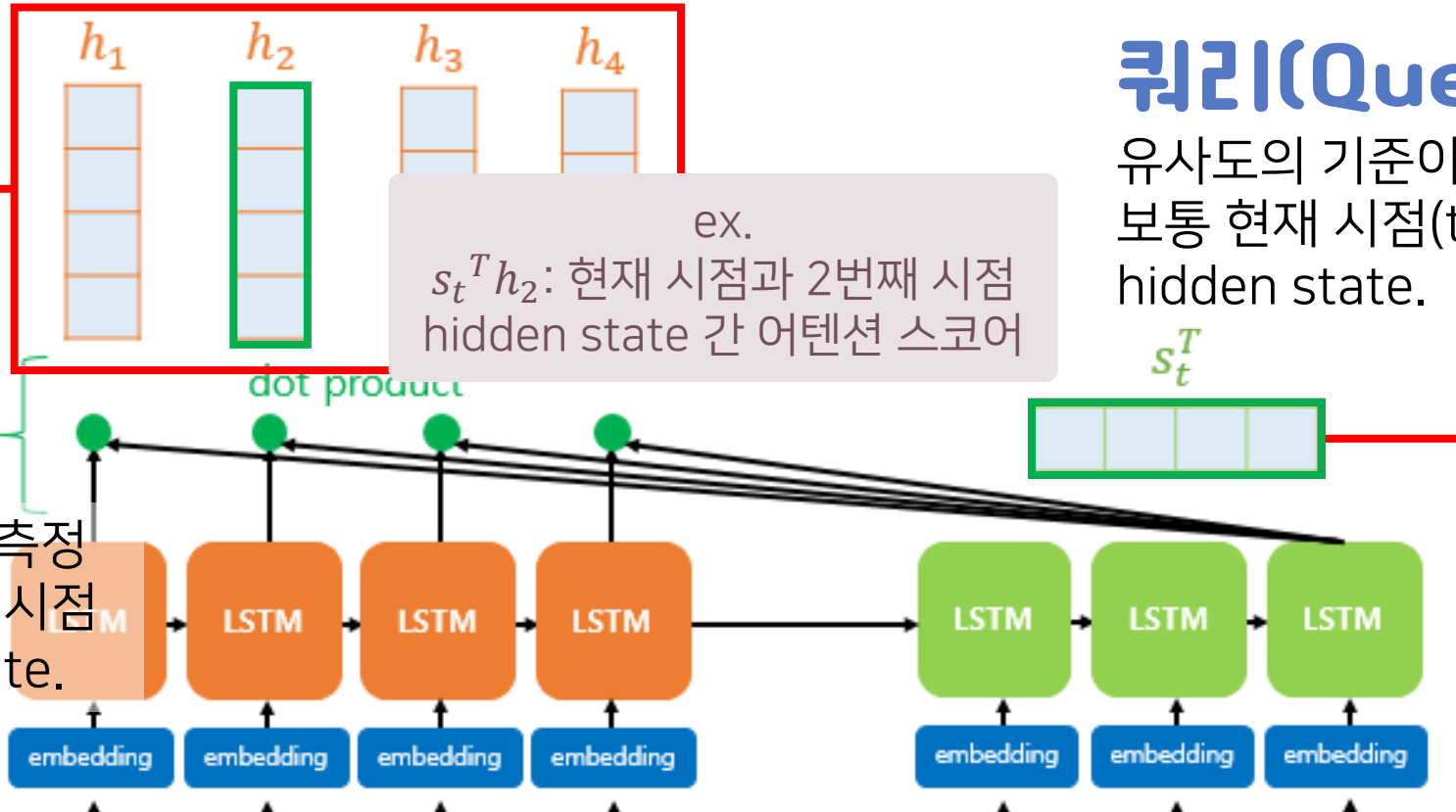
어텐션 매커니즘에서 t 번째 디코더 셀은 $t-1$ 번째 셀의 hidden state, output, 그리고 t 번째 어텐션 값을 입력값으로 받는다.

어텐션 값 구하기

키(Key)

쿼리와 비교하여 유사도를 측정하는 단어(벡터). 보통 모든 시점의 인코더 셀의 hidden state.

