자연어처리_seq2seq

리미애오

목차

1. RNN
2. Seq2Seq
3. Seq2Seq with attention

RNN
(Recurrent Neural Network)

백혜림 hreeee@yonsei.ac.kr

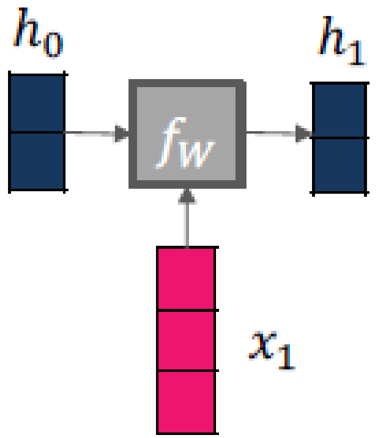
- Sequence형태의 데이터를 뉴럴 네트워크가 학습할 수 있게 하는 것?

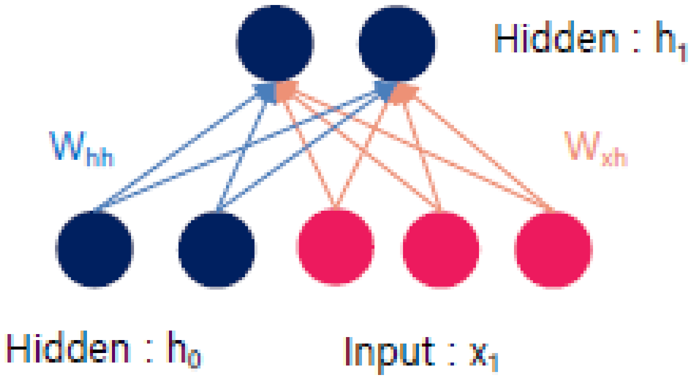
Recurrent Neural Networks

RNN : Computational Graph

$$h_t = f_W(h_{\{t-1\}}, x_t)$$

[가정]
- input : (3, 1)
- Hidden : (2, 1)
- Output : (3, 1)

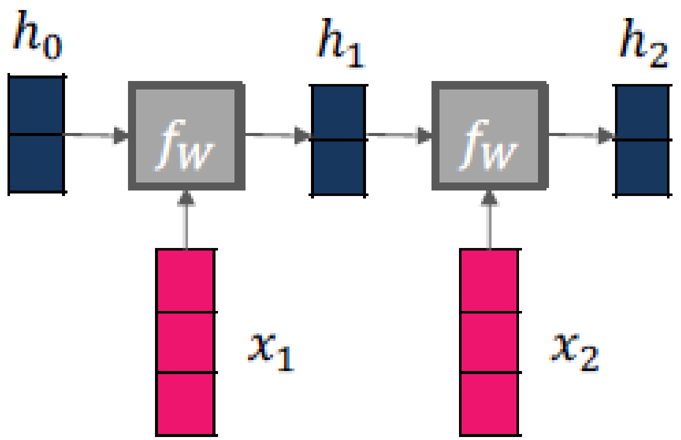


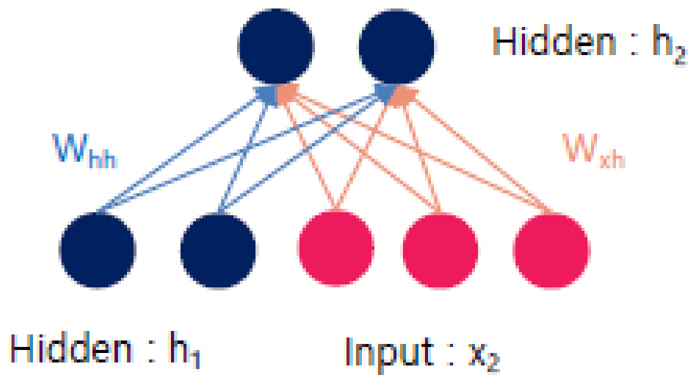


RNN : Computational Graph

$$h_t = f_W(h_{\{t-1\}}, x_t)$$

[가정]
- input : (3, 1)
- Hidden : (2, 1)
- Output : (3, 1)

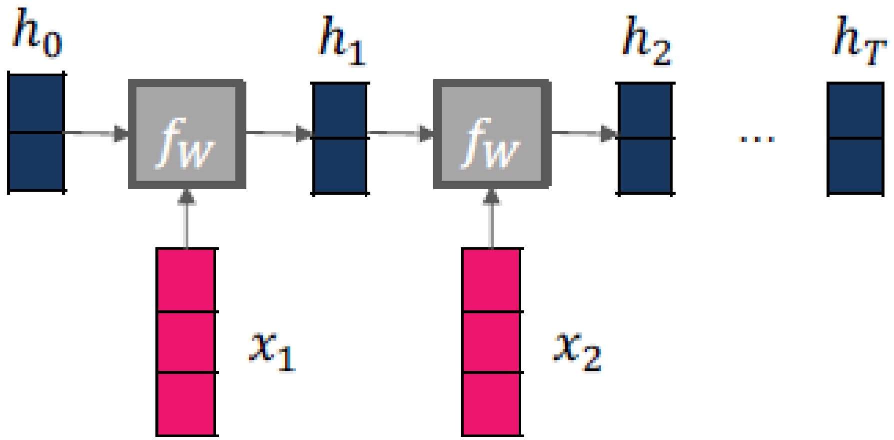




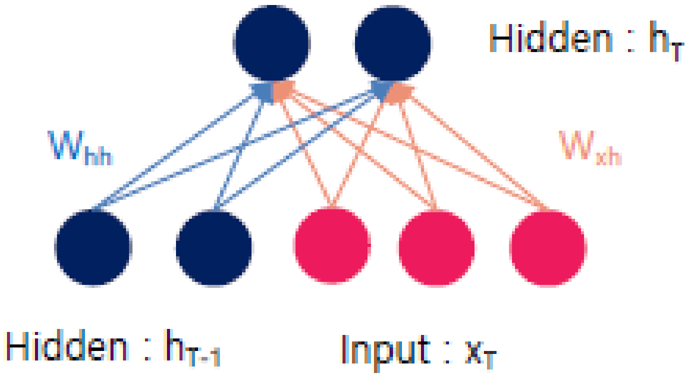
RNN : Computational Graph

$$h_t = f_W(h_{\{t-1\}}, x_t)$$

[가정]
- input : (3, 1)
- Hidden : (2, 1)
- Output : (3, 1)



- f_w는 하나 이다.
- 모든 timestep에서 동일한 W_{hh}, W_{xh}, W_{hy}가 사용 된다.
- 데이터만 바뀌는 것



RNN : 자동완성 예시

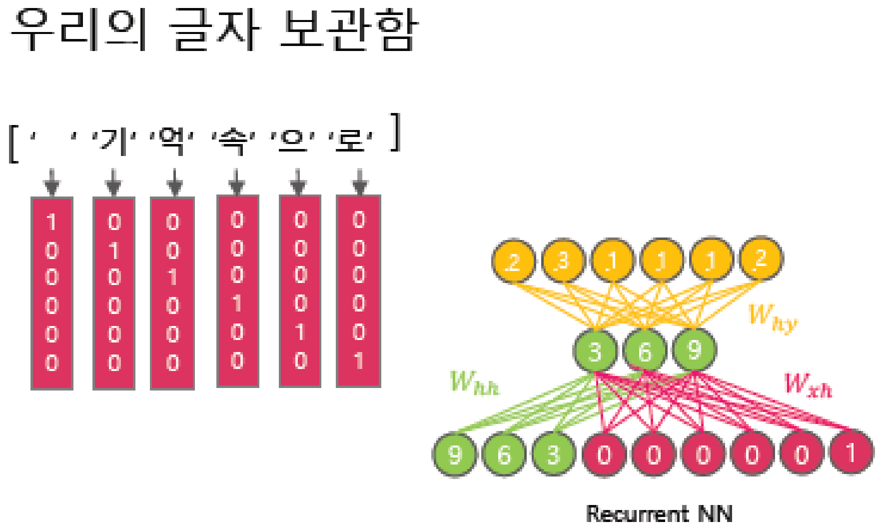
'기' 를 입력 했을 때 자동으로
'기억 속으로 '
가 나오는 모델을 학습 한다고 가정 해 봅시다

