

NIPS 2014

Sequence to Sequence Learning with Neural Networks

2022.09.23

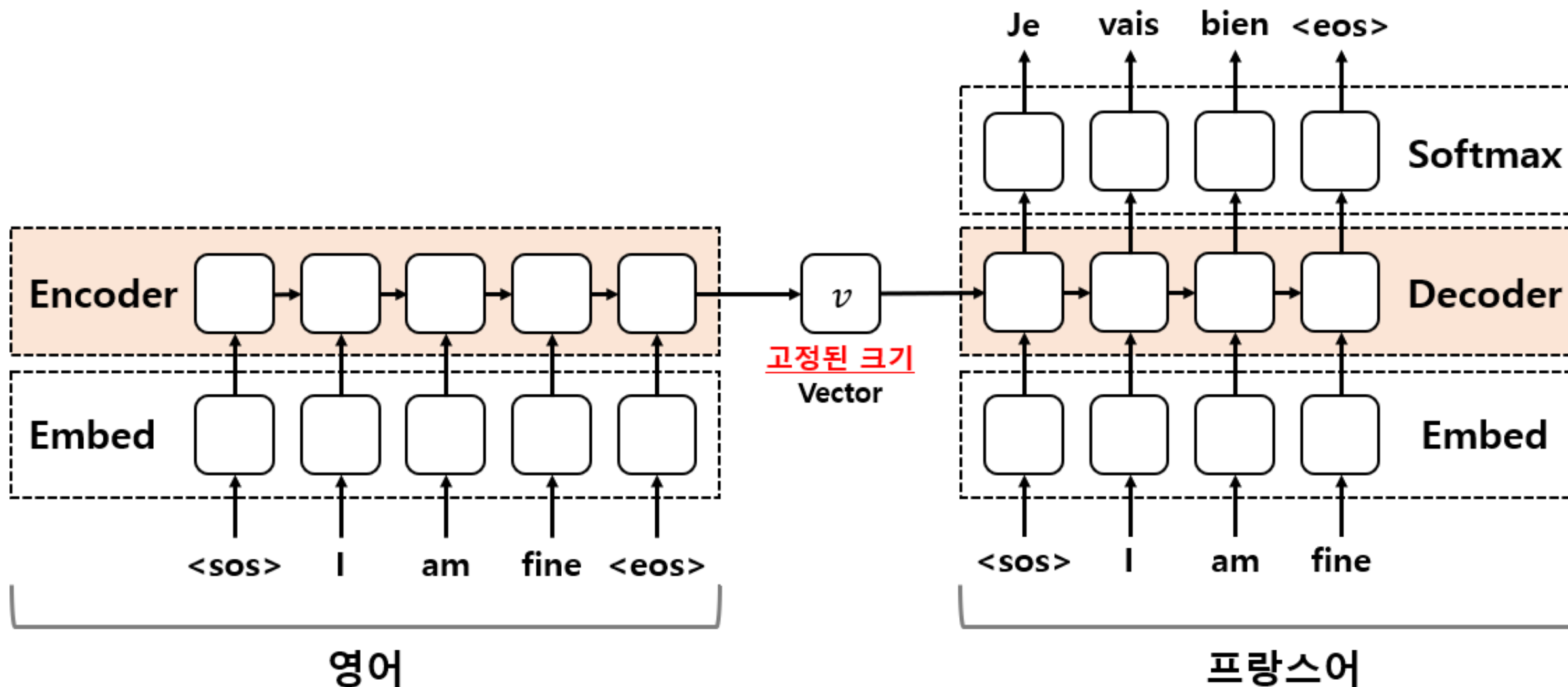
논문 리뷰

배성훈

Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- Research Background:

- 기존 DNN의 한계: Sequence 간의 매핑에 사용하기 어려움 (미리 문장의 길이를 알 수 없는 문제에 적용X)
- 이러한 한계를 극복하기 위해 문장 구조에 최소한의 가정만 하는 LSTM을 활용한 End-to-End 접근법을 제시



Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- Background Information:

- Language Model
- 문장에 확률을 부여하는 모델
- 특정 상황에 알맞은 문장, 단어 예측 가능

Machine Translation

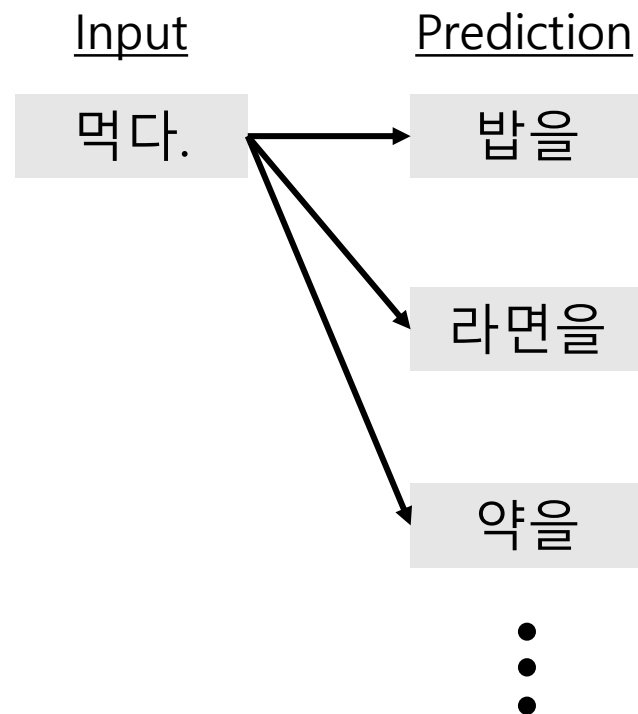
$P(\text{나는 피자를 먹는다.} \mid I \text{ eat pizza})$



$P(\text{나는 피자를 먹지 않는다.} \mid I \text{ eat pizza})$



Word Prediction



Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- Background Information:

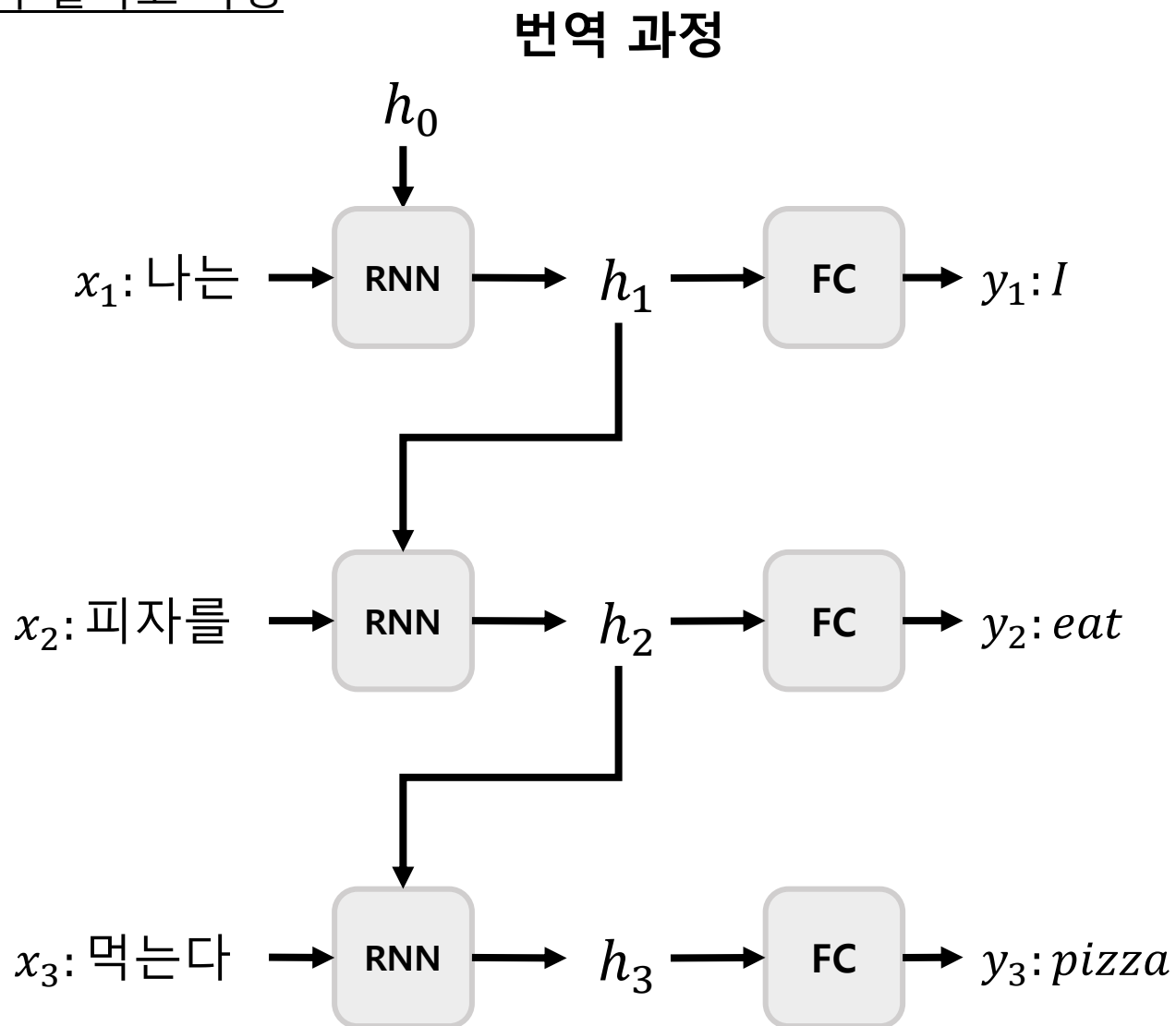
- RNN based Machine Translation
- RNN 기반 기계 번역은 입력과 출력의 크기가 같다고 가정