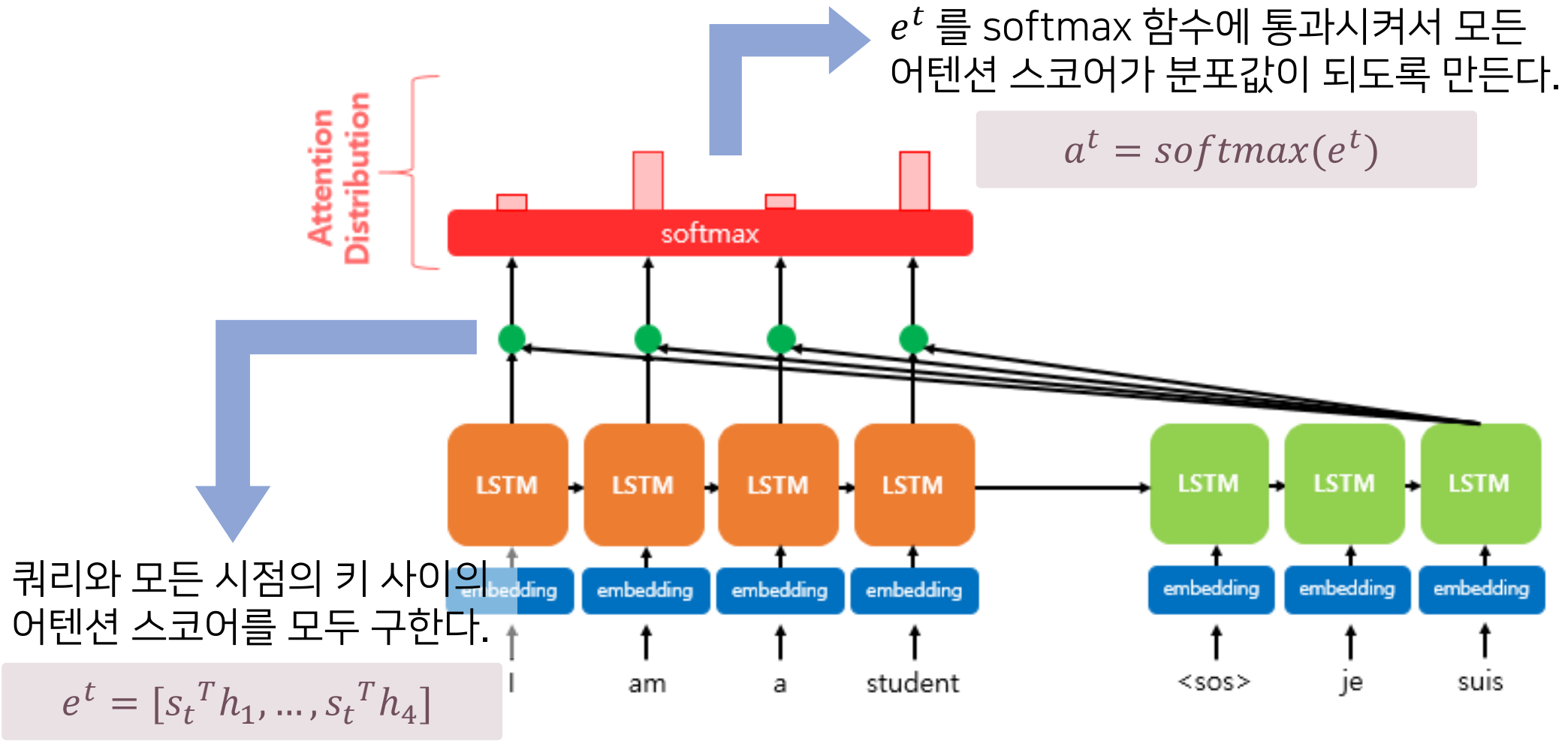
Attention Score



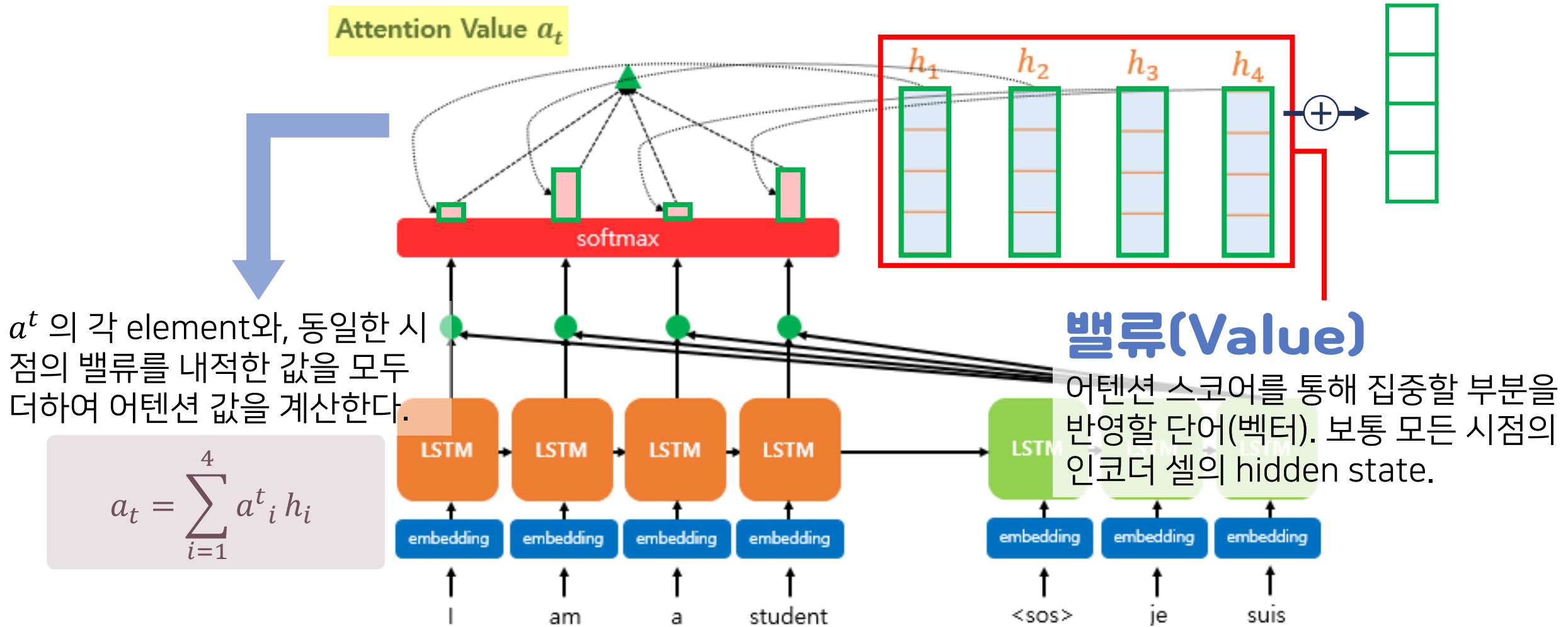
쿼리(Query)

유사도의 기준이 되는 단어(벡터). 보통 현재 시점(t)의 디코더 셀의 hidden state.