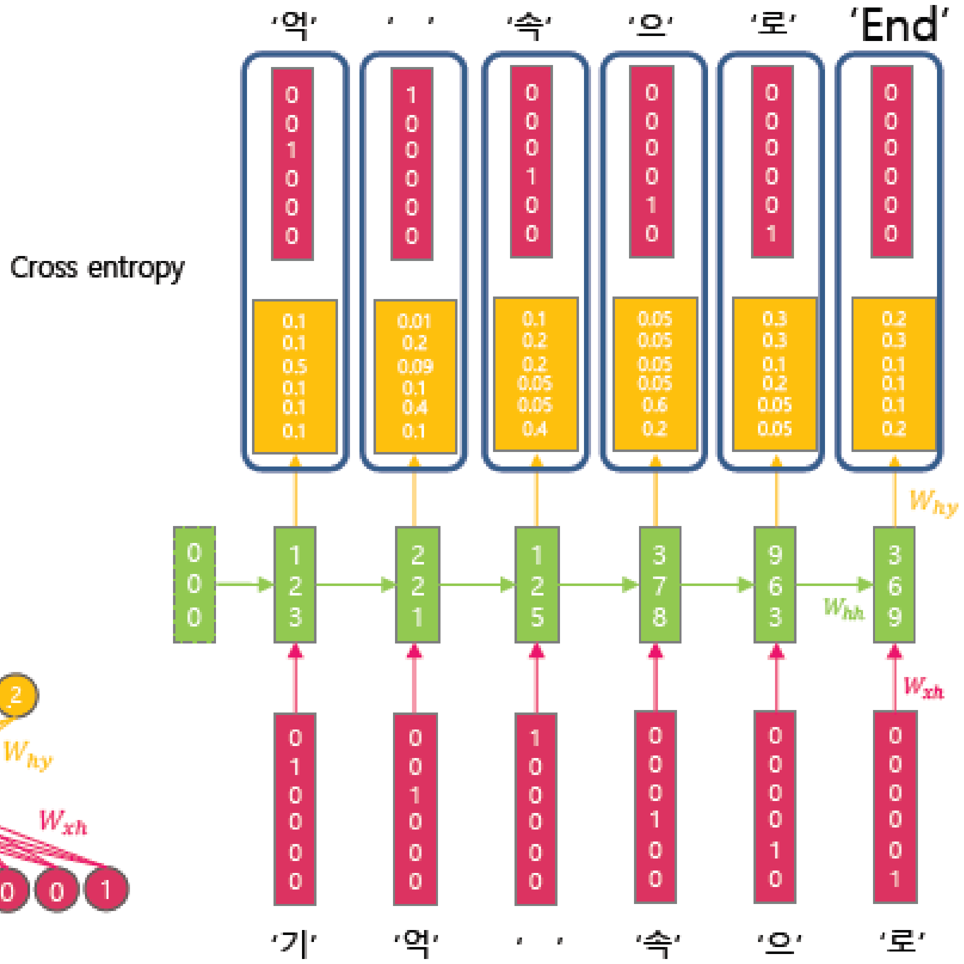
'기' '억' ' ' '속' '으' '로'

