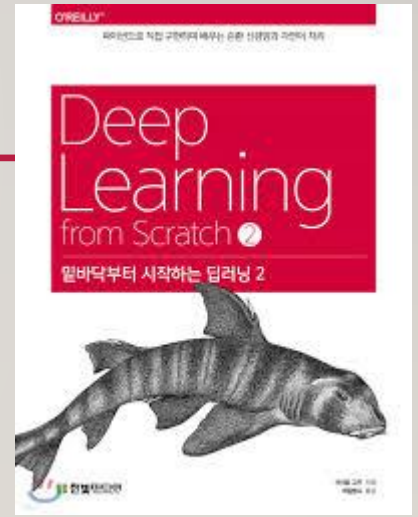
밑바닥부터 시작하는 딥러닝2



7장. RNN을 사용한 문장 생성

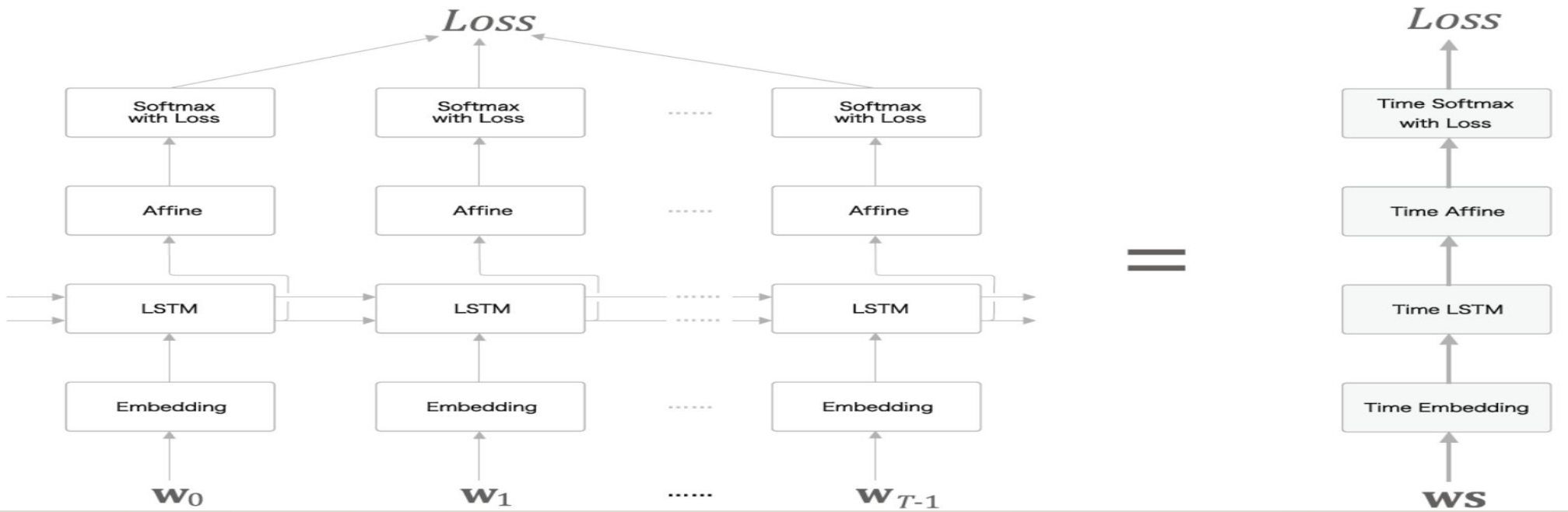
8장. 어텐션

AI융합학부 20193124 고경빈

목차

REVIEW

그림 7-1 앞 장에서 구현한 언어 모델: 오른쪽은 시계열 데이터를 한꺼번에 처리하는 Time 계층을 사용했고, 왼쪽은 같은 구성을 펼친 모습



RNN을 사용한 문장 생성 순서

그림 7-2 언어 모델은 다음에 출현할 단어의 확률분포를 출력한다.

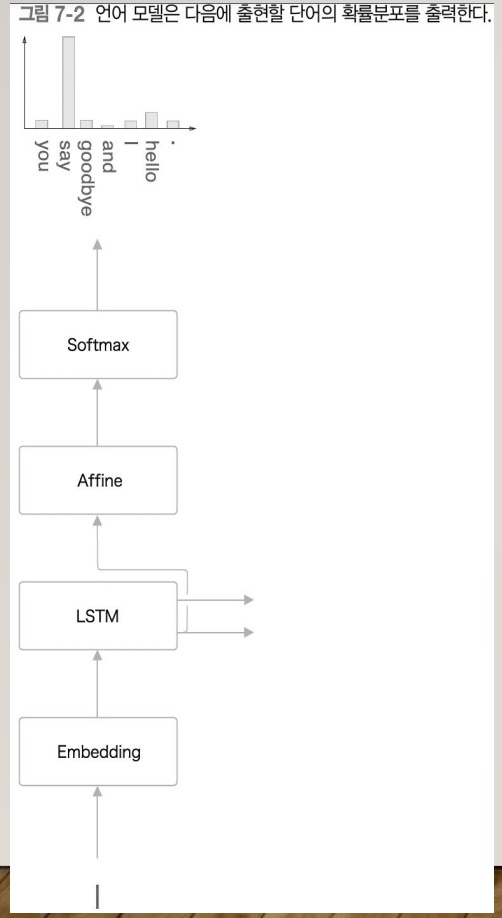


그림 7-3 확률분포대로 단어를 하나 샘플링한다.

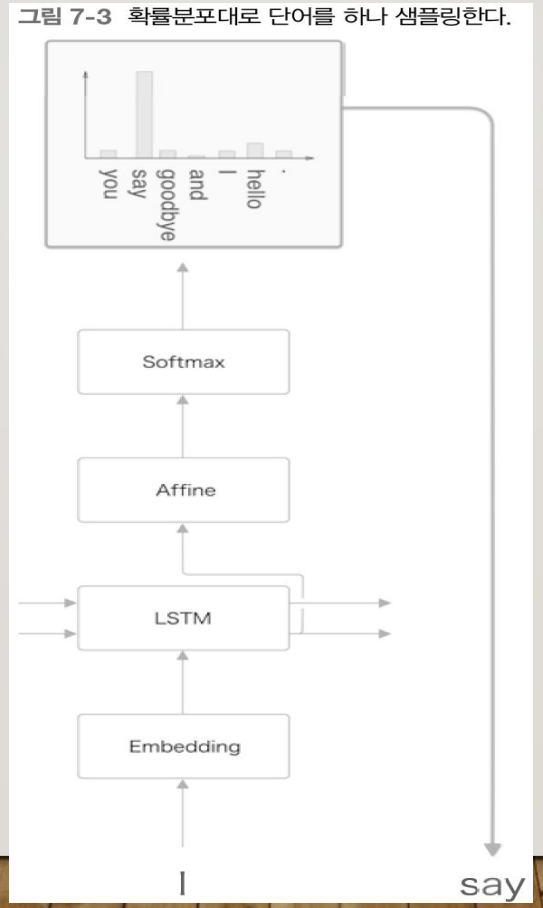
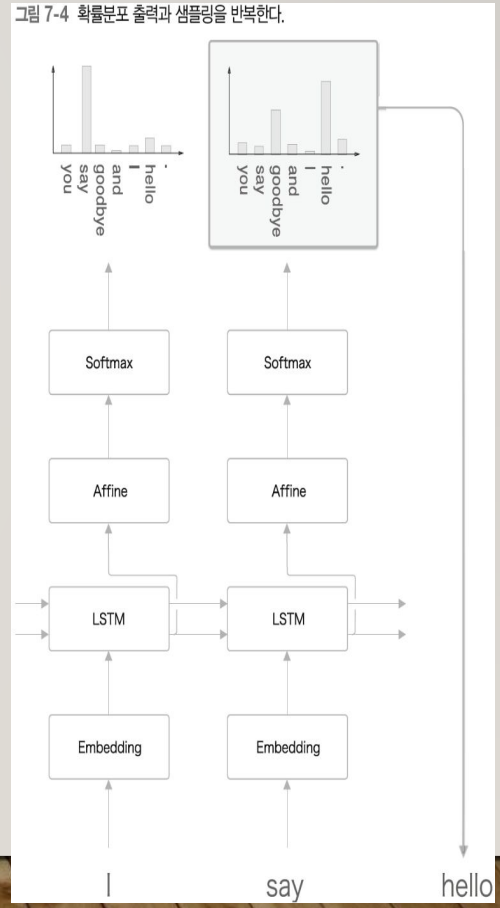


그림 7-4 확률분포 출력과 샘플링을 반복한다.