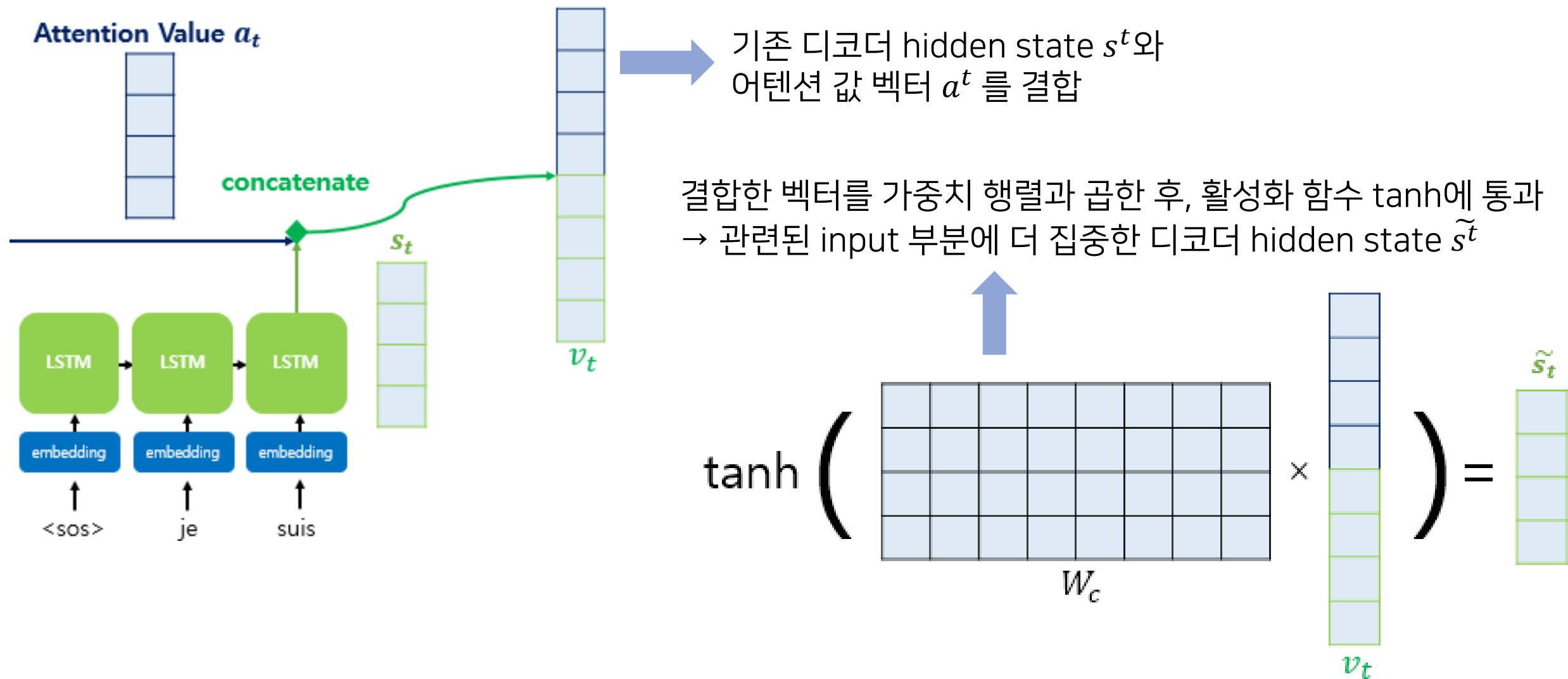
어텐션 값 구하기



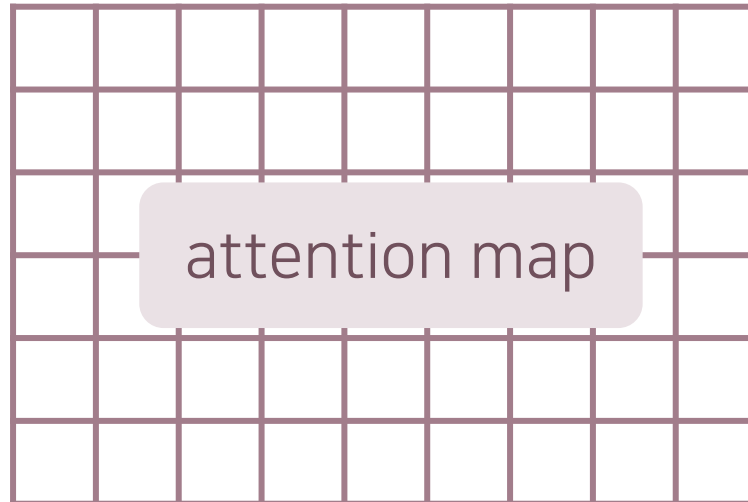
어텐션 값 구하기



디코더에 어텐션 값 반영하기