단순한 예시를 만들기 위해

한글 글자가 '기', '억', ' ', '속', '으', '로' 만 있다고 가정 해 봅니다

RNN : 자동완성 예시

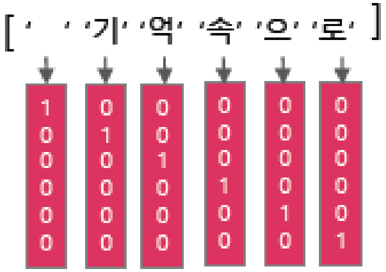


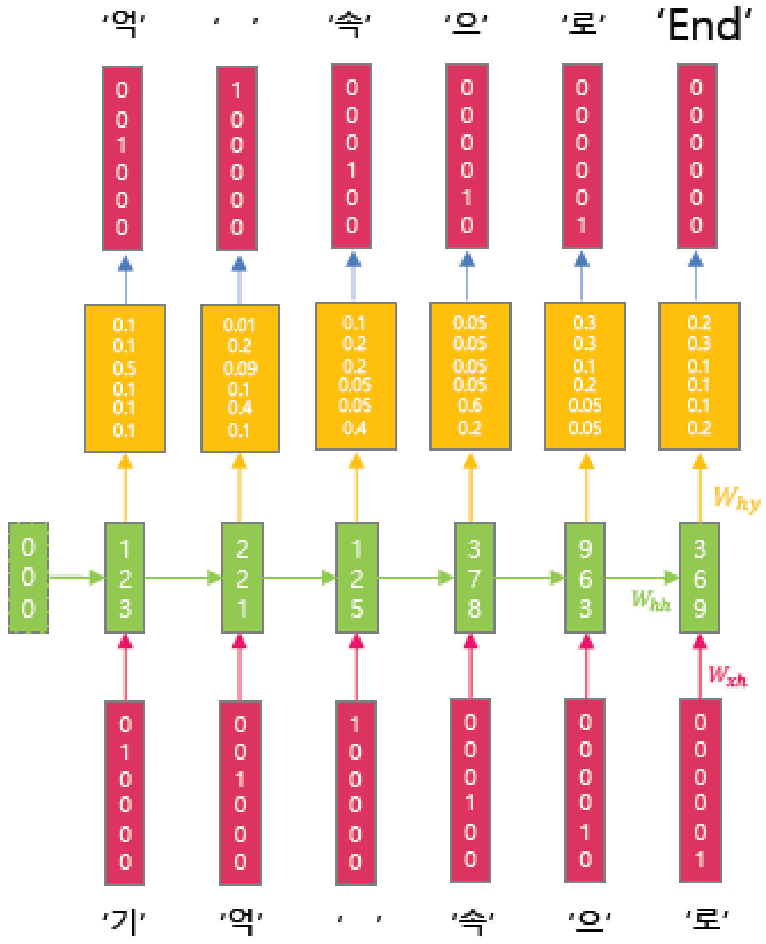


RNN : 자동완성 예시

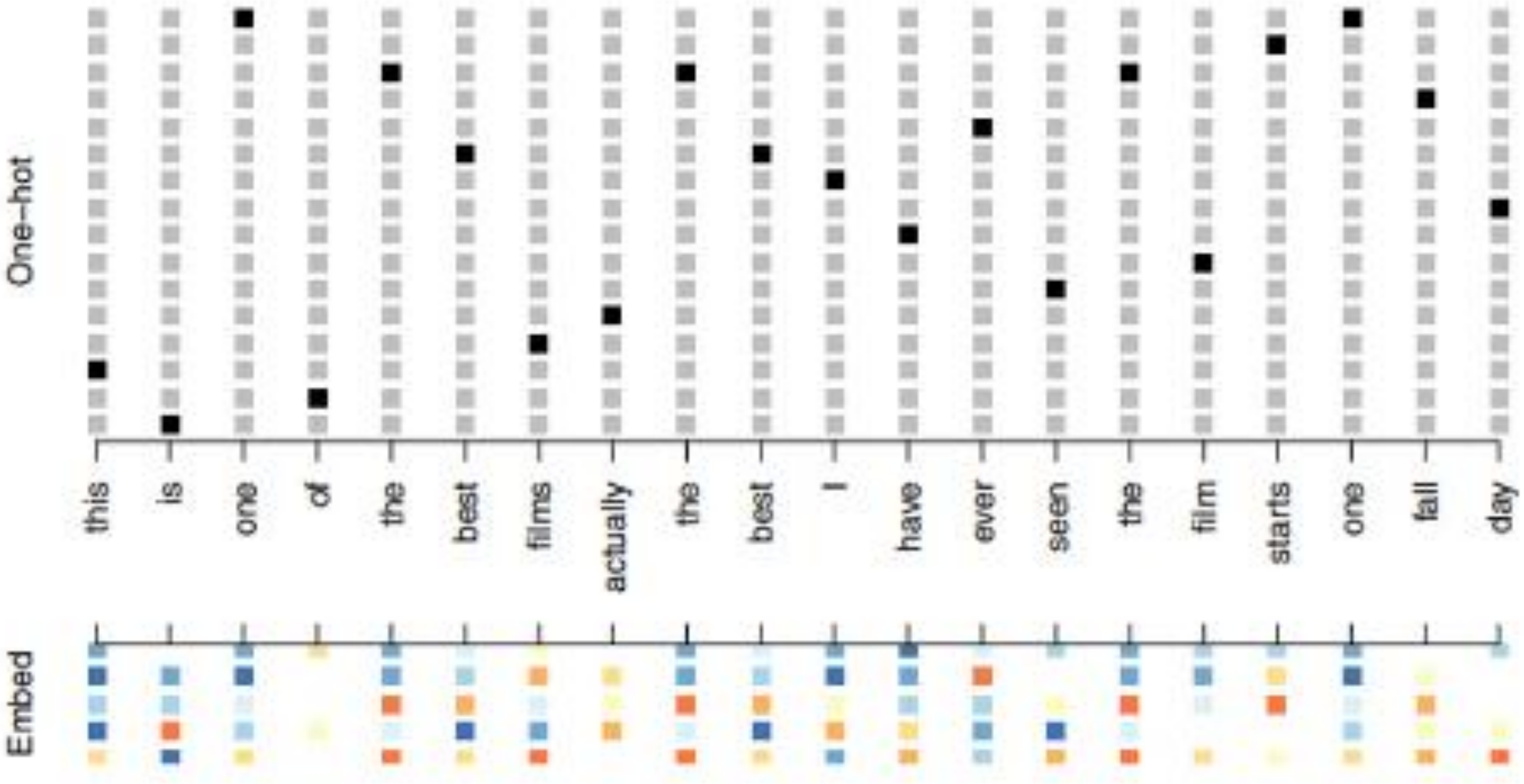
학습 완료 후

우리의 글자 보관함



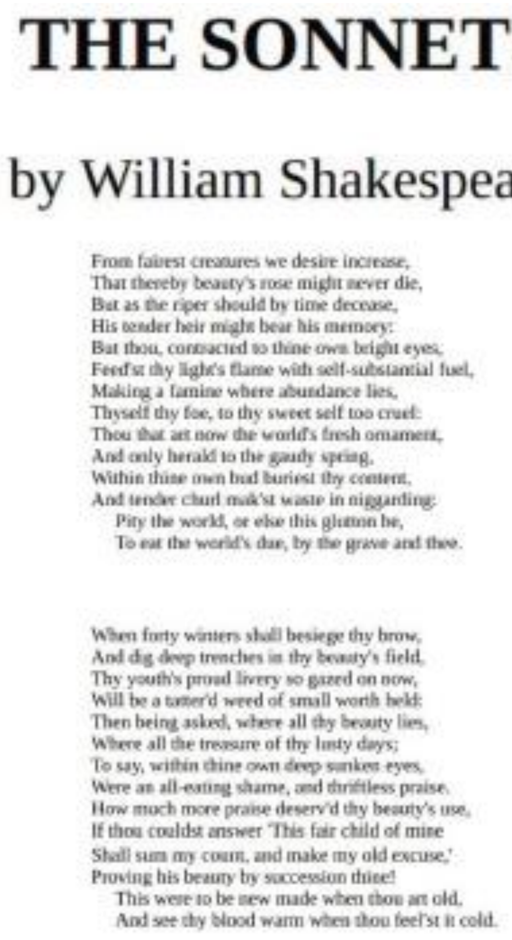


RNN : 자동완성 예시



One-hot
Embed
this
is
one
of
the
best
films
actually
the
best
I
have
ever
seen
the
film
starts
one
fall
day

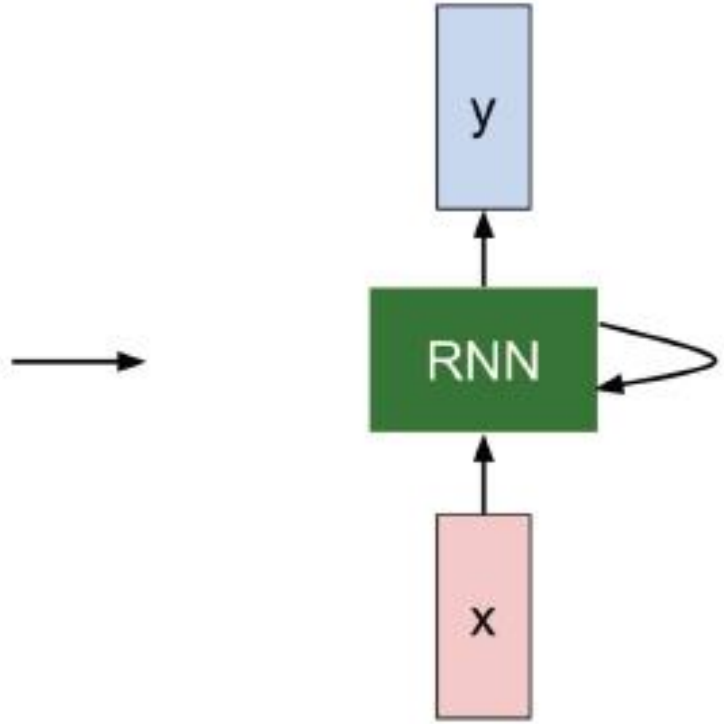
RNN : 소설쓰기


THE SONNETS

by William Shakespeare

From fairest creatures we desire increase,
That thereby beauty's rose might never die,
But as the riper should by time decease,
His tender heir might bear his memory:
But thou, contracted to thine own bright eyes,
Feed'st thy light's flame with self-substantial fuel,
Making a famine where abundance lies,
Thyself thy foe, to thy sweet self too cruel:
Thou that art now the world's fresh ornament,
And only herald to the gaudy spring,
Within thine own bud buriest thy content,
And tender churl mak'st waste in niggarding:
Pity the world, or else this glutton be,
To eat the world's due, by the grave and thee.

When forty winters shall besiege thy brow,
And dig deep trenches in thy beauty's field,
Thy youth's proud livery so gazed on now,
Will be a tatter'd weed of small worth held:
Then being asked, where all thy beauty lies,
Where all the treasure of thy lusty days;
To say, within thine own deep sunken eyes,
Were an all-eating shame, and thriftless praise.
How much more praise deserv'd thy beauty's use,
If thou couldst answer 'This fair child of mine
Shall sum my count, and make my old excuse,'
Proving his beauty by succession thine!
This were to be new made when thou art old,
And see thy blood warm when thou feel'st it cold.



RNN : 소설쓰기

at first:

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

train more

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

train more

```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

train more

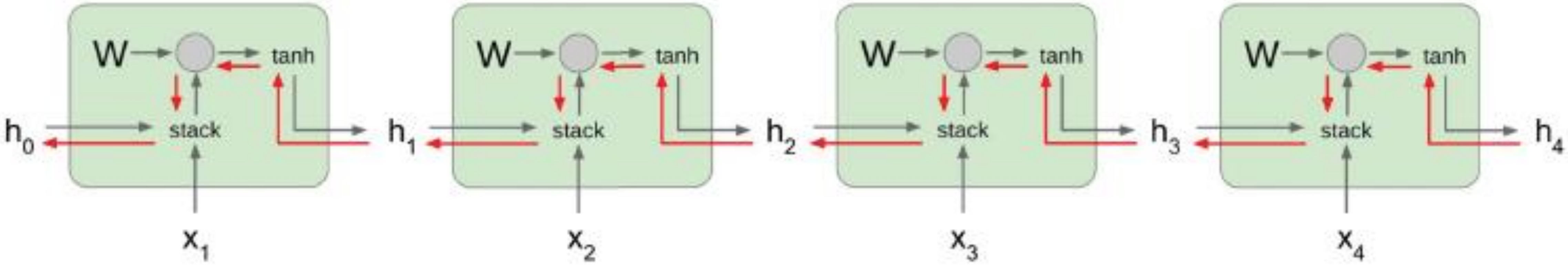
```
"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
```

RNN : 소설쓰기

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

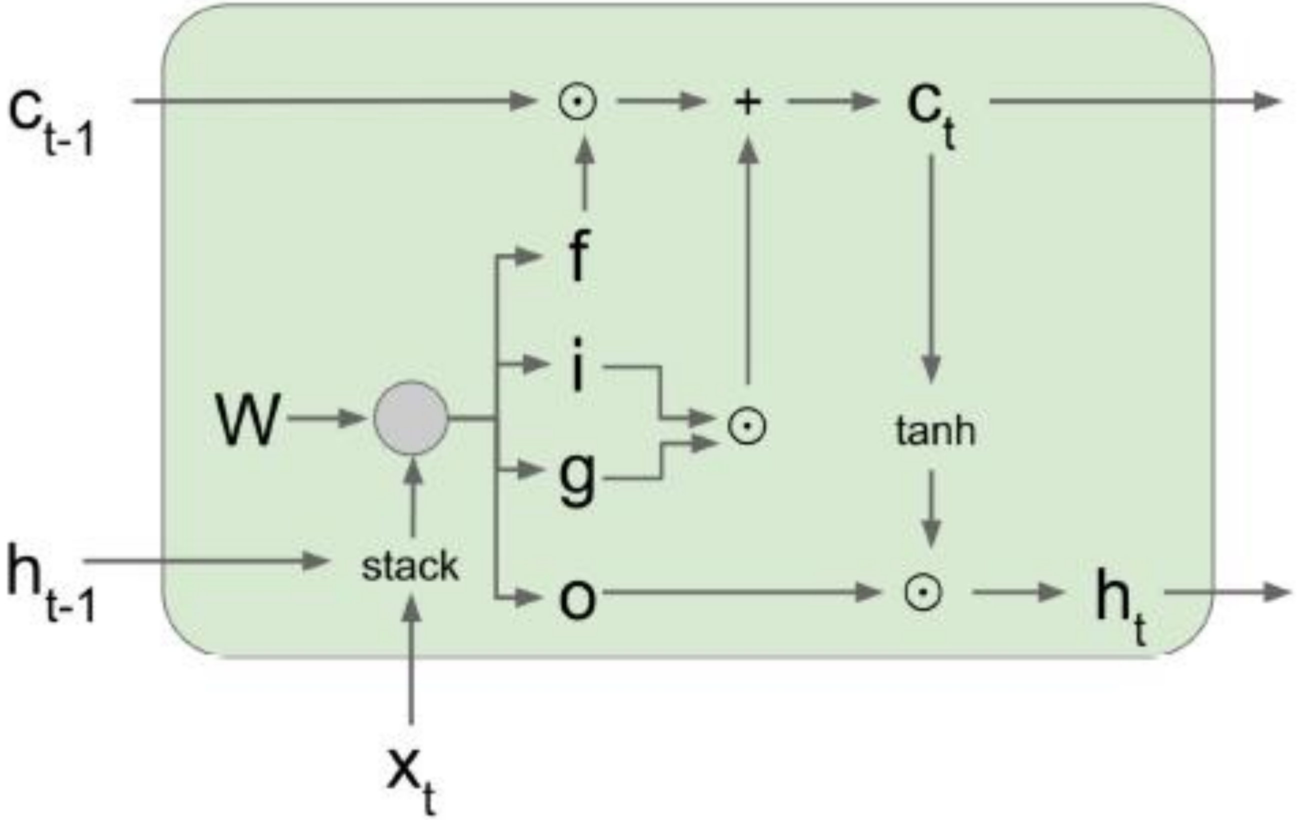
```
VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.
```

RNN Flow



W
tanh
stack
h0
h1
x1
W
tanh
stack
h2
x2
W
tanh
stack
h3
x3
W
tanh
stack
h4
x4

LSTM (Long Short Term Memory) [Hochreiter et al., 1997]

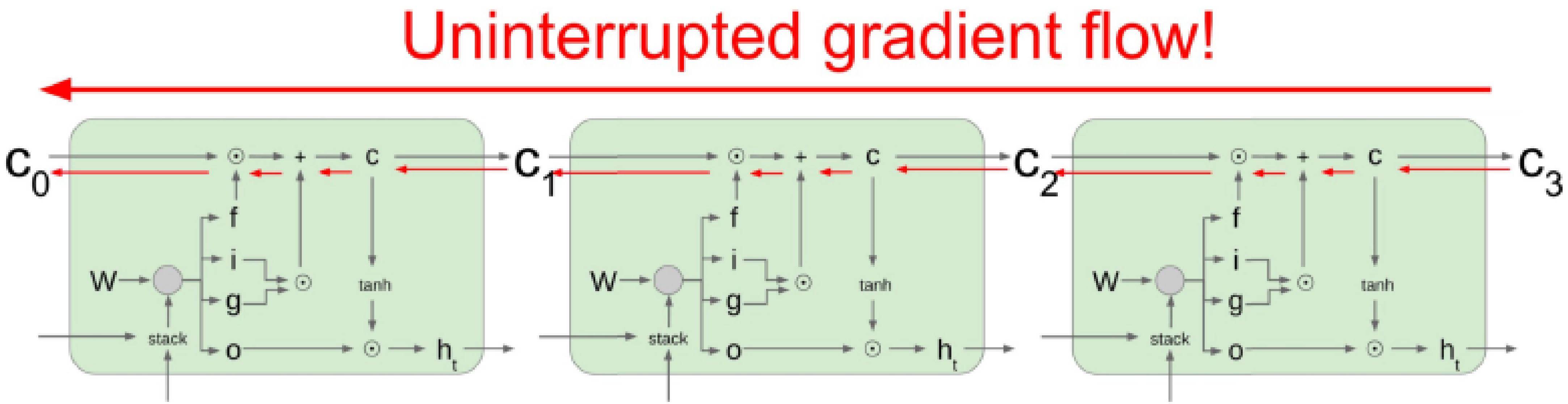


$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot g$$

$$h_t = o \odot \tanh(c_t)$$

LSTM (Long Short Term Memory) [Hochreiter et al., 1997]



Seq2Seq

백혜림 hreeee@yonsei.ac.kr