입력: (x_1, x_2, \dots, x_T)

출력: (y_1, y_2, \dots, y_T)

➤ $h_t = \text{sigmoid}(W^{hx}x_t + W^{hh}h_{t-1})$

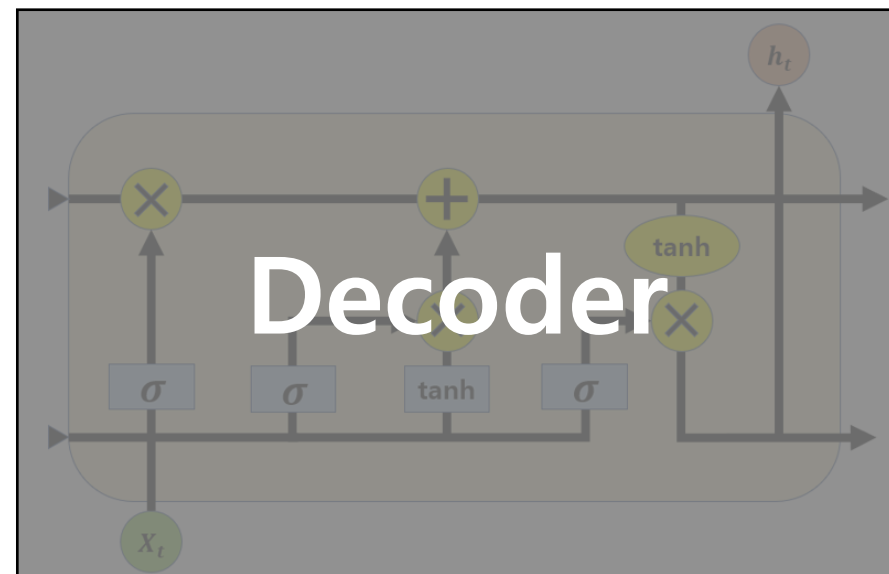
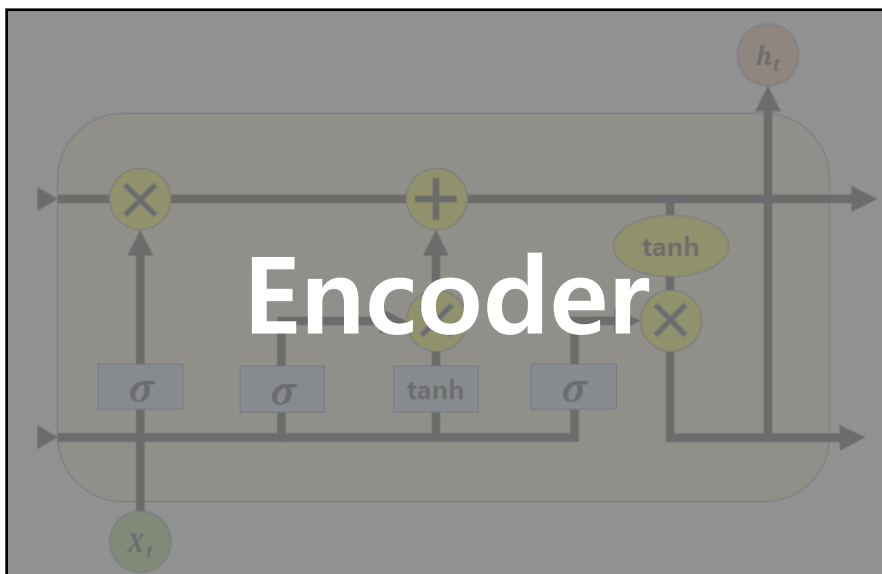
➤ $y_t = W^{yh}h_t$



Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- **Method:**

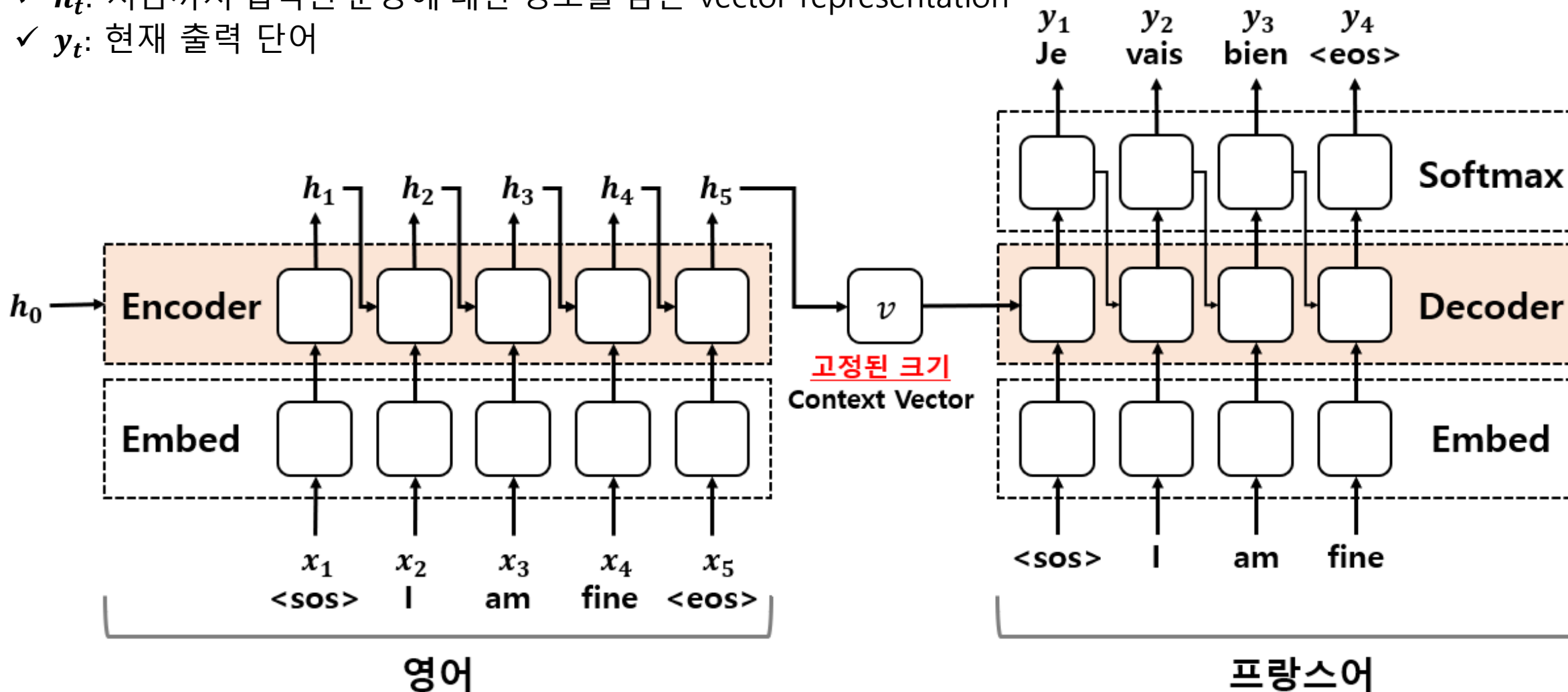
- RNN based sequence to sequence model 설계 (**End-to-End, LSTM 활용**)
- Encoder: 고정된 크기의 context vector 추출 → One LSTM
- Decoder: context vector로부터 번역 결과 추론 → Another LSTM
 - Encoder, Decoder는 서로 다른 파라미터를 가짐



Sequence to Sequence Learning with Neural Networks (NIPS 2014)

• Method:

- 제안한 모델은 4-layered LSTM 사용 → Deeper LSTM이 더 우수한 성능 보임
- ✓ x_t : 현재 입력 단어
- ✓ h_t : 지금까지 입력된 문장에 대한 정보를 담은 vector representation
- ✓ y_t : 현재 출력 단어



Sequence to Sequence Learning with Neural Networks (NIPS 2014)

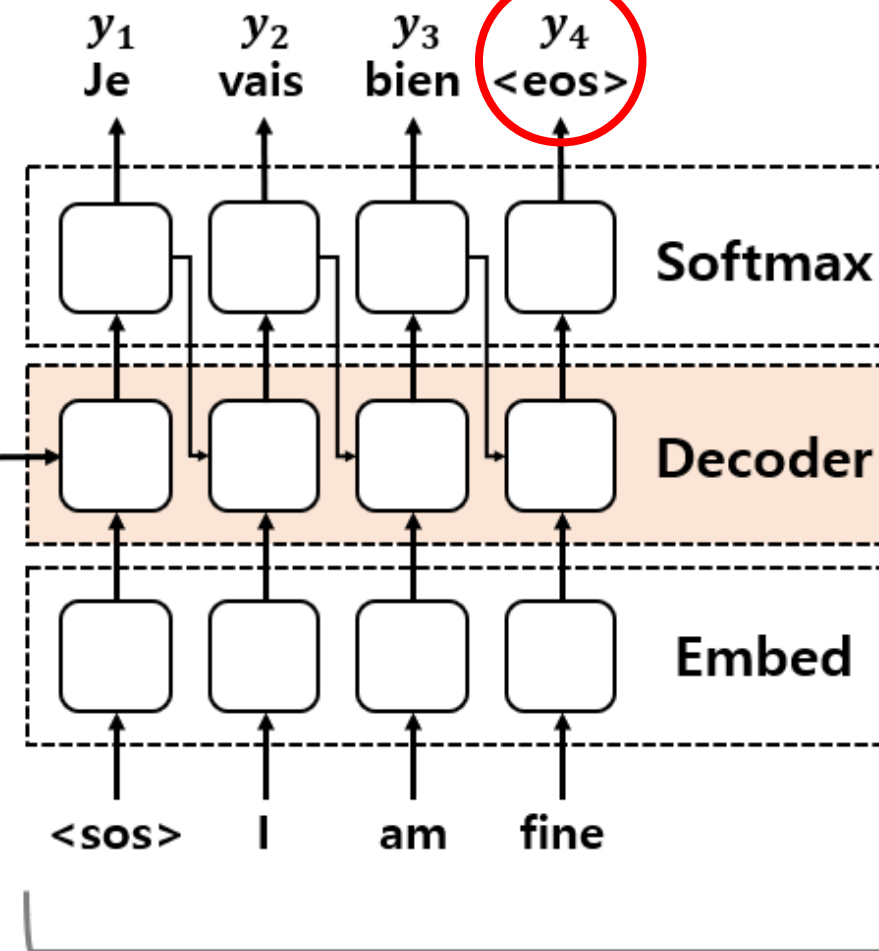
- Method:

추론 종료 시점

$$p(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod_{t=1}^{T'} p(y_t | v, y_1, \dots, y_{t-1})$$

- v : $\langle x_1, \dots, x_T \rangle$ 정보를 담은 vector representation
- y_t : 현재 출력 단어

고정된 크기
Context Vector

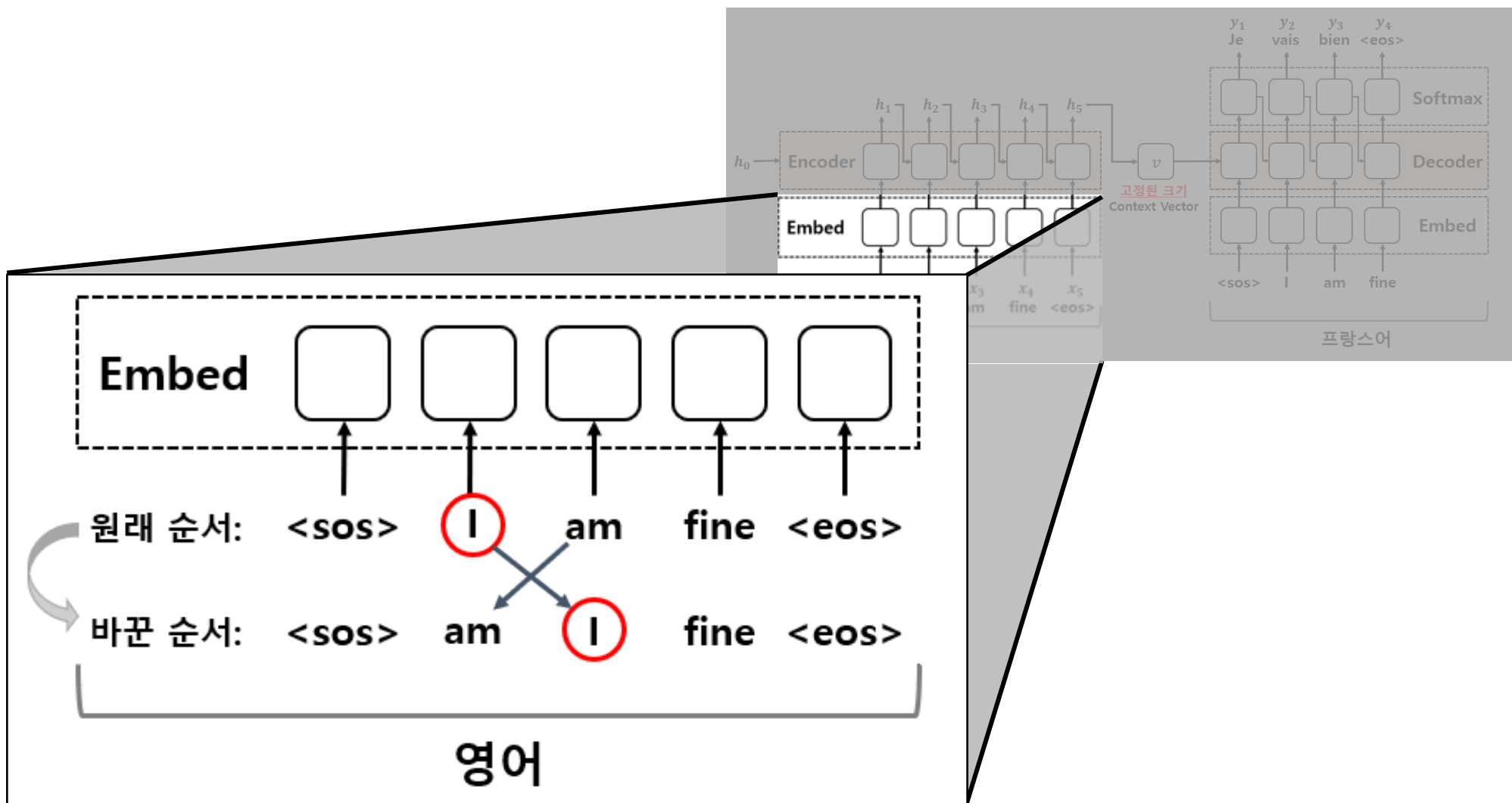


프랑스어

Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- Method:

- 입력 문장의 순서를 거꾸로 했을 때 더 높은 정확도를 보임 (출력 문장의 순서는 바꾸지 않음)
- 데이터의 많은 short term dependencies가 최적화 문제를 더 쉽게 만듦



Sequence to Sequence Learning with Neural Networks (NIPS 2014)

• Experiment:

- **Dataset:** WMT' 14 English to French dataset
- **목적 함수 (objective function)**를 사용해 log probability를 높이도록 학습한다.
- 학습이 끝나면, 번역을 생성하는 Inference 과정에서 확률이 가장 높은 output sequence를 찾는데 left-to-right beam search decoder 사용

Objective Function

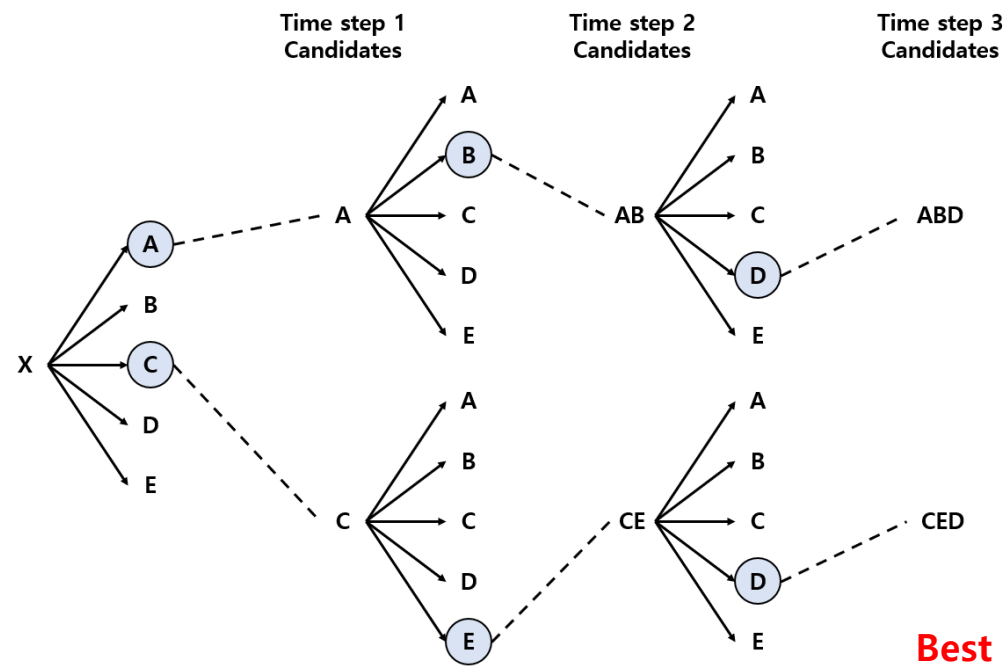
$$\frac{1}{|S|} \sum_{(T,S) \in S} \log p(T|S)$$

T : correct translation

S : source sentence

Beam Search

해당 시점에서 가장 높은 예측 확률을 선택



Best Performance

$$\hat{T} = \underset{T}{\operatorname{argmax}} p(T|S)$$

Beam 크기: 2
<EOS> 토큰이 나오는 문장: 결과 문장

Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- **Experiment:**

- Source sentences의 순서를 바꿨을 때, LSTM의 test perplexity는 감소 (5.8 \rightarrow 4.7)
BLEU scores는 증가 (25.9 \rightarrow 30.6)
- 제안한 방법은 Ensemble of 5 Reversed LSTMs + Beam size=12일 때, baseline 모델보다 좋은 성능을 보임

The performance of the LSTM on WMT'14 English to French test set

Method	test BLEU score (ntst14)
Bahdanau et al. [2]	28.45
Baseline System [29]	33.30
Single forward LSTM, beam size 12	26.17
Single reversed LSTM, beam size 12	30.59
Ensemble of 5 reversed LSTMs, beam size 1	33.00
Ensemble of 2 reversed LSTMs, beam size 12	33.27
Ensemble of 5 reversed LSTMs, beam size 2	34.50
Ensemble of 5 reversed LSTMs, beam size 12	34.81

Method	test BLEU score (ntst14)
Baseline System [29]	33.30
Cho et al. [5]	34.54
State of the art [9]	37.0
Rescoring the baseline 1000-best with a single forward LSTM	35.61
Rescoring the baseline 1000-best with a single reversed LSTM	35.85
Rescoring the baseline 1000-best with an ensemble of 5 reversed LSTMs	36.5
Oracle Rescoring of the Baseline 1000-best lists	~ 45