코드 구현

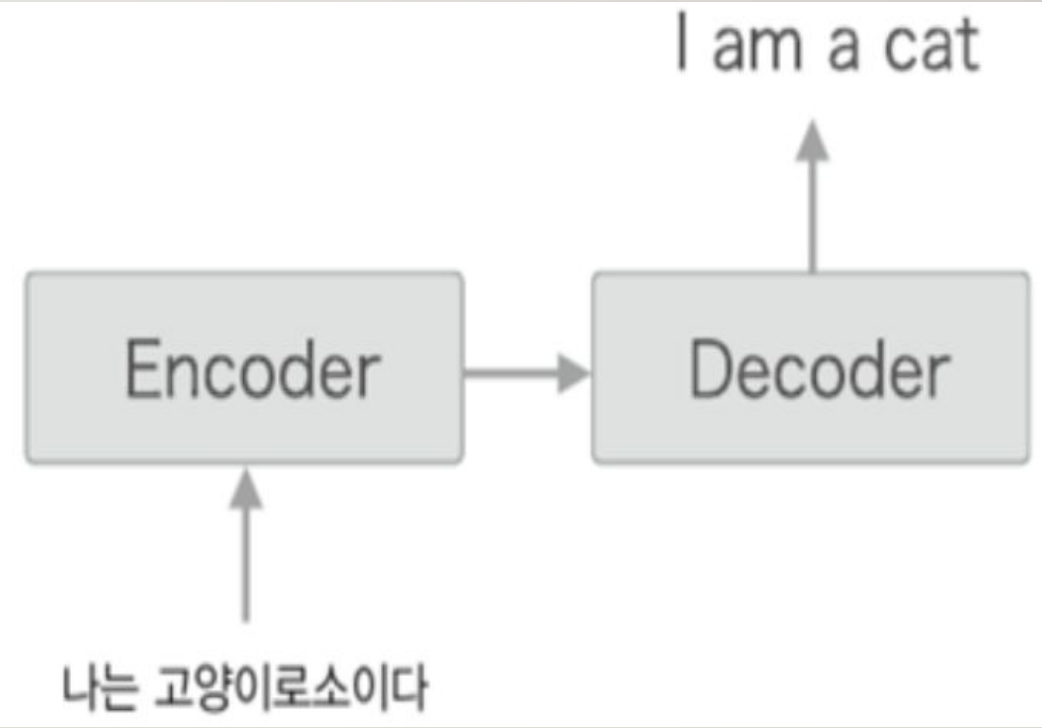
```
you seek trades.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 practiced they monitored other such computers corp. with two greater conditions based.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 practiced they monitored other such computers corp. with two greater conditions based.
 the buyer is located paying its standards.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 practiced they monitored other such computers corp. with two greater conditions based.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 but the rtc is slowing of recent stable goods.
 practiced they monitored other such computers corp. with two greater conditions based.
 the buyer is located paying its standards.
 when messages david garrison a european disaster rooms a vice computer who presidential makes him as he had annual sales of taking demonstrations
.
 if they would provide such a acceleration in an effort to have a lot of people who be working into the crandall way of the industry is closer to
the reward marsh & lewis enough to come on graduates.
 currently
```

더 좋은 문장으로

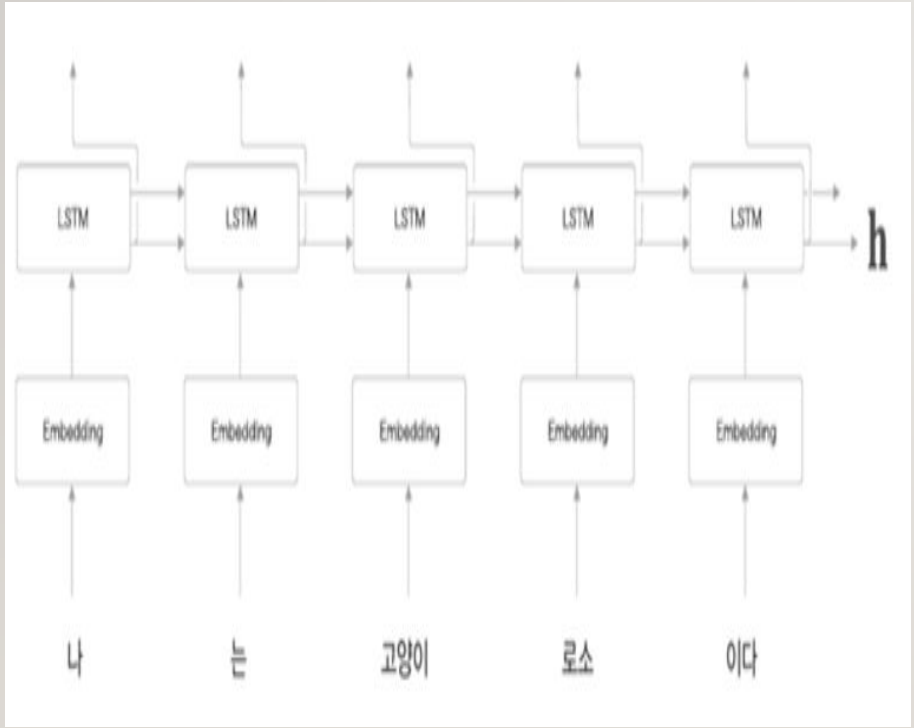
```
you 've seen two families and the women and two other women of students.

the principles of investors that prompted a bipartisan rule of which had a withdrawn target of black

the appeal was to deny steady increases in the operation of dna and educational damage in the 1950s.
```

SEQ2SEQ의 원리

- Encoder
  - 입력 데이터를 인코딩(부호화)하고, 정보를 어떤 규칙에 따라 변환하는 것
- Decoder
  - 인코딩된 데이터를 디코딩(복호화)하는 것

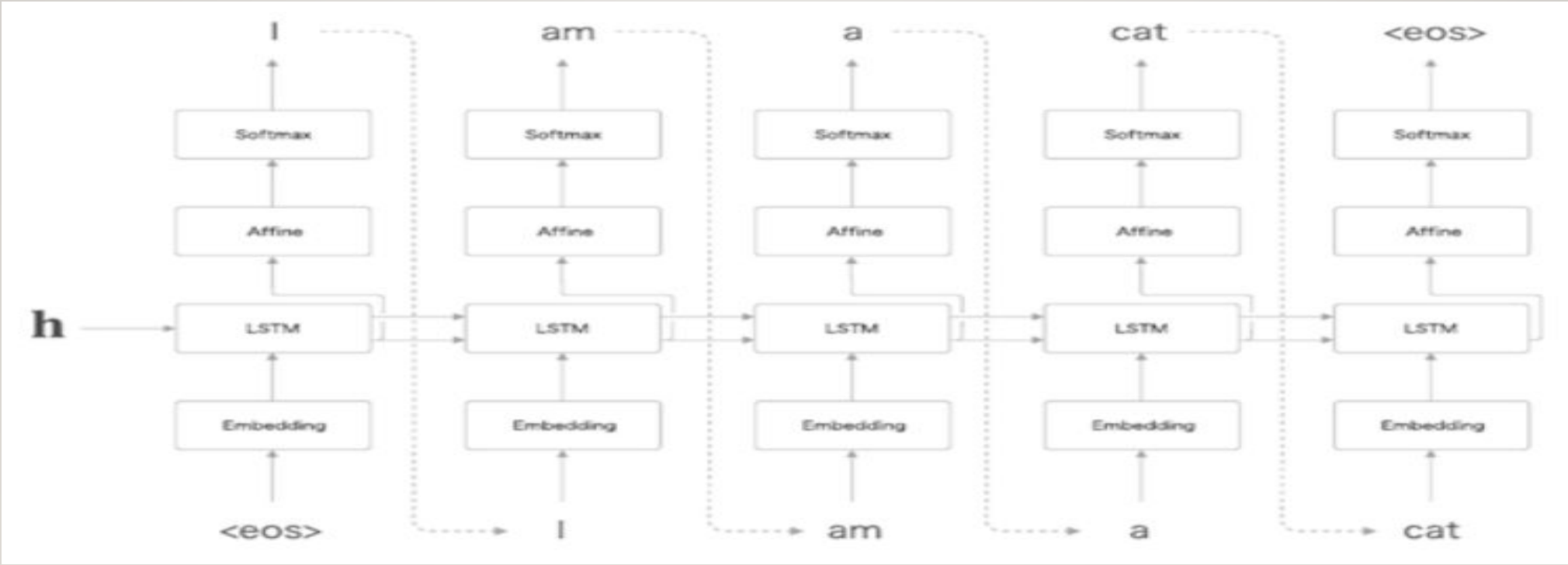


ENCODER 계층

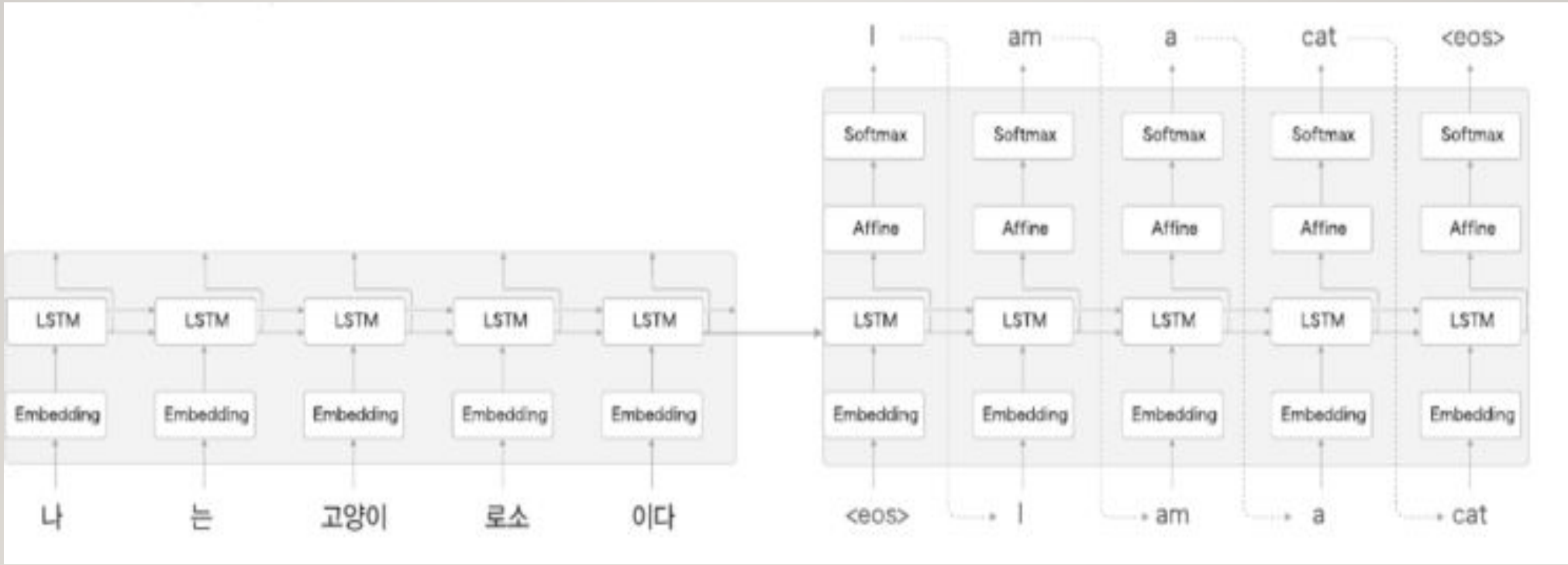

LSTM
LSTM
LSTM
LSTM
LSTM
h
Embedding
Embedding
Embedding
Embedding
Embedding
나
는
고양이
로소
이다