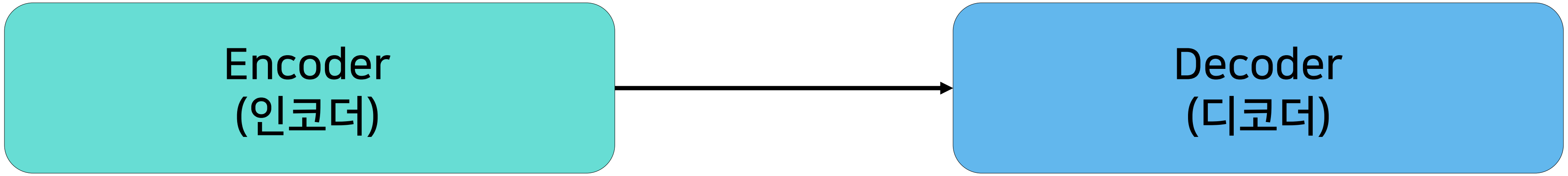
Seq2Seq

- Sequence to Sequence의 약어
- 주어진 시퀀스를 조건(contition)으로 하여 새로운 시퀀스를 만들어내는 작업
- Sequence는 주로 문자로 이루어진 문장을 말하지만, 음성데이터, 한달간의 날씨와 같은 데이터도 뜻한다.

예를 들어

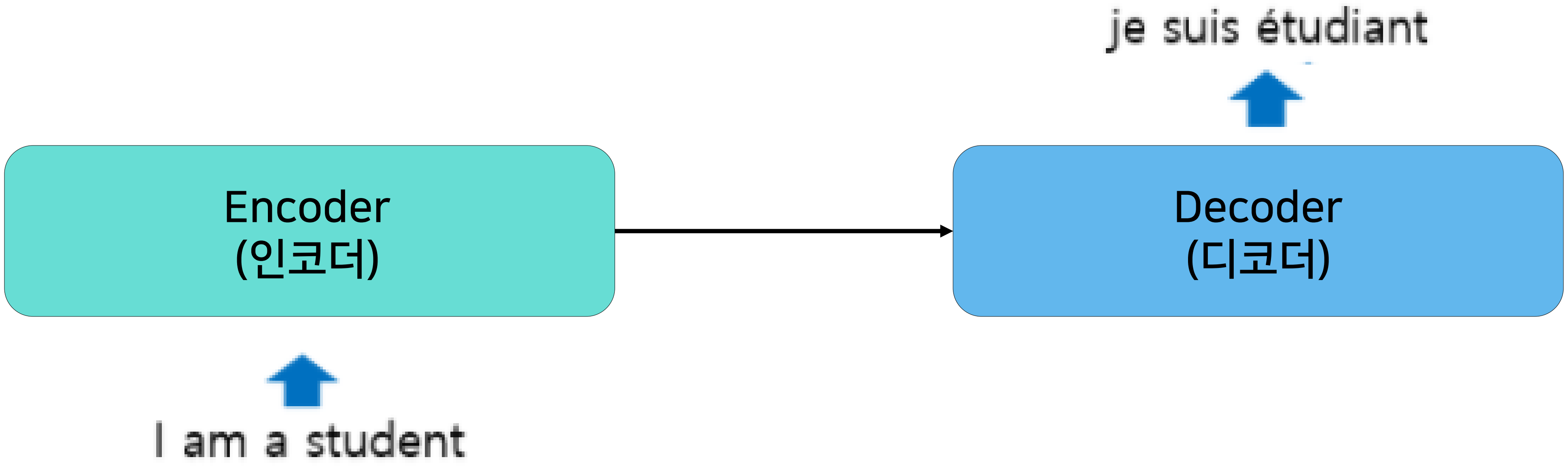
- 번역작업(프랑스어 문장-영어 문장)
- 챗봇 (물음-대답)
- TTS (문장-음성)
- 날씨 예측 (과거 날씨-미래날씨)등의 예를 들 수 있다.

Seq2Seq



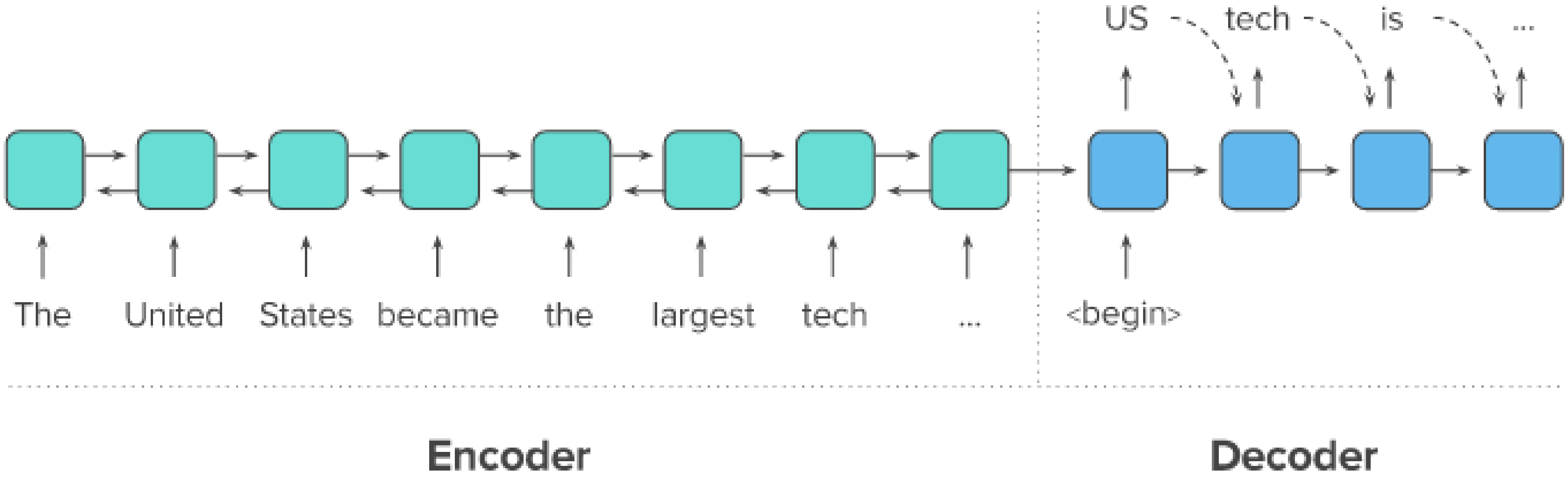
- 번역작업(프랑스어 문장-영어 문장)
- 챗봇 (물음-대답)
- TTS (문장-음성)
- 날씨 예측 (과거 날씨-미래날씨)등의 예를 들 수 있다.

Seq2Seq

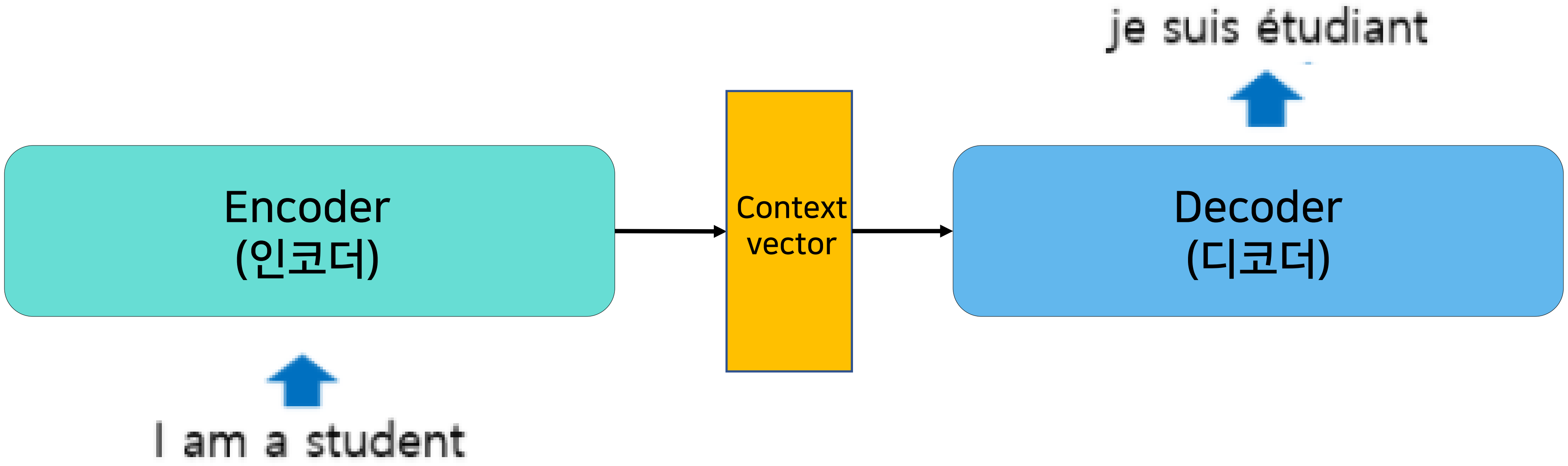


- 번역작업(프랑스어 문장-영어 문장)
- 챗봇 (물음-대답)
- TTS (문장-음성)
- 날씨 예측 (과거 날씨-미래날씨)등의 예를 들 수 있다.

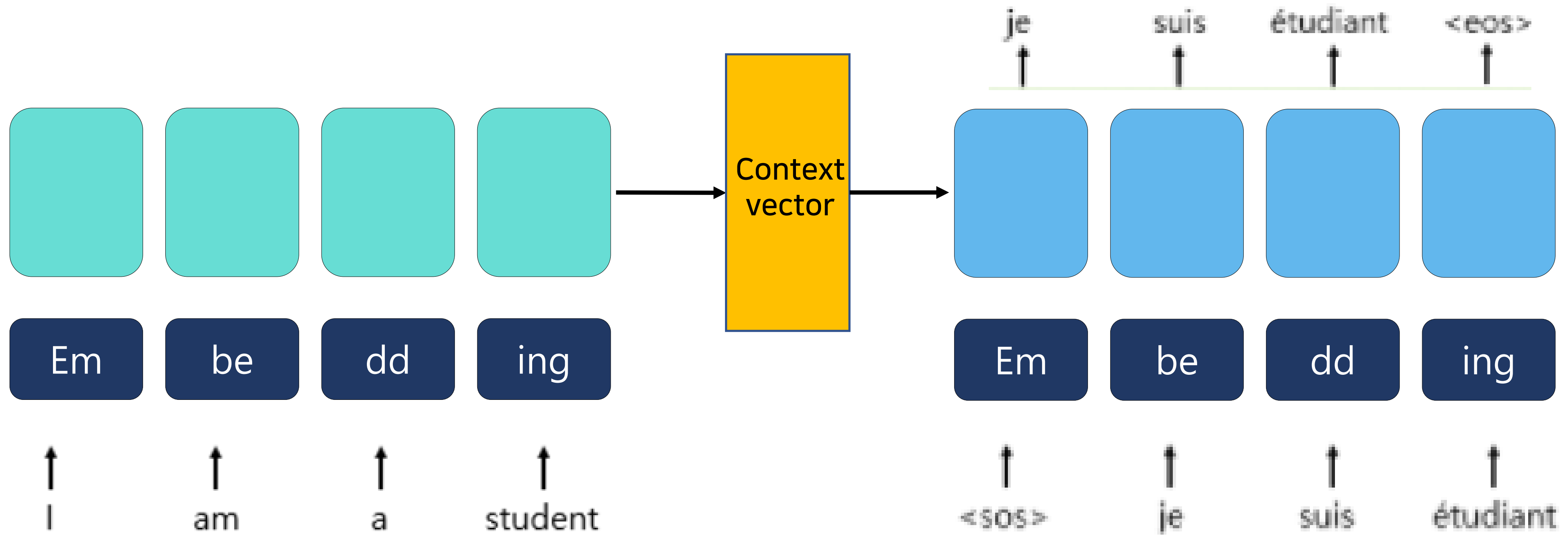
Seq2Seq


US
tech
is
...
The
United
States
became
the
largest
tech
...
<begin>
Encoder
Decoder

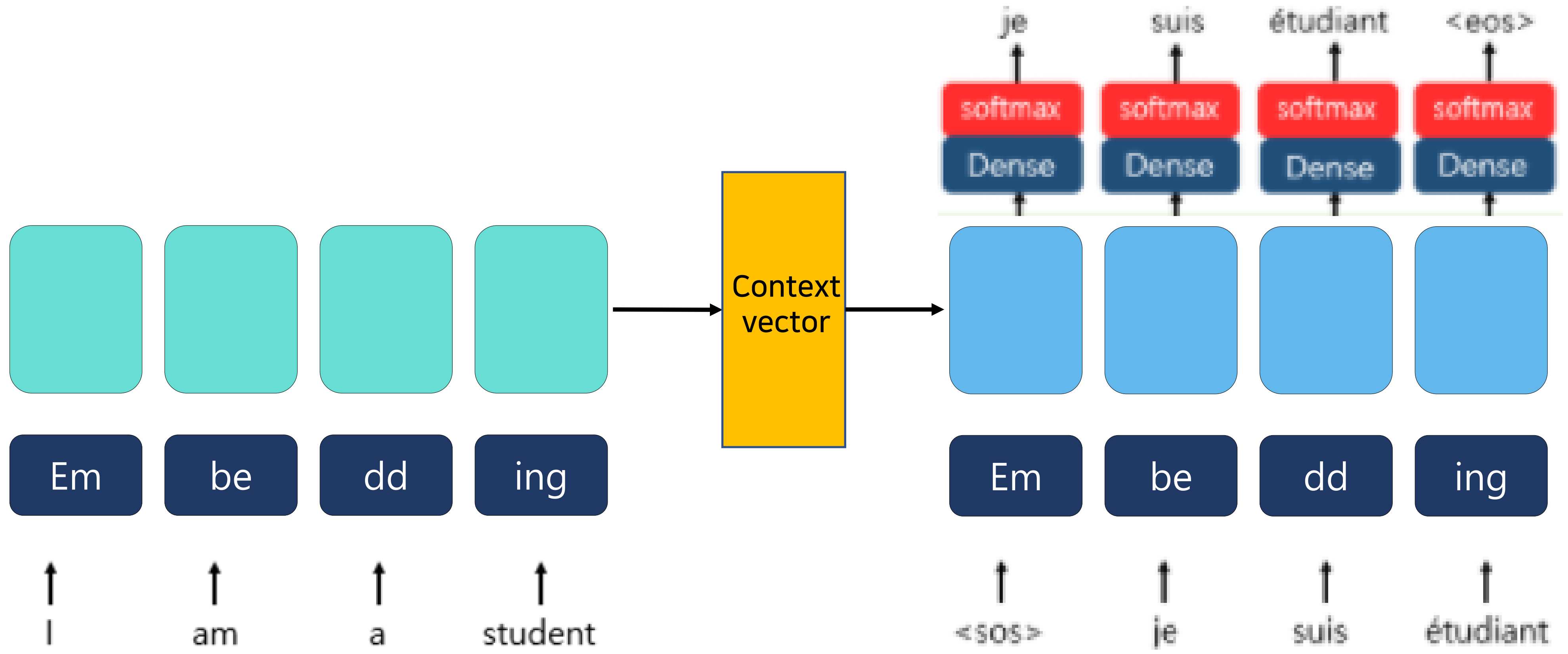
Seq2Seq


je suis étudiant
Encoder
(인코더)
Context
vector
Decoder
(디코더)
I am a student