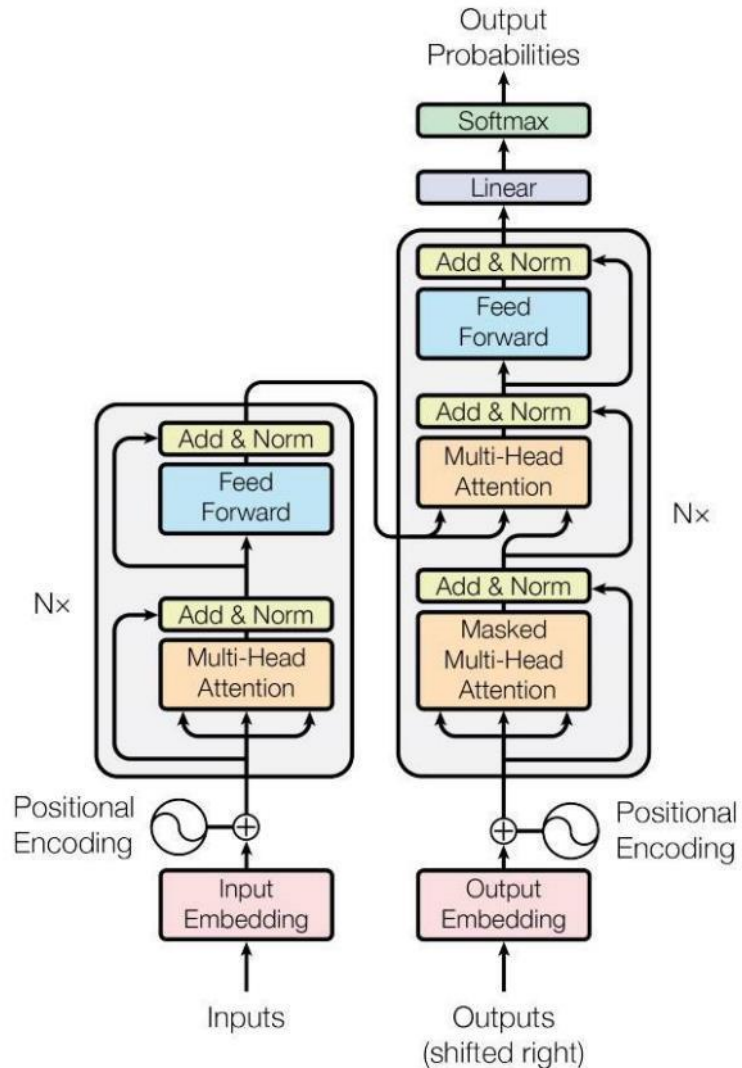


이미지에 어텐션 값 반영하기



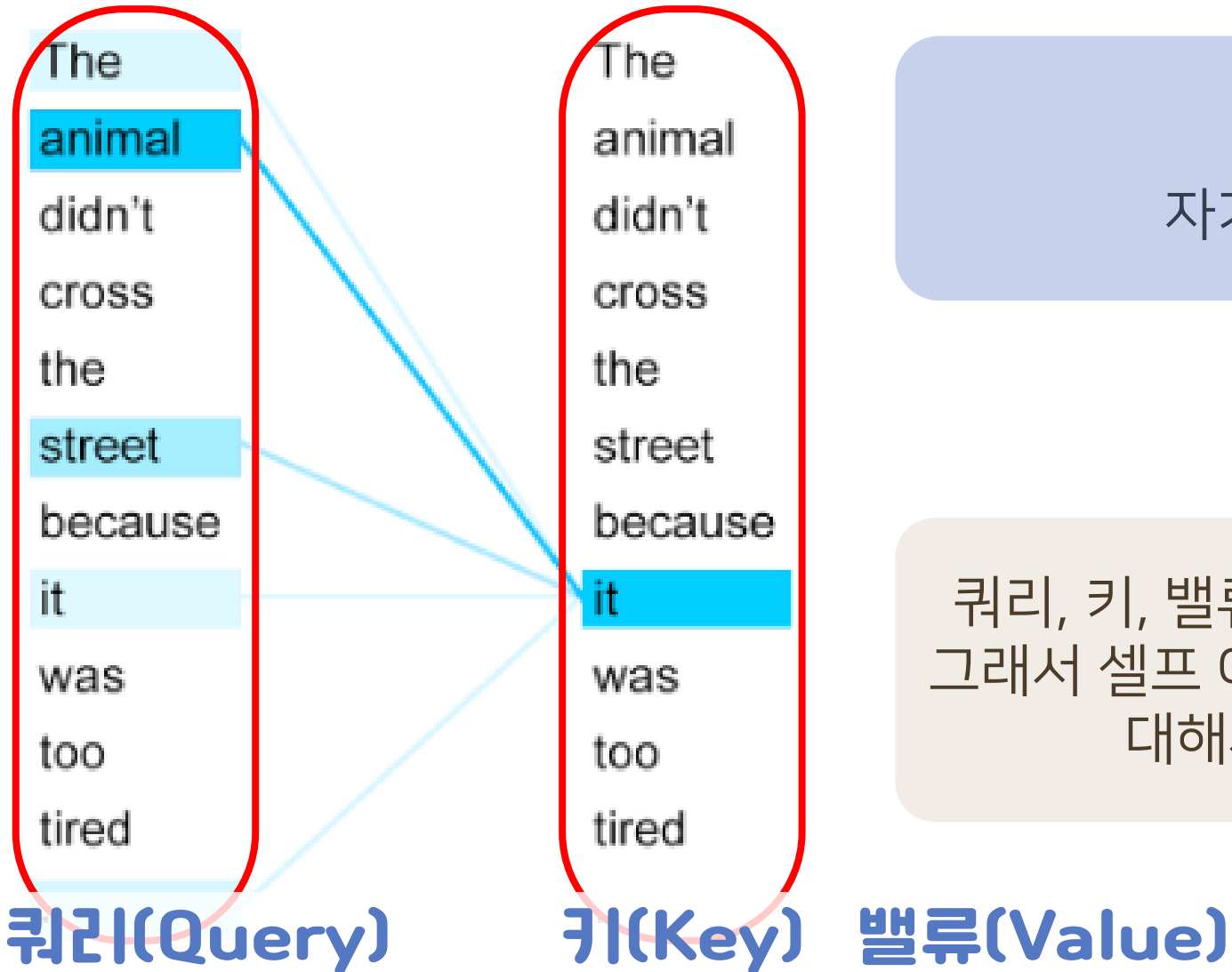
트랜스포머

트랜스포머 모델



- seq2seq 모델처럼 인코더-디코더 구조를 갖고 있으나, RNN을 사용하지 않고 오직 어텐션만 사용하여 만든 모델
- 오직 어텐션만 사용해도 딥러닝 모델을 만들 수 있으며, 오히려 RNN보다 성능이 좋았다는 점에서 큰 관심을 받고 있다.
- 처음에는 NLP 부문에서 등장했으나, 시간이 지나면서 CV 분야까지 진출하였다.