h
나는 고양이로소이다
Encoder
이름은 아직 없어
Encoder
어디서 태어났는지 도무지 모르겠네
Encoder

DECODER 계층

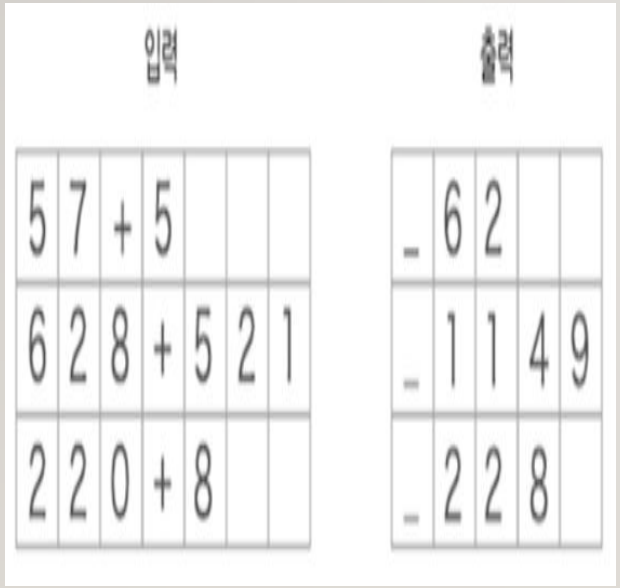


DECODER와 ENCODER 연결한 계층



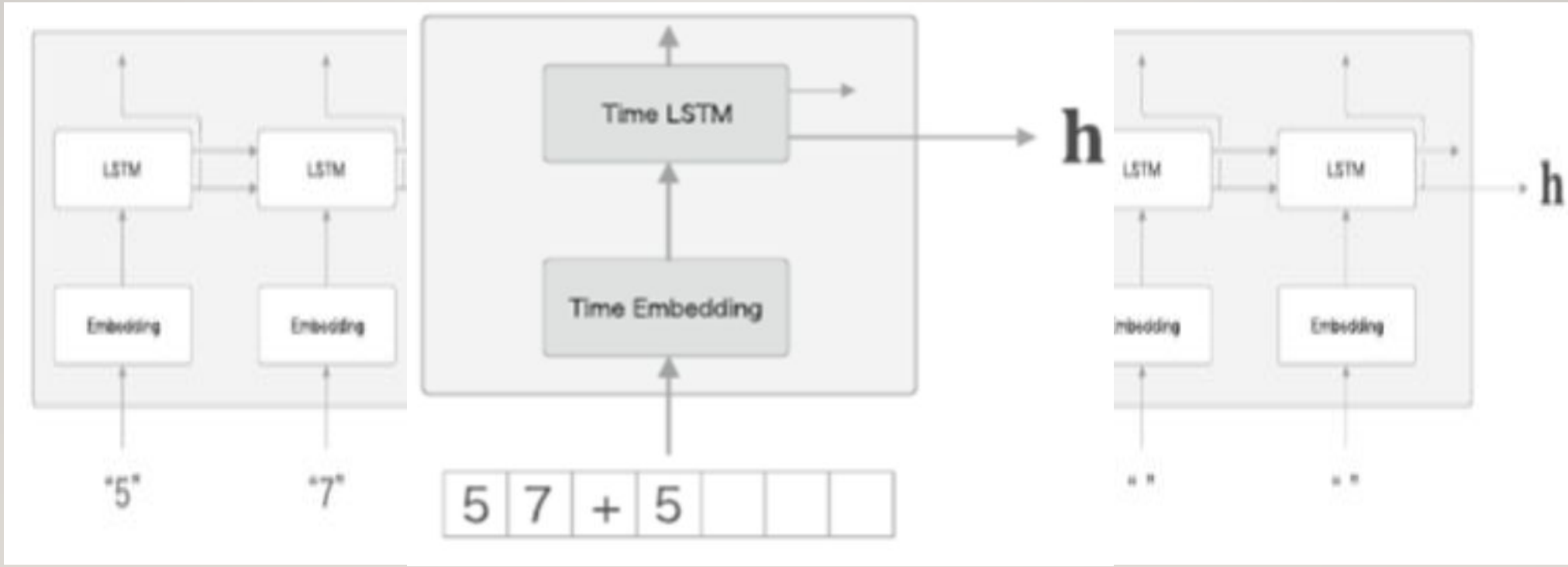
시계열 데이터 변환용 장난감 문제

가변 길이 시계열 데이터

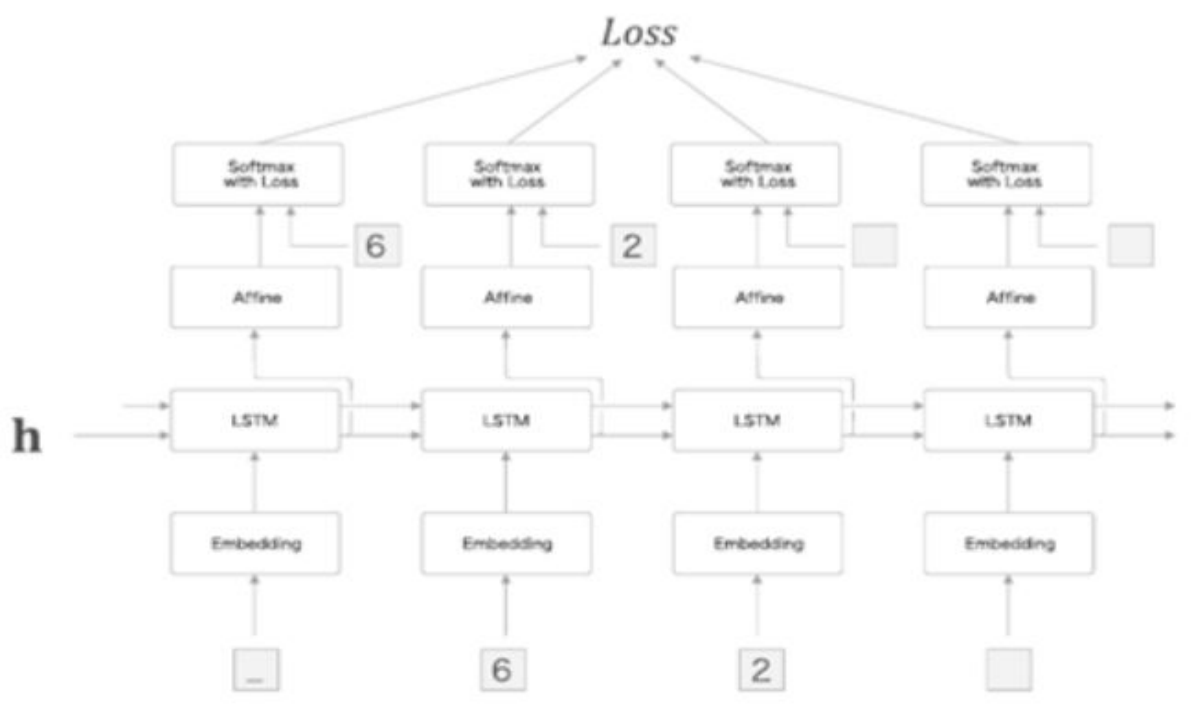


- **문제 조건**
  - **0 ~ 999** 사이의 숫자 **2**개만을 더함
  - 정답 데이터에도 패딩 수행
  - 출력 앞에 구분자로 밑줄 붙임
- Problem
  - 존재하지 않던 패딩용 문자까지 **seq2seq**가 처리하게 됨

SEQ2SEQ 구현 I

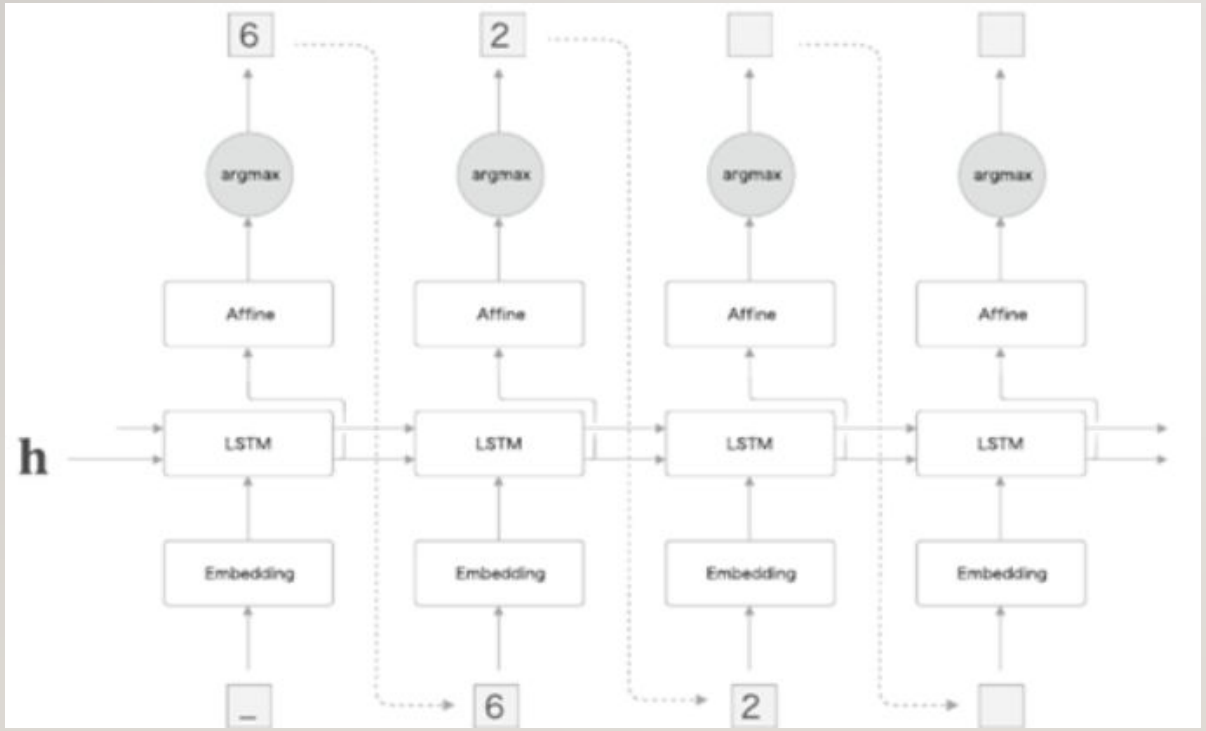

LSTM
LSTM
Embedding
Embedding
"5"
"7"
Time LSTM
Time Embedding
5 7 + 5
h
LSTM
LSTM
Embedding
h