Seq2Seq


je
suis
étudiant
<eos>
Context vector
Em
be
dd
ing
Em
be
dd
ing
I
am
a
student
<sos>
je
suis
étudiant

Seq2Seq



je
suis
étudiant
<eos>
softmax
softmax
softmax
softmax
Dense
Dense
Dense
Dense
Context vector
Em
be
dd
ing
Em
be
dd
ing
I
am
a
student
<sos>
je
suis
étudiant

Seq2Seq with attention

백혜림 hreeee@yonsei.ac.kr

Seq2seq의 문제점

- 하나의 고정된 크기의 벡터에 모든 정보를 압축하려고 하니까 정보손실이 발생
- RNN의 고질적인 문제인 기울기 소실(Vanishing Gradient)문제가 존재

Attention

- Encoder에 의해 encoding된 정보를 효과적으로 decoder에서 사용할 수 있도록 attention이라는 메커니즘을 도입

Attention Idea

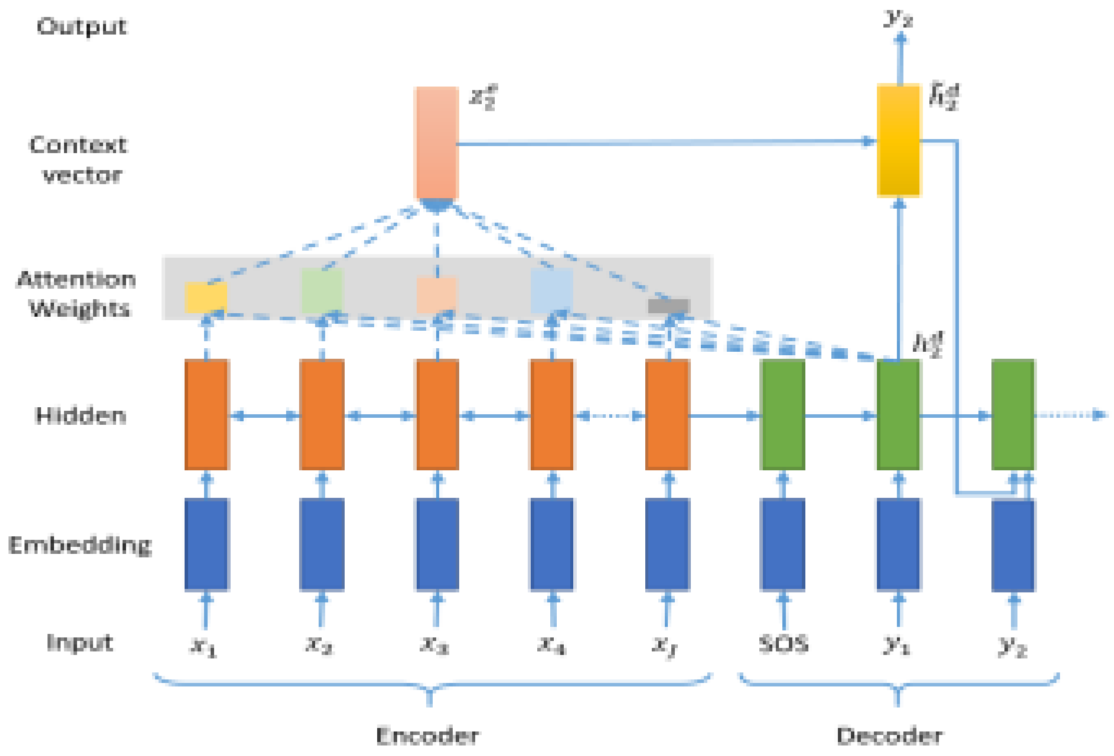
- 디코더에서 출력 단어를 예측하는 매 시점(time step)마다, 인코더에서의 전체 입력문장을 다시 한 번 참고한다는 점
- 전체 입력 문장을 전부 다 동일한 비율로 참고하는 것이 아니라, 해당 시점에서 예측해야할 단어와 연관이 있는 입력 단어 부분을 좀 더 집중(attention)해서 보게 되는 것

Attention

1. Decoder의 한 step의 hidden vector와 encoder의 모든 hidden vector들 간에 어떠한 연산을 수행하여 attention weights를 만든다.