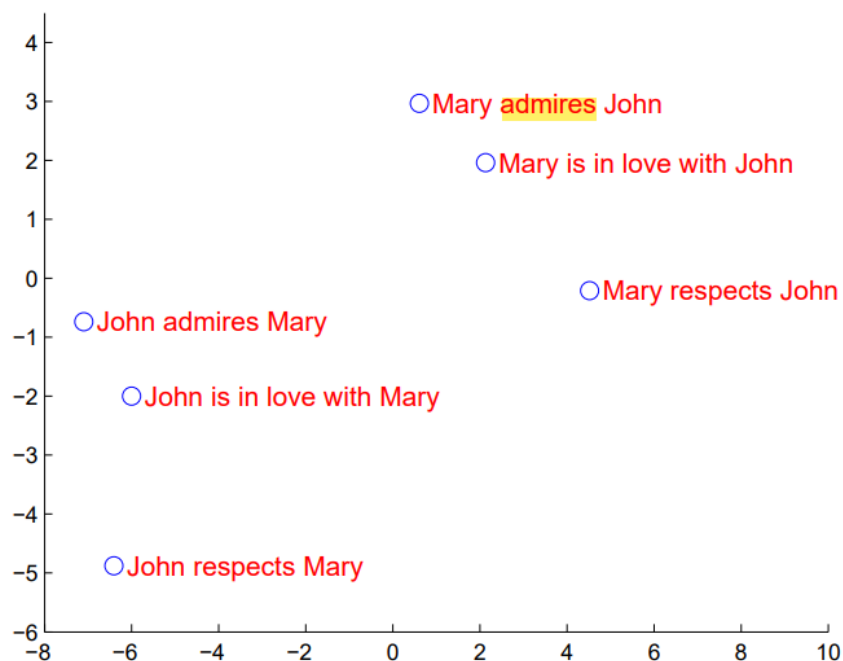
SOTA보다
0.5 낮지만
유사한 성능

Sequence to Sequence Learning with Neural Networks (NIPS 2014)

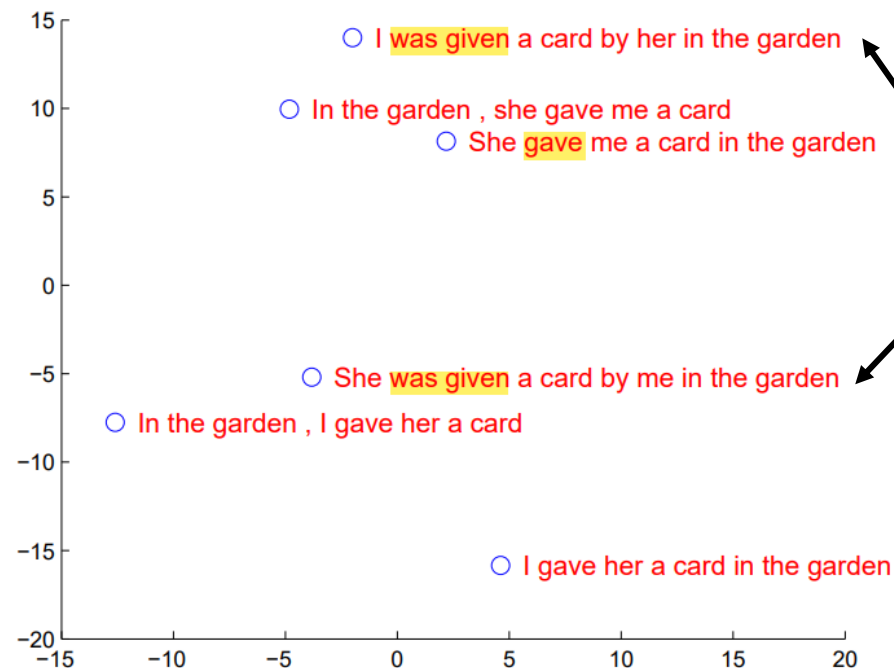
- Experiment:

- 모델 분석
- 능동태, 수동태에 상관없이 단어 순서에 민감함
- 능동태를 수동태로 대체하는데 둔감함

A 2-dimensional PCA projection of the LSTM hidden states



Active voice
(능동태)