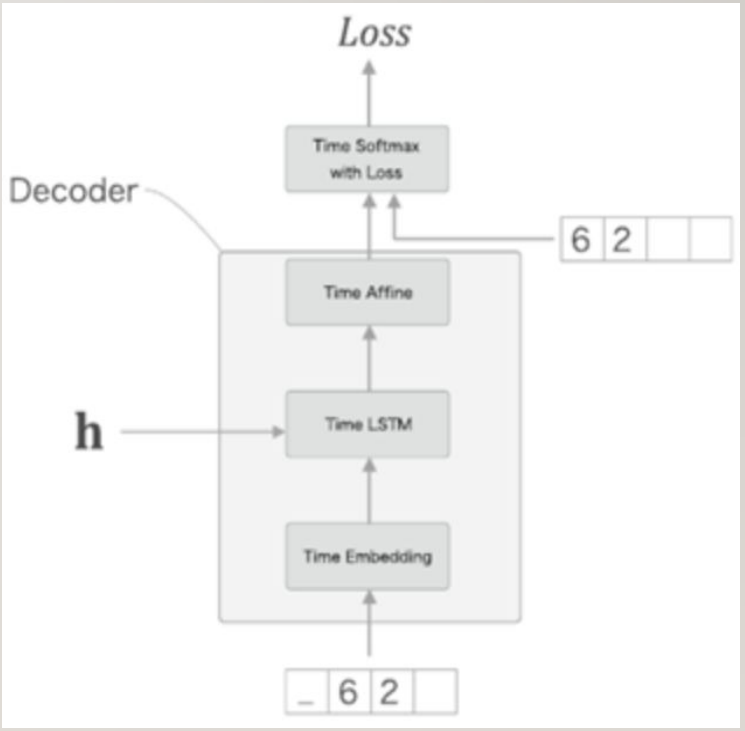
SEQ2SEQ 구현2



- 학습
  - 시계열 방향 데이터 한번에 추출 가능
- 생성
  - 최초 시작을 알리는 구문 문자 하나만 줌
  - 그 출력으로부터 문자를 하나 샘플링해 그 문자를 다음 입력을 사용하는 과정 반복

DECODER의 문자열 생성 순서





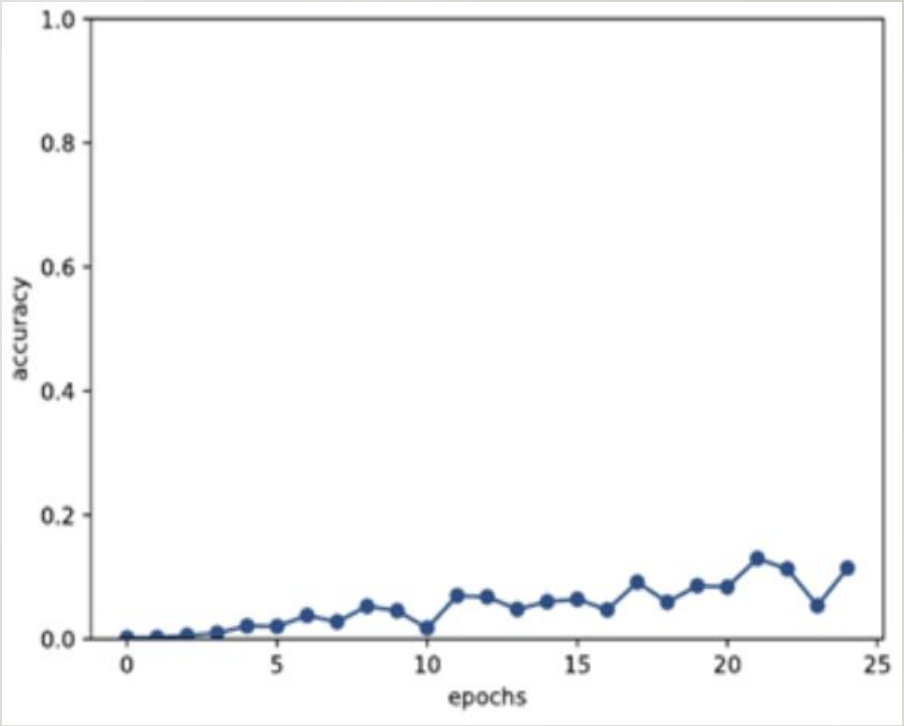
SEQ2SEQ 평가

```
Q 77+85
T 162
X 100
---
Q 975+164
T 1139
X 1000
---
Q 582+84
T 666
X 1000
---
Q 8+155
T 163
X 100
---
Q 367+55
T 422
X 1000
---
Q 600+257
T 857
X 1000
```

```
Q 77+85
T 162
X 994
---
Q 975+164
T 1139
X 1000
---
Q 582+84
T 666
X 700
---
Q 8+155
T 163
X 100
---
Q 367+55
T 422
X 400
---
Q 600+257
T 857
X 1000
```

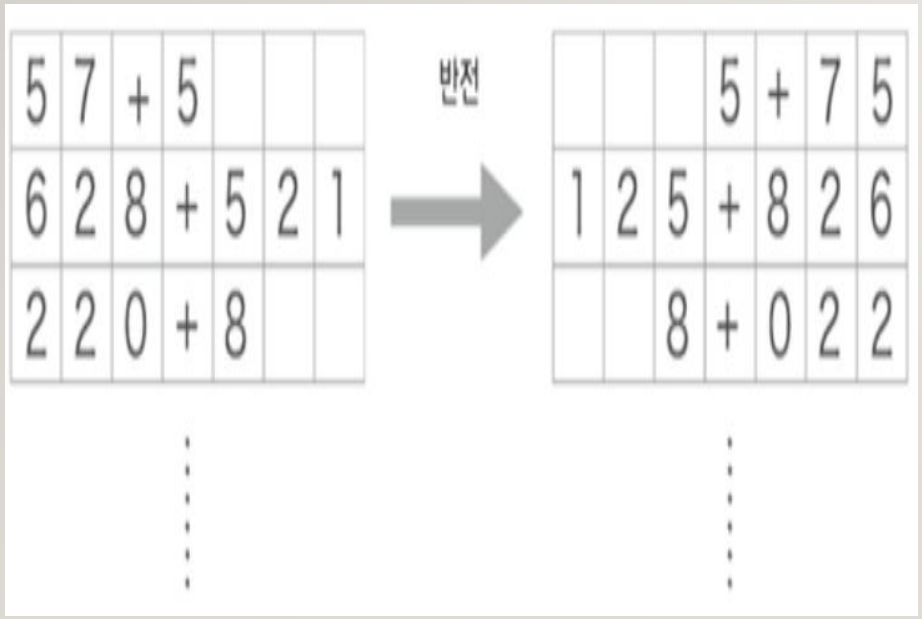
```
Q 77+85
T 162
X 166
---
Q 975+164
T 1139
X 1169
---
Q 582+84
T 666
O 666
---
Q 8+155
T 163
X 174
---
Q 367+55
T 422
X 411
---
Q 600+257
T 857
X 864
```

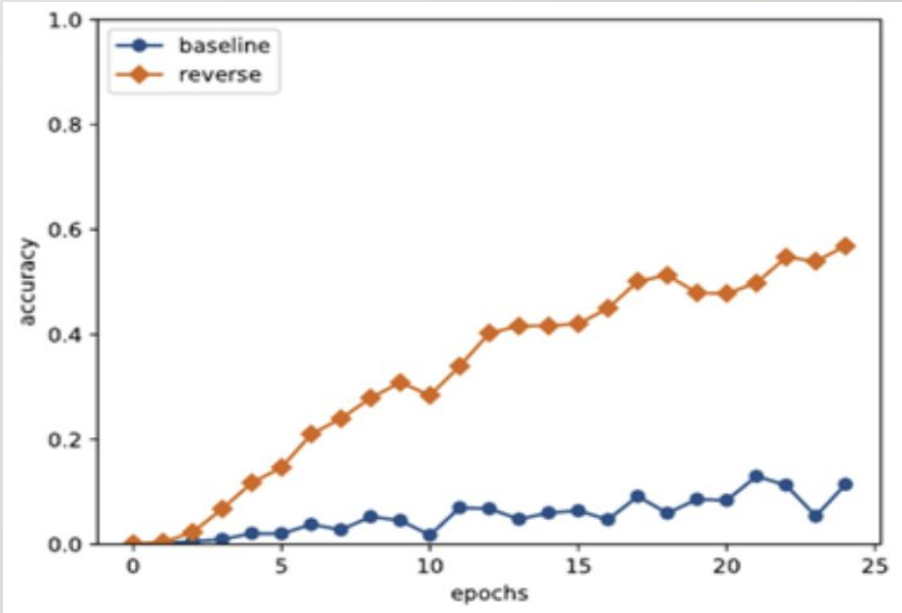
```
Q 77+85
T 162
X 164
---
Q 975+164
T 1139
X 1160
---
Q 582+84
T 666
X 668
---
Q 8+155
T 163
X 173
---
Q 367+55
T 422
X 411
---
Q 600+257
T 857
X 870
```