셀프 어텐션

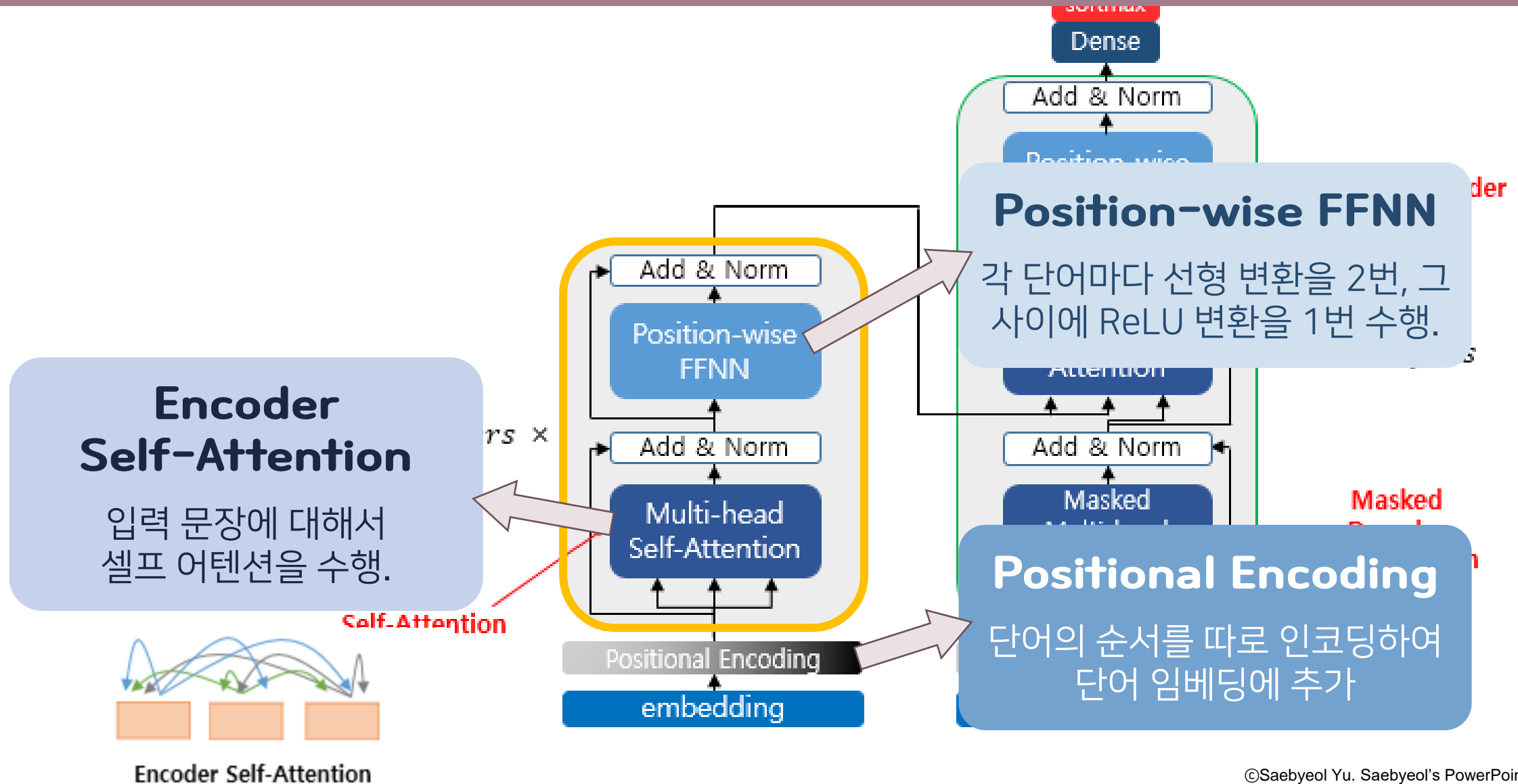


셀프 어텐션

자기자신에게 적용하는 어텐션

쿼리, 키, 밸류 모두 입력 문장의 단어 벡터들이다.
그래서 셀프 어텐션은 모든 시점의 hidden state에
대해서 어텐션 스코어를 계산한다.

인코더



디코더

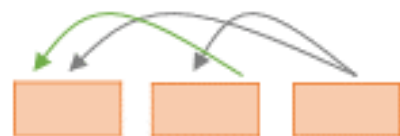
Encoder-Decoder Attention

쿼리는 디코더의 단어 벡터들, 키와 밸류는 인코더의 단어 벡터들인 어텐션 수행.