2. Attention weights를 비율로 하여 encoder의 hidden vector들을 weighted sum하여 context vector를 만들어 낸다.

3. Context vector를 decoder의 hidden vector와 concat하여 최종 output을 하기 위해 사용한다.



Attention

Attention을 하기 위한 energy e_{ij}는 다음과 같이 계산합니다.

$$e_{ij} = a\left(s_{i-1}, h_j\right) = v_a^{\mathsf{T}} \tanh\left(W_a s_{i-1} + U_a h_j\right)$$

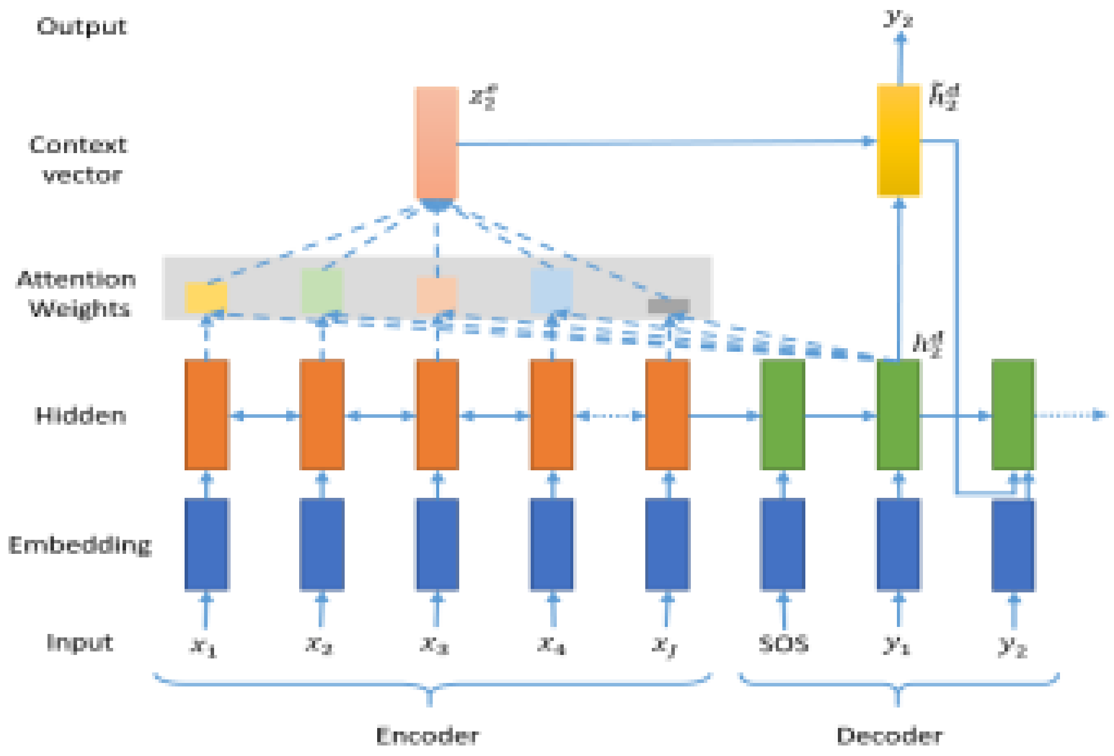
s_{i-1}: i-1시점의 decoder hidden state
h_j: j시점의 encoder hidden state
v_a^T, W_a, U_a: trainable parameters

Attention weights는 energy e_{ij}를 softmax 연산을하여 확률분포의 형태를 만들어 구합니다.

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_k \exp\left(e_{ik}\right)}$$

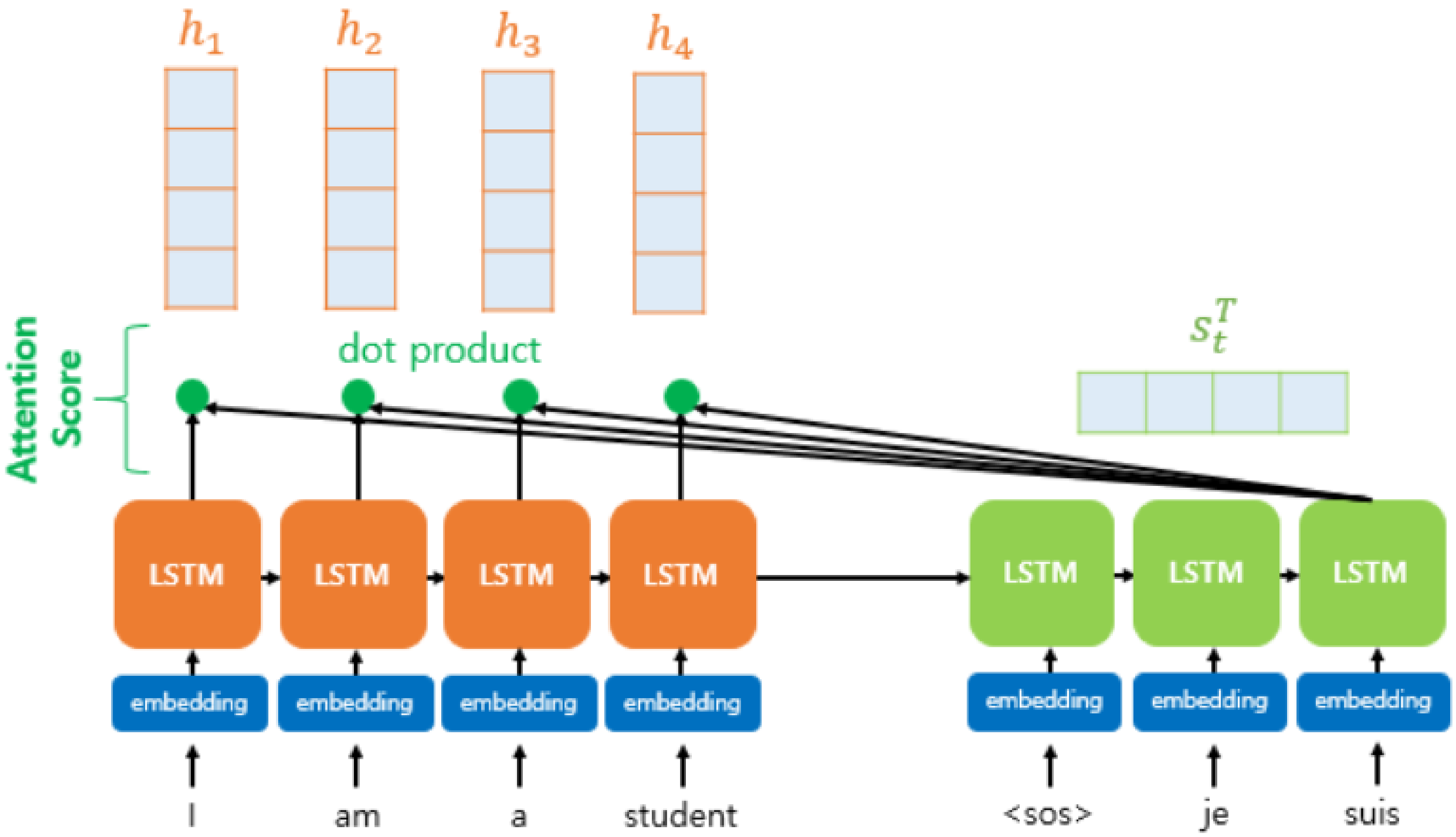
마지막으로 context vector는 attention weight α_{ij}와 encoder hidden state들을 weighted sum하여 구한다.

$$c_i = \sum_j \alpha_{ij} h_j$$

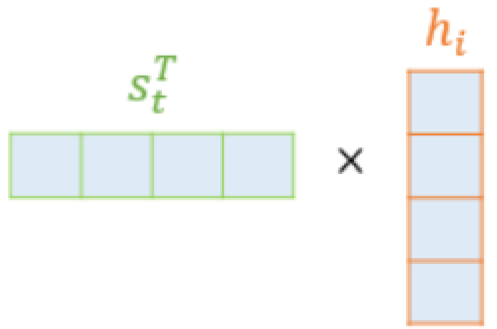


Attention