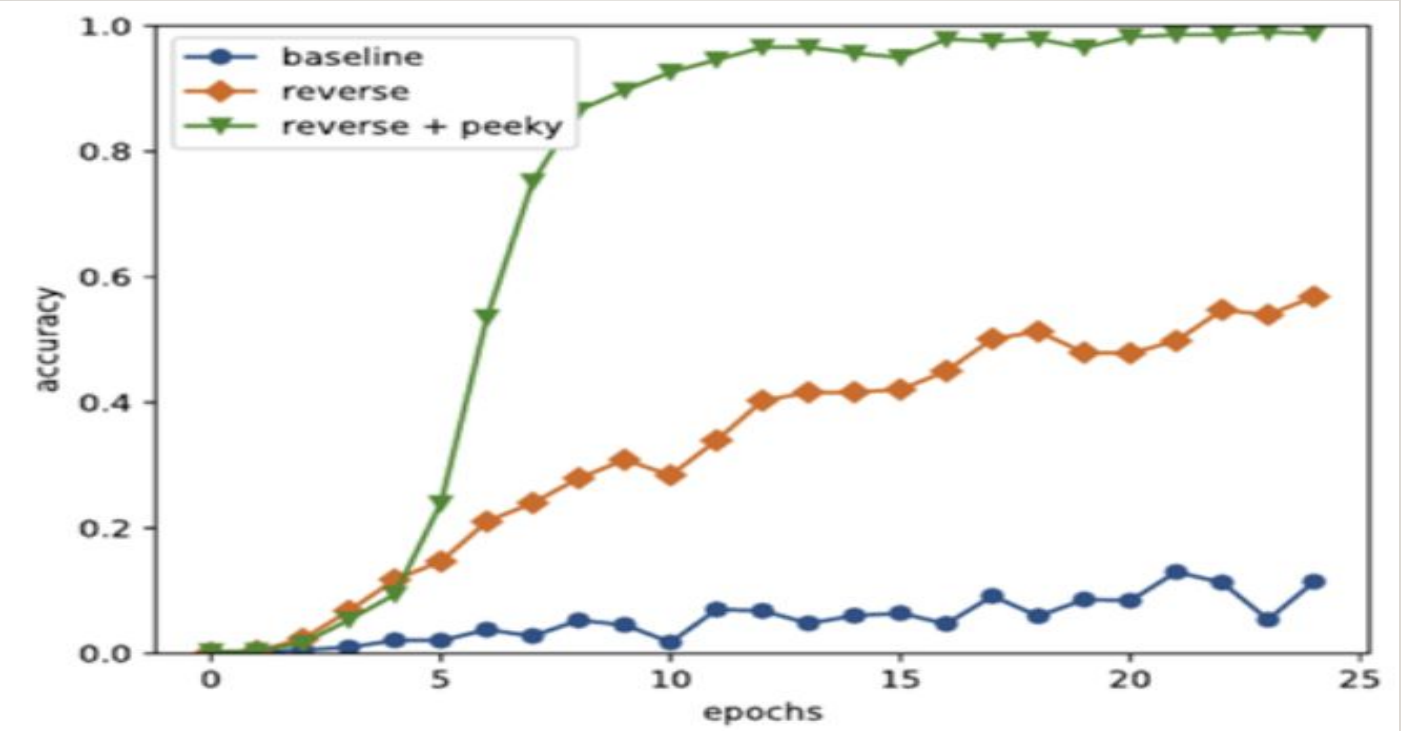
SEQ2SEQ 개선 I

- 입력 데이터 반전(Reverse)





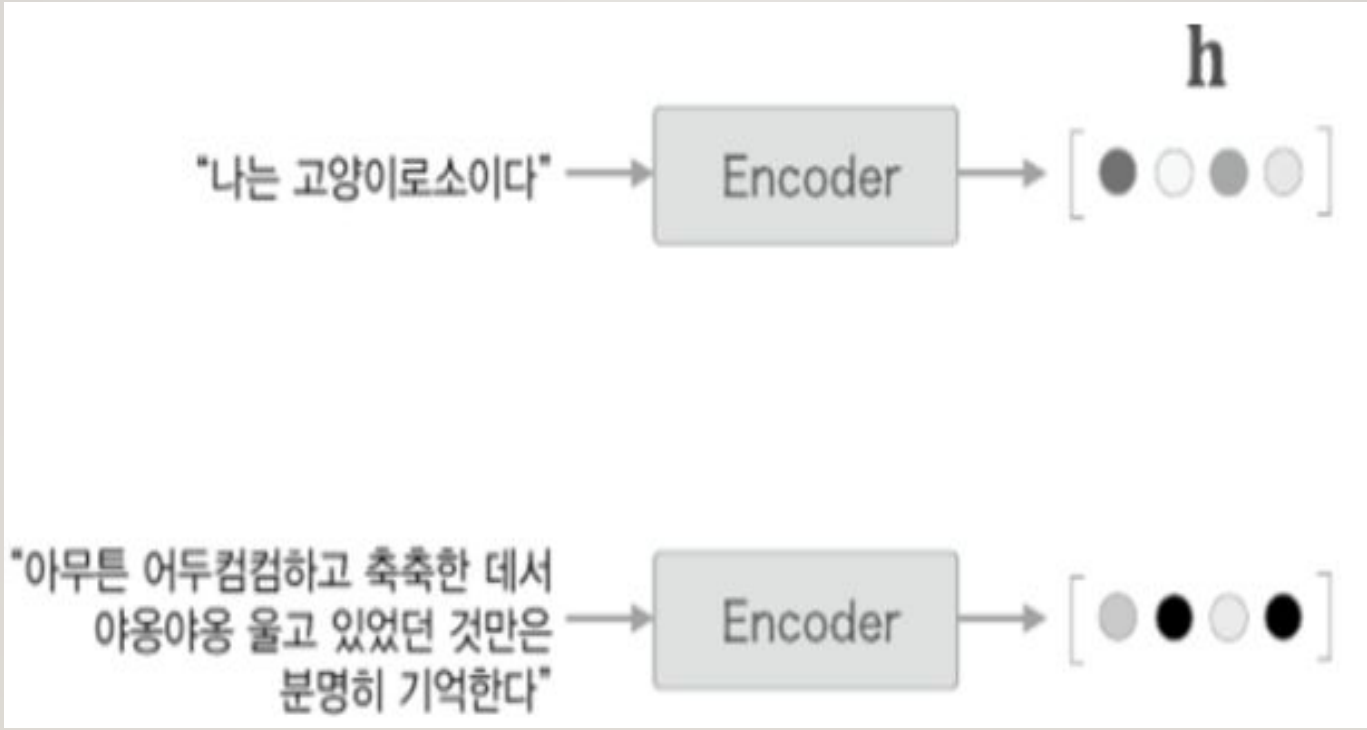
SEQ2SEQ 개선2


baseline
reverse
reverse + peeky
accuracy
epochs
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25

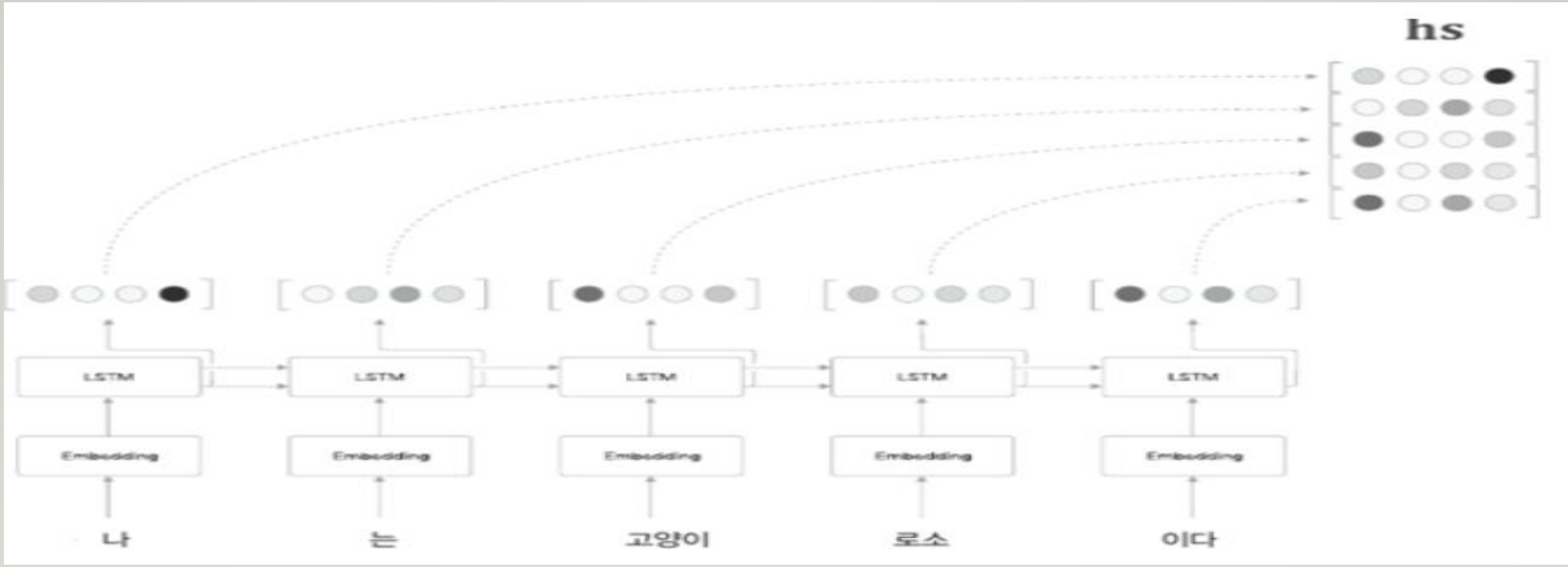
SEQ2SEQ를 이용하는 애플리케이션

- 챗봇
  - 상대의 말을 자신의 말로 변환하는 문제
- 알고리즘 학습
  - 소스 코드는 문자로 쓰여진 시계열 데이터
- 이미지 캡셔닝
  - 이미지를 문장으로 변환하는 문제