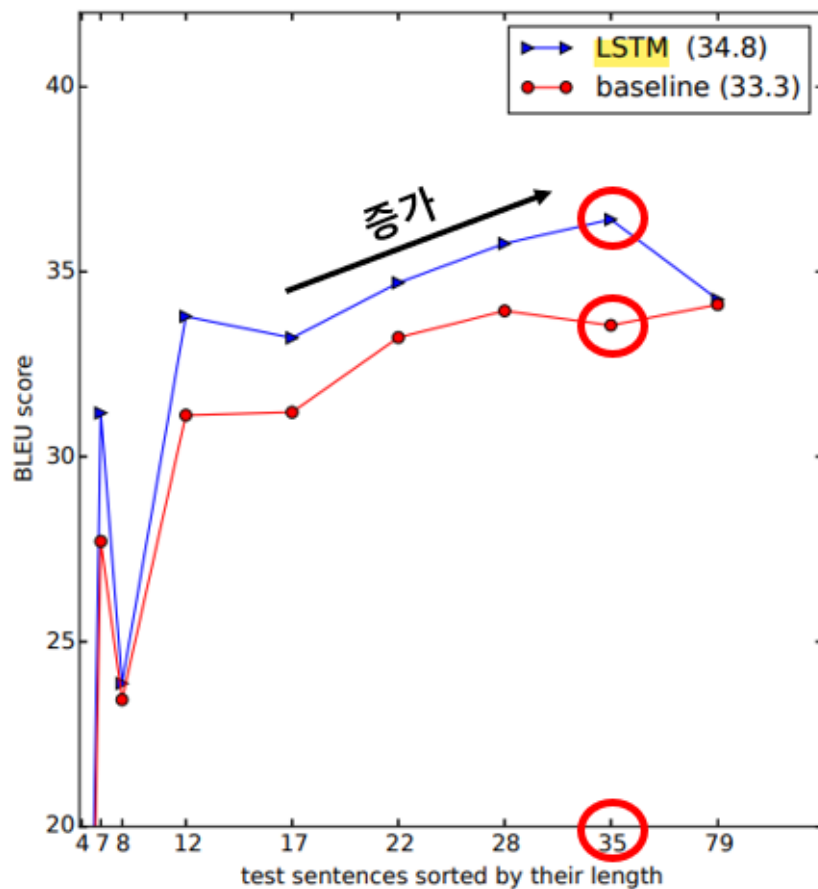
Passive voice
(수동태)

문법 오류

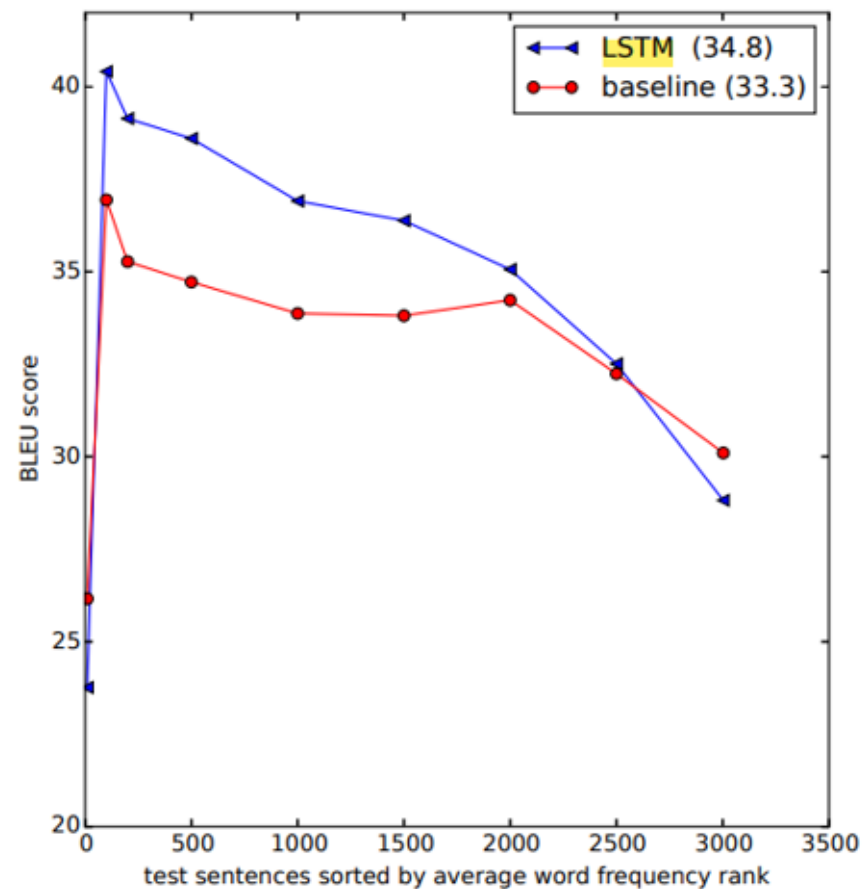
Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- Experiment:

- Long sentences에서도 좋은 성능을 보임



문장 길이에 따른 성능 비교



단어 빈도 순위에 따른 성능 비교

Sequence to Sequence Learning with Neural Networks (NIPS 2014)

- 한줄평:
 - ✓ 본 논문은 RNN 기반의 encoder-decoder 구조에 Deep LSTM을 활용하고, reversed order input sequence를 사용해 기존 DNN의 한계점을 개선했다.
 - ✓ 무엇보다 입력 문장의 순서를 바꿔 성능을 개선한 아이디어는 가장 기억에 남는 것 같다.
 - ✓ 최적화되지 않은 Deep LSTM이 Baseline (SMT)보다 뛰어난 성능을 나타내는 것을 발견한 점은 관련 연구가 더욱 발전할 수 있다는 점에서 상당히 의미있는 논문이라 생각한다