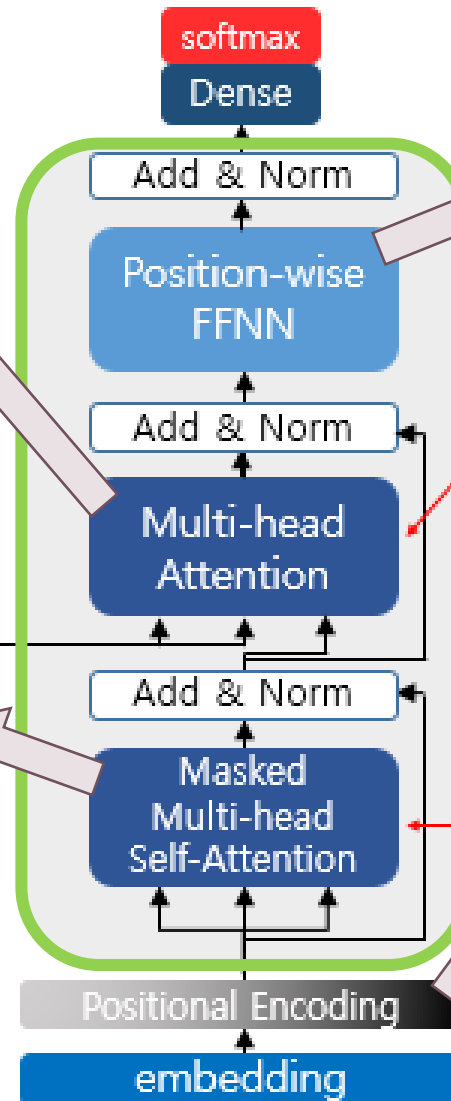
Masked Decoder Self-Attention

입력 문장에 대해서 셀프 어텐션을 수행하되, 현재 시점보다 이후의 단어는 참고하지 못하도록 마스킹

E
Self



Masked Decoder Self-Attention



Position-wise FFNN

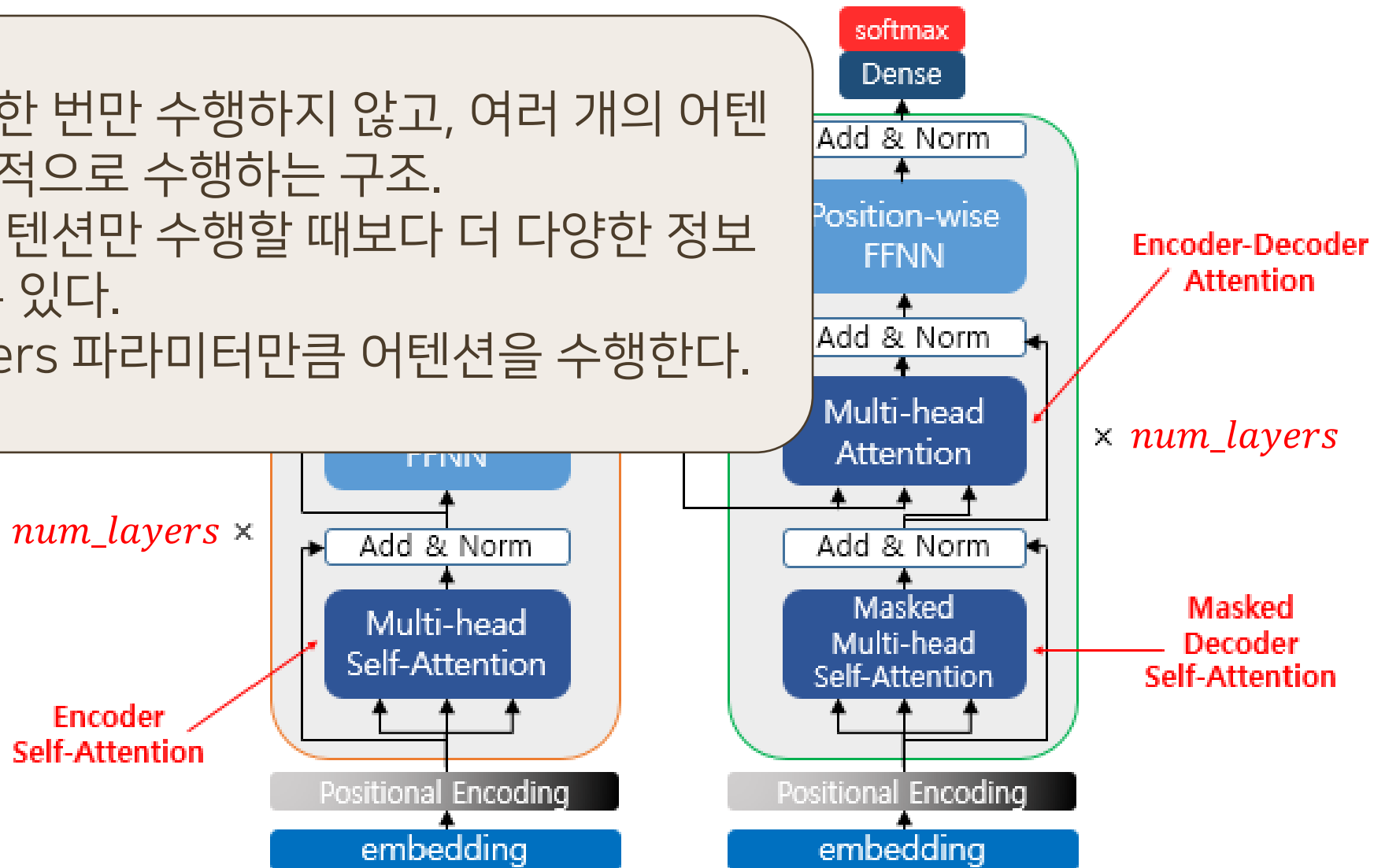
각 단어마다 선형 변환을 2번, 그 사이에 ReLU 변환을 1번 수행.

Positional Encoding

단어의 순서를 따로 인코딩하여 단어 임베딩에 추가

멀티 헤드 어텐션

- 어텐션을 한 번만 수행하지 않고, 여러 개의 어텐션을 병렬적으로 수행하는 구조.
- 한 개의 어텐션만 수행할 때보다 더 다양한 정보를 얻을 수 있다.
- `num_layers` 파라미터만큼 어텐션을 수행한다.



**이것으로 겨울방학 딥러닝 분반 수업 마치겠습니다.
방학동안 정말 수고 많으셨습니다~!!**