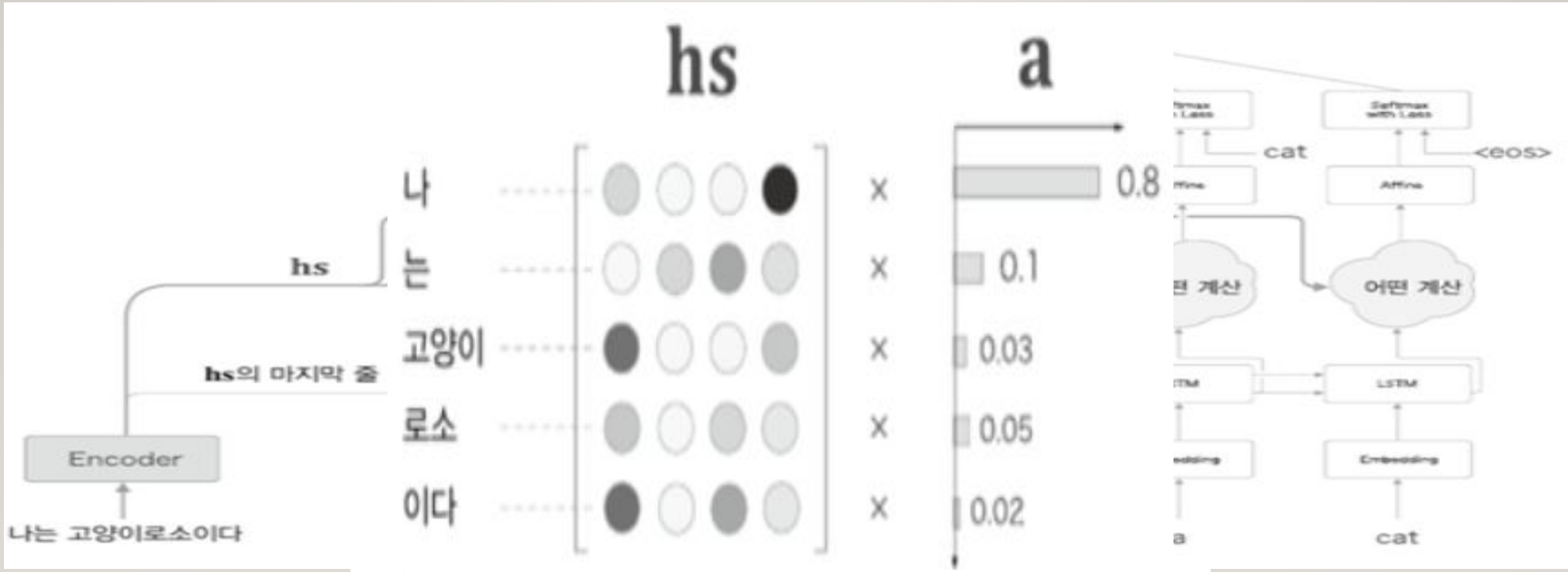
SEQ2SEQ의 문제점



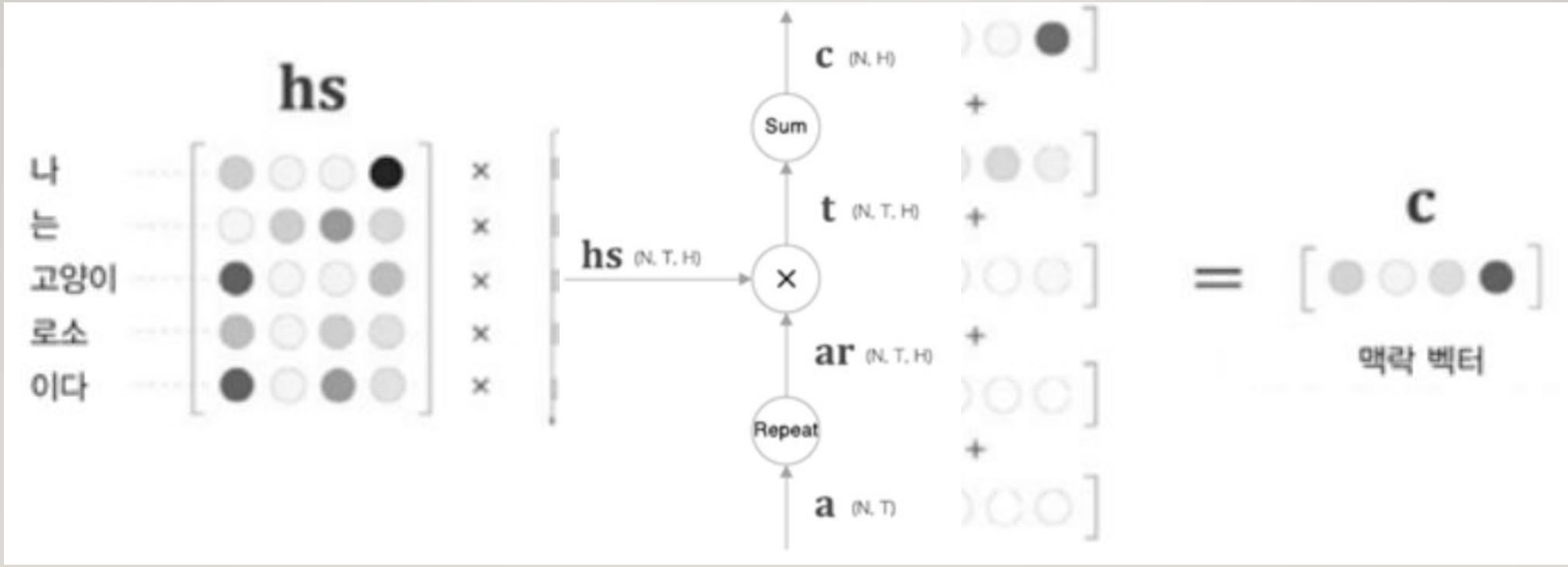
ENCODER 개선


hs
LSTM
LSTM
LSTM
LSTM
LSTM
Embedding
Embedding
Embedding
Embedding
Embedding
나
는
고양이
로소
이다

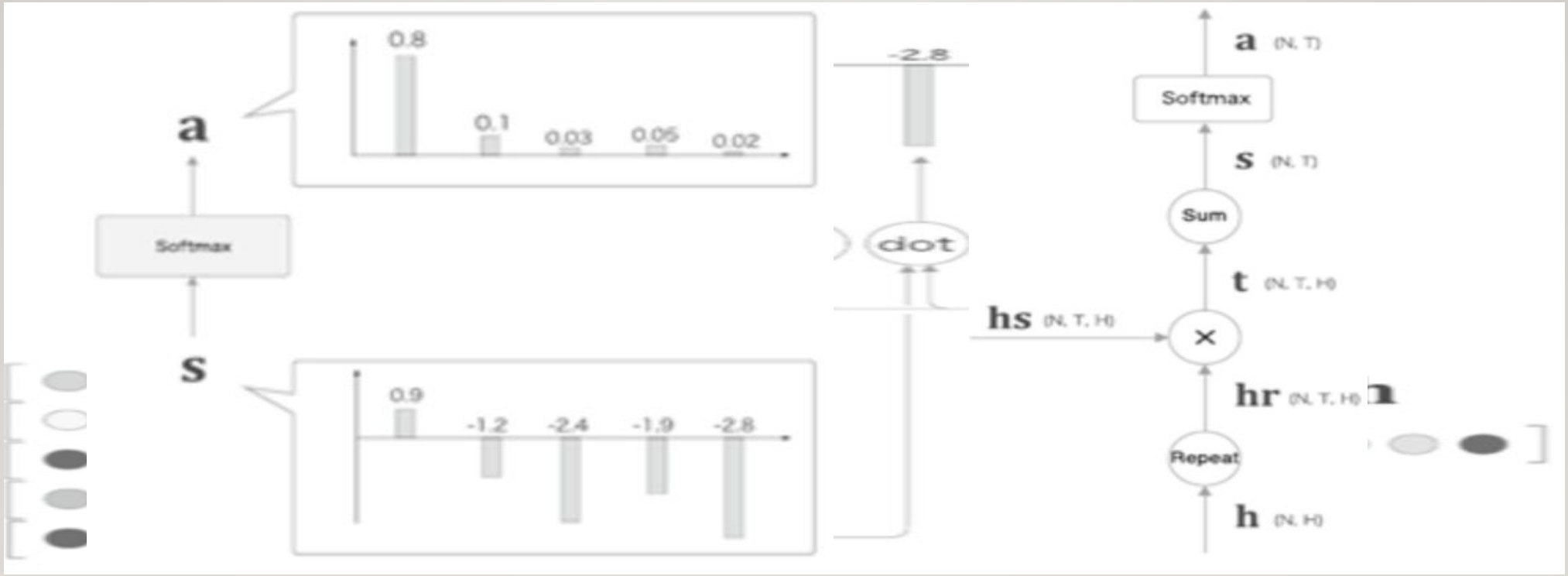
DECODER 개선 I


hs
a
나
는
고양이
로소
이다
0.8
0.1
0.03
0.05
0.02
hs
hs의 마지막 줄
Encoder
나는 고양이로소이다
cat
<eos>
Affine
어떤 계산
LSTM
Embedding
cat