1) 어텐션 스코어(Attention Score)를 구한다.



$$score(s_t,\ h_i) = s_t^T h_i$$

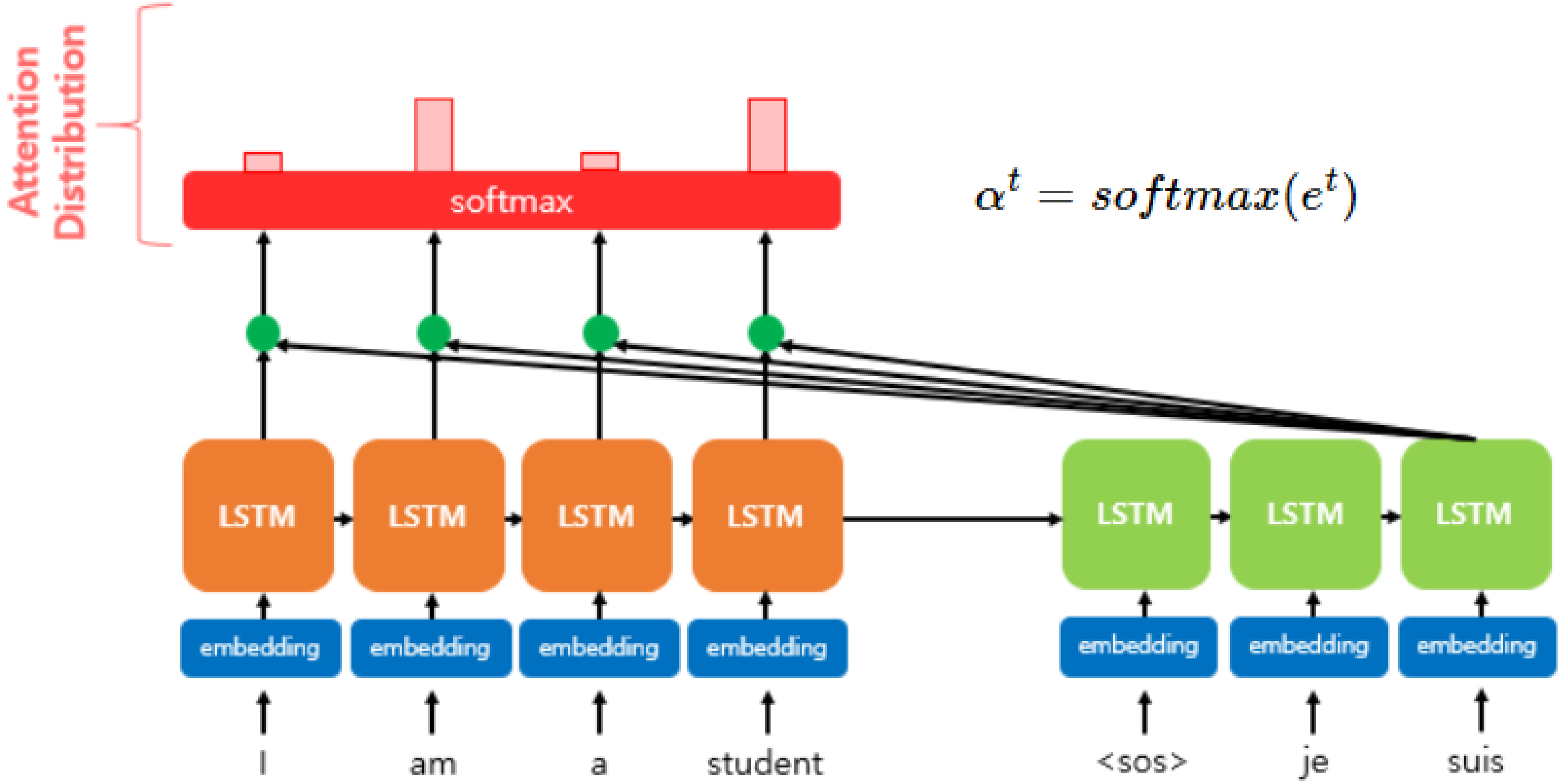


어텐션 스코어의 모음집

$$e^t = [s_t^T h_1, \ldots, s_t^T h_N]$$

Attention

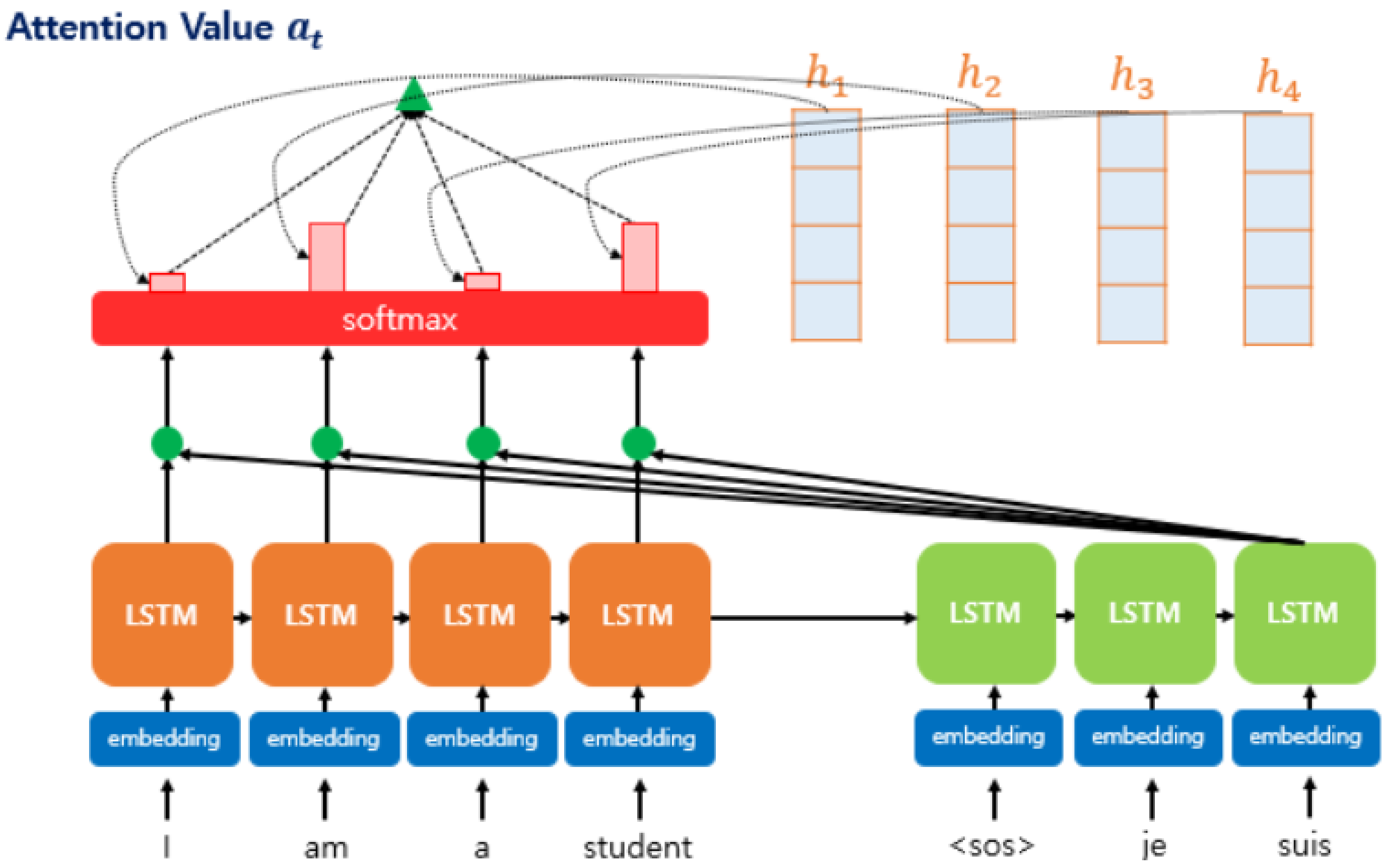
2) 소프트맥스(softmax) 함수를 통해 어텐션 분포(Attention Distribution)를 구한다.



$$\alpha^t = softmax(e^t)$$

Attention

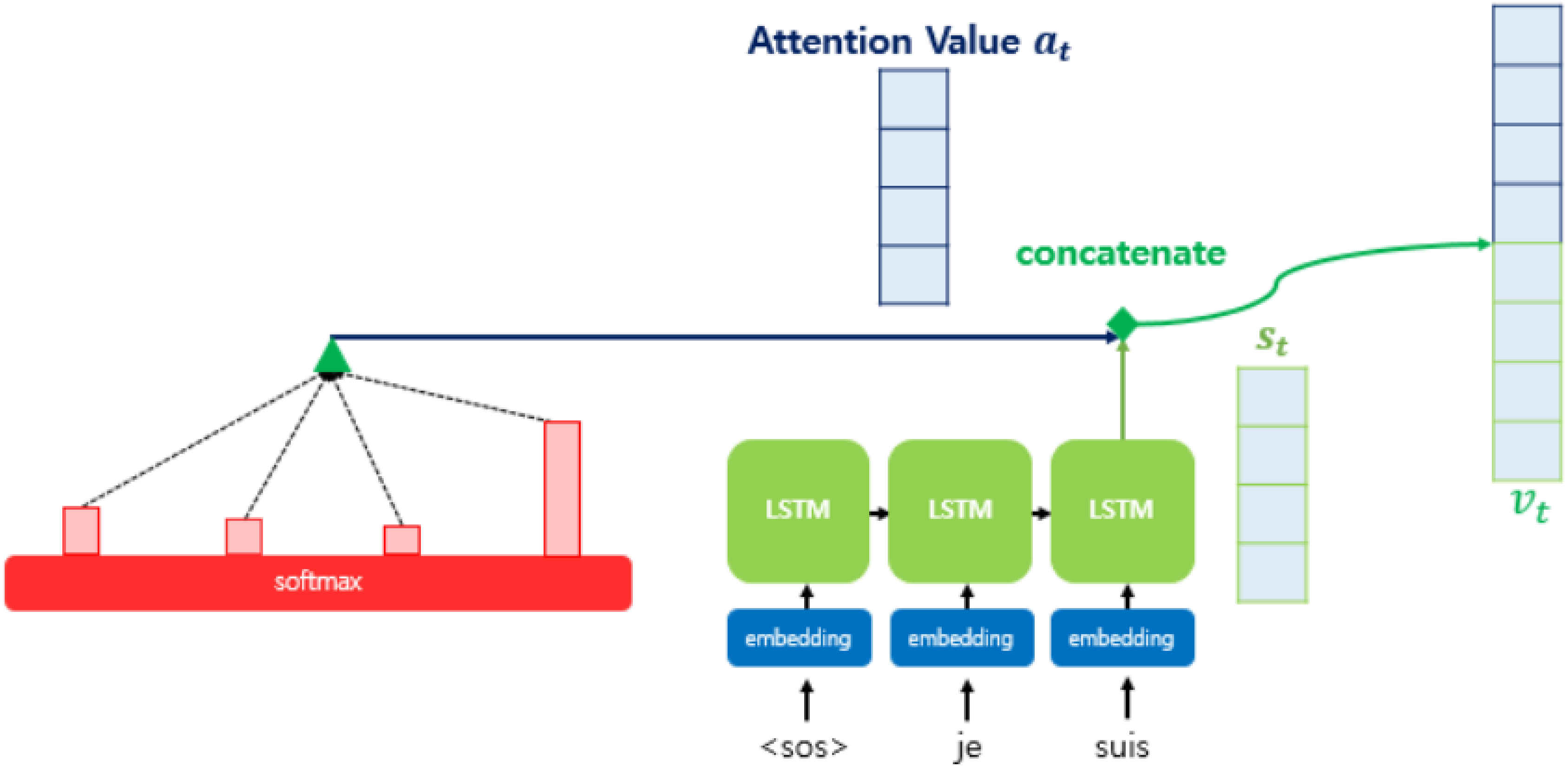
3) 각 인코더의 어텐션 가중치와 은닉 상태를 가중합하여 어텐션 값(Attention Value)을 구한다.



$$a_t = \sum_{i=1}^{N} \alpha_i^t h_i$$

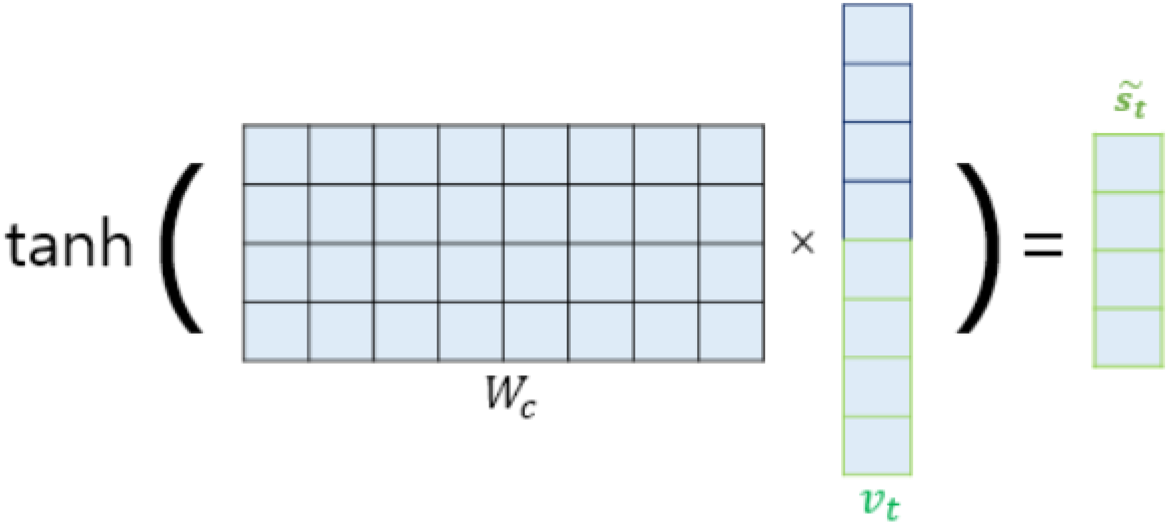
Attention

4) 어텐션 값과 디코더의 t 시점의 은닉 상태를 연결한다.(Concatenate)



Attention

5) 출력층 연산의 입력이 되는 $\tilde{s}_t$를 계산합니다.



$$\tilde{s}_t = \tanh\left(\mathbf{W_c}[a_t; s_t] + b_c\right)$$

Attention

6) $\tilde{s}_t$를 출력층의 입력으로 사용합니다.

$\tilde{s}_t$를 출력층의 입력으로 사용하여 예측 벡터를 얻습니다.

$$\hat{y}_t = \mathrm{Softmax}\left(W_y \tilde{s}_t + b_y\right)$$

수고하셨습니다.