맥락 벡터 구하기



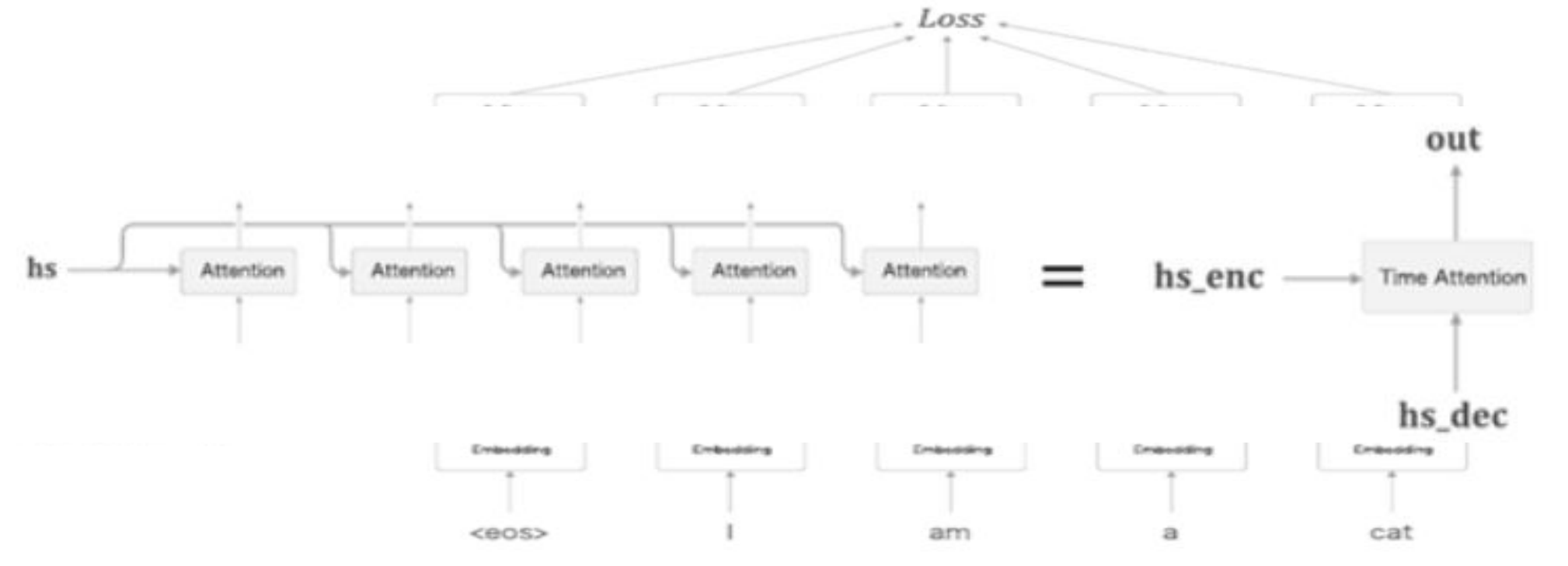
DECODER 개선2


a
0.8
0.1
0.03
0.05
0.02
Softmax
s
0.9
-1.2
-2.4
-1.9
-2.8
-2.8
dot
hs (N, T, H)
a (N, T)
Softmax
s (N, T)
Sum
t (N, T, H)
×
hr (N, T, H)
Repeat
h (N, H)

DECODER 개선3

- Attention Weight 계층
  - Encoder가 출력하는 각 단어의 벡터 hs에 주목하여 해당 단어의 가중치 a를 구함
- Weight Sum 계층
  - 계층이 a와 hs의 가중합을 구하고, 그 결과를 맥락 벡터 c로 출력
- Attention 계층
  - 두 계층을 하나로 결합

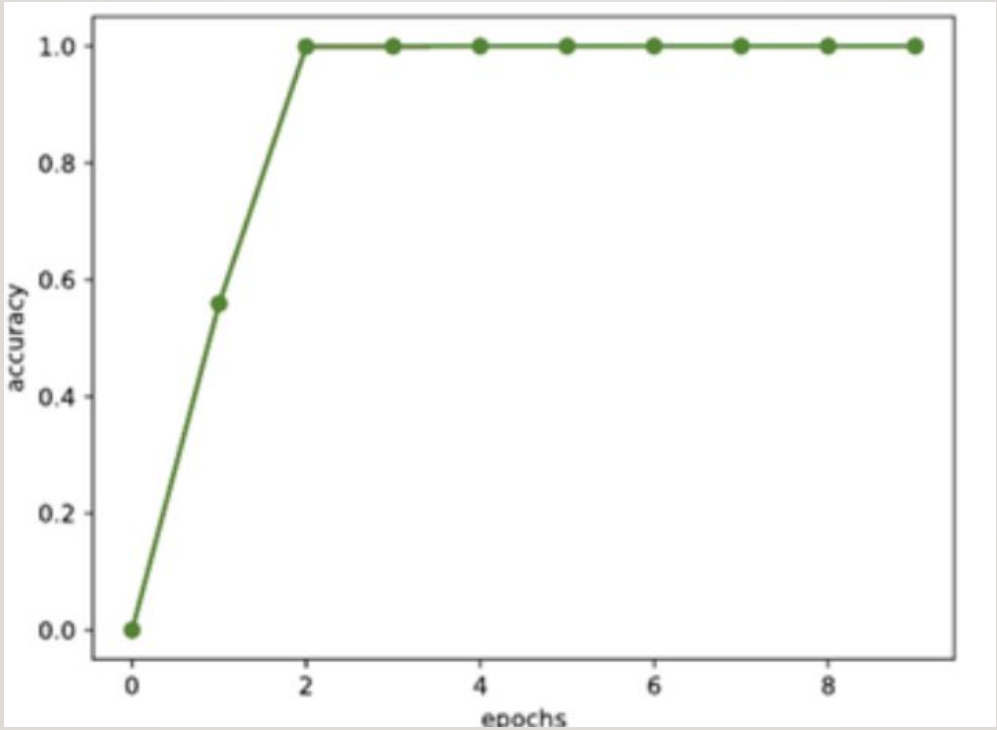
ATTENTION 계층



ATTENTION 평가

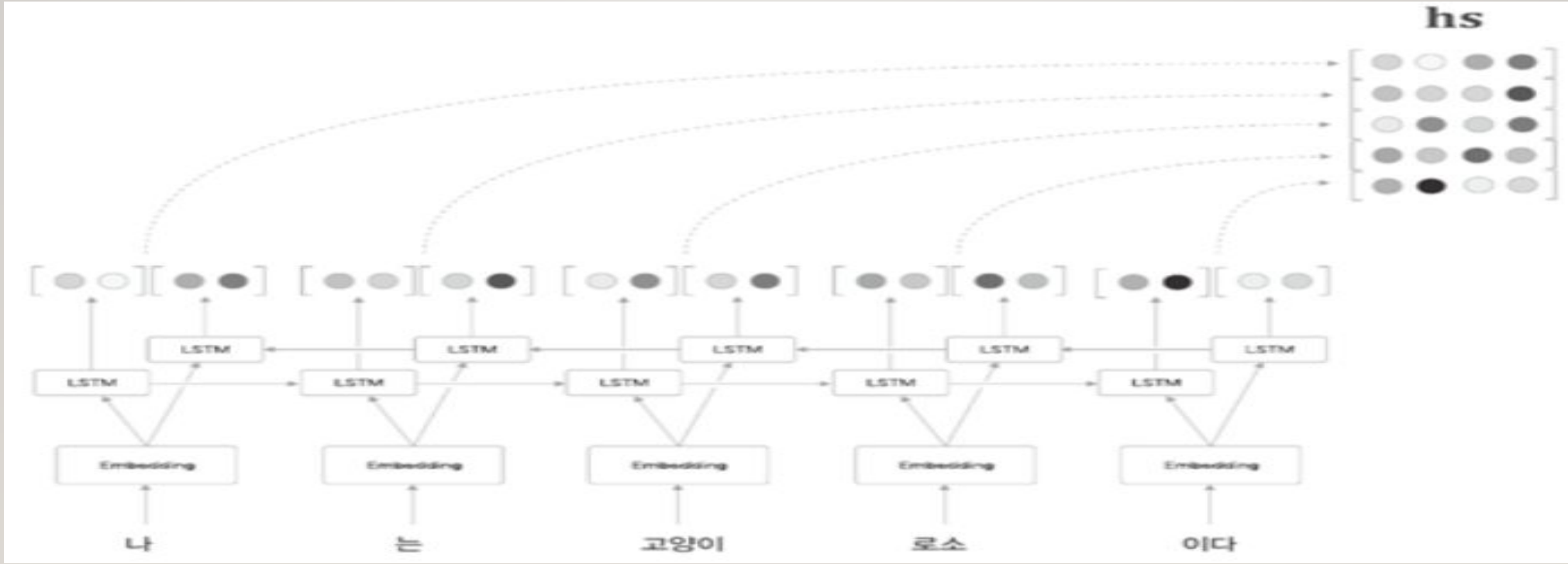
- 겉보기만큼 간단하지 않다
- 입력과 출력 사이에 알기 쉬운 대응 관계가 있다

어텐션을 갖춘 SEQ2SEQ의 학습



어텐션 시각화

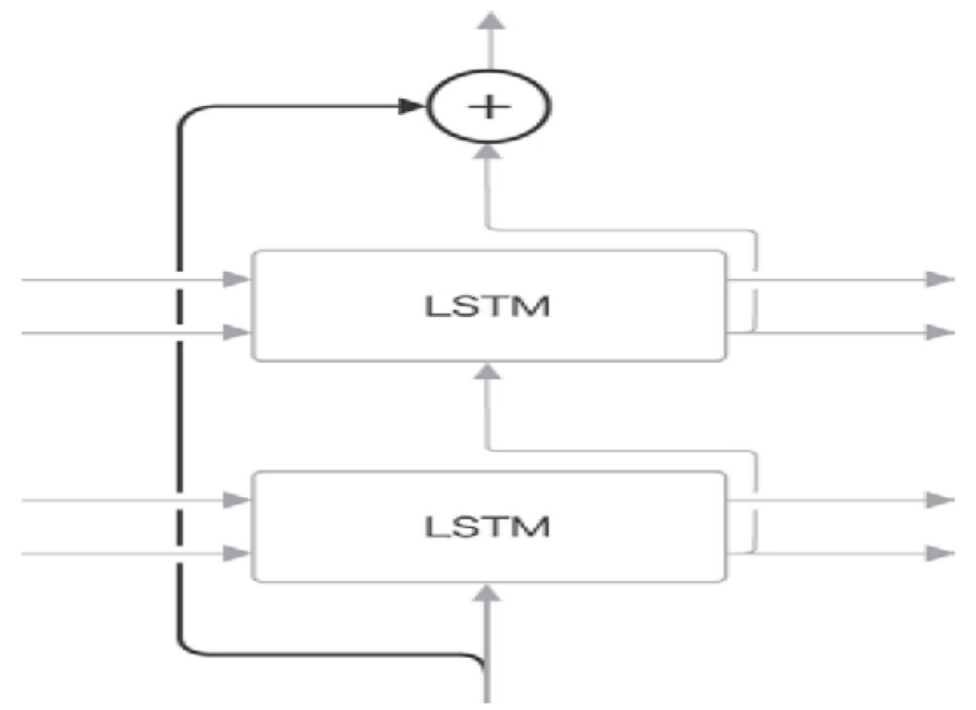
양방향 RNN


hs
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
Embedding
Embedding
Embedding
Embedding
Embedding
나
는
고양이
로소
이다

어텐션 계층 사용 방법

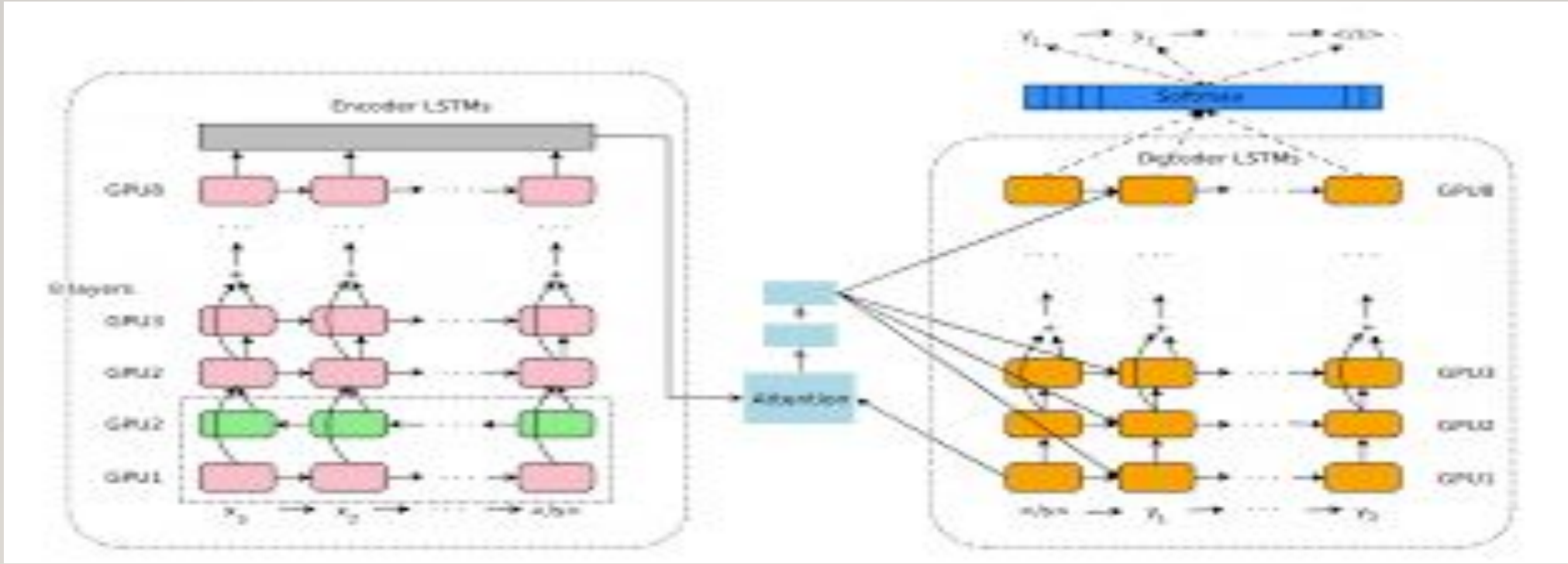

Loss
Softmax with Loss
Softmax with Loss
I
am
Affine
Affine
hs
hs
Attention
Attention
Encoder
hs의 마지막 줄
LSTM
LSTM
나는 고양이로소이다
Embedding
Embedding
<eos>
I

SEQ2SEQ 심층화와 SKIP 연결

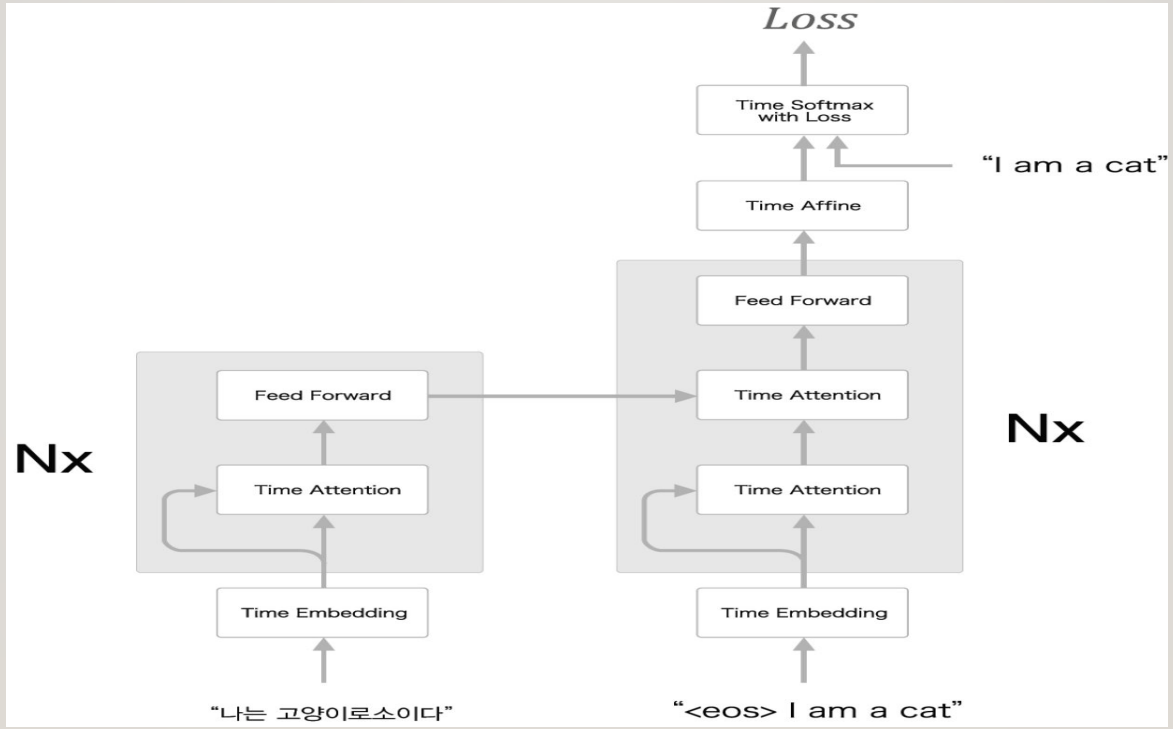


- Skip 연결
  - 층을 깊게 할 때 사용되는 기법
  - 계층을 건너뜀
  - 접속부에서 원소 별 덧셈
    - 역전파 시 기울기를 그래도 흘려보내므로 기울기 손실/폭발 위험 줄여줌
  - RNN 시간 방향에서 기울기 소실
  - 기울기 폭발: 기울기 클리핑
  - 기울기 소실: skip 연결

구글 신경망 기계번역(GNMT)



트랜스포머


Loss
Time Softmax with Loss
"I am a cat"
Time Affine
Feed Forward
Time Attention
Time Attention
Time Embedding
Nx
Feed Forward
Time Attention
Time Embedding
Nx
"나는 고양이로소이다"
"<eos> I am a cat"

뉴럴 튜링 머신(NTM)

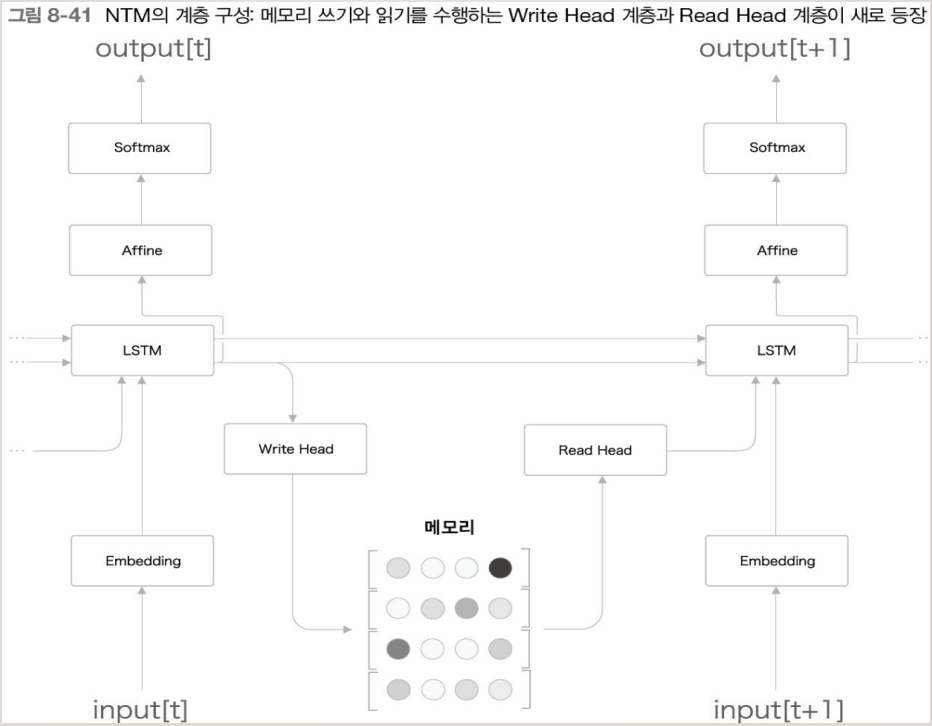


그림 8-41 NTM의 계층 구성: 메모리 쓰기와 읽기를 수행하는 Write Head 계층과 Read Head 계층이 새로 등장
output[t]
output[t+1]
Softmax
Softmax
Affine
Affine
LSTM
LSTM
Write Head
Read Head
메모리
Embedding
Embedding
input[t]
input[t+1]