

Sequence to Sequence Learning with Neural Networks

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le.
Advances in neural information processing systems 27 (2014).

2024.01.18
유하영

- 이전 연구의 한계

DNN

입력과 출력의 차원이 고정된 문제에서만 사용 가능

sequence와 같이 가변적인 길이의 데이터를 처리하는 데에는 제한적

Ex) 음성인식, 기계번역

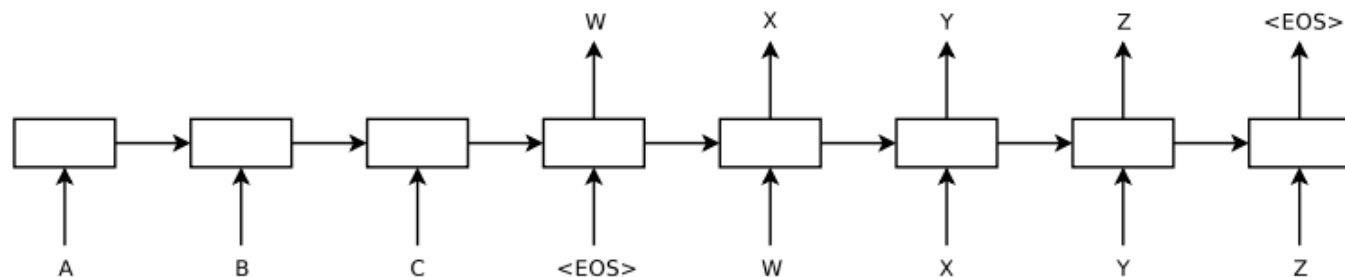
- 아이디어 제시

LSTM 구조로 **sequence to sequence** 문제 해결이 가능하다고 봄

한 개의 **LSTM**이 한번에 한 **timestep** 씩 **input sequence**를 읽어서 거대한 고정 차원수의 **vector representation**을 얻고

또 다른 **LSTM**이 **vector representation**으로부터 **target sequence** 추출

-> 시간에 의존적인 긴 데이터 학습이 가능해짐



- RNN

inputs sequence가 주어질 때,
표준 RNN은 다음의 방정식을 반복하면서 output sequence을 계산해낸다

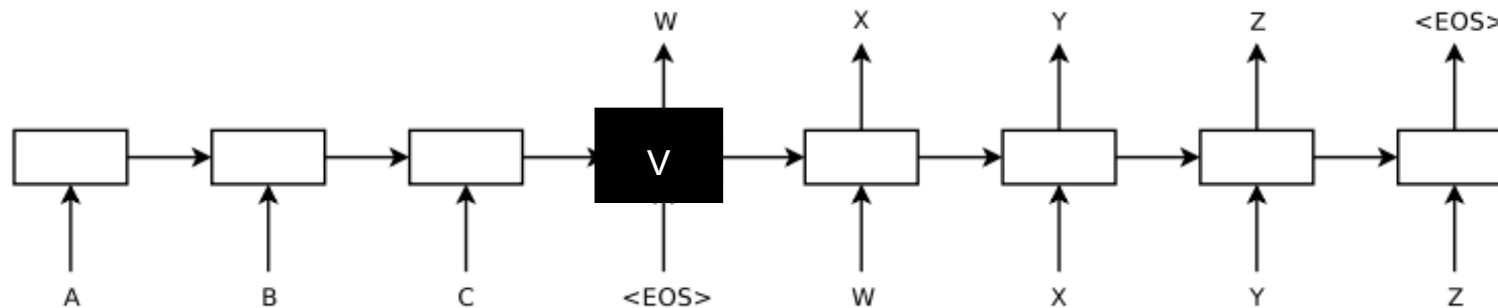
$$h_t = \text{sigm}(W^{hx}x_t + W^{hh}h_{t-1})$$
$$y_t = W^{yh}h_t$$

- LSTM

LSTM의 최종 목표는 input sequence와 output sequence에 대해

조건부 확률 $p(y_1, \dots, y_{T'} | x_1, \dots, x_T)$ 측정하는 것

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



- LSTM

1. 입력과 출력에 대한 서로 다른 **2개의 LSTM**을 사용한다.

2. Deep LSTM의 Shallow LSTM 보다 좋은 성능을 제공한다.
(layer 4 deep LSTM)

3. 입력의 순서를 뒤집어서 제공한다.

A B C -> D E F

C B A -> D E F

Reversing the Source Sentences

- LSTM learns much better when the source sentences are **reversed** solving problems with **long term dependencies**

명확한 설명은 X

가설)

- 초반 단어의 영향이 최종 **hidden state**에 미치는 영향 증가
(**short term dependencies**을 도입한 것)
- 입력 문장을 뒤집어도 **source word** 와 **target word** 사이 거리의 평균을 동일하게 유지된다.

하지만, "**minimal time lag**" 가 크게 줄어들기 때문에 성능의 향상이 있을 수 있다고 설명

perplexity(혼란도) : 5.8-> 4.7

번역문의 test BLEU 점수 : 25.9-> 30.6

perplexity:

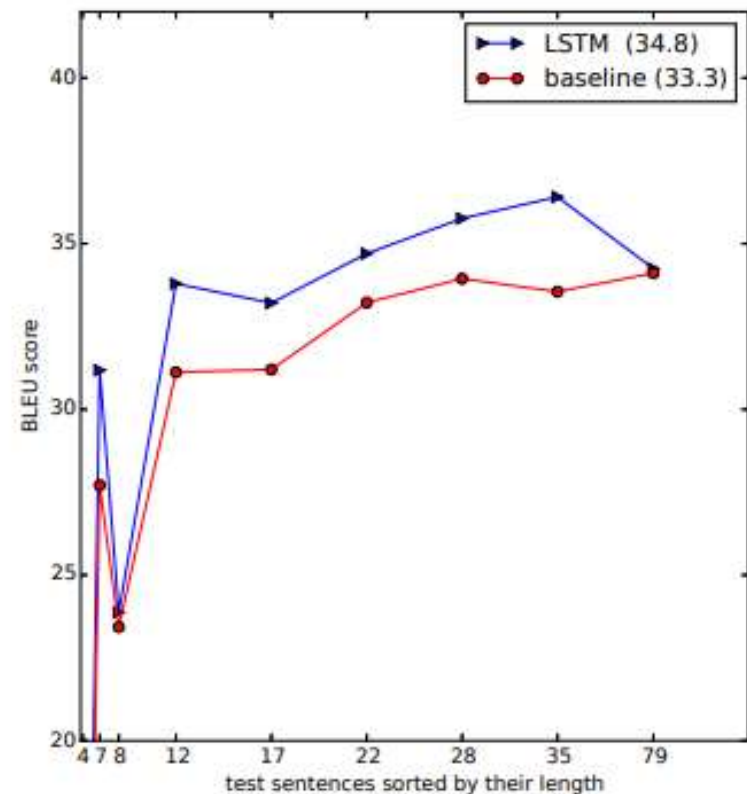
언어 모델 성능 측정 지표 중 하나로 모델이 내놓은 답의 혼란한 정

도

Experimental Results

- 각 LSTM layer 마다 서로 다른 초기화 과정을 거치고 minibatch를 랜덤하게 섞어 학습한 LSTM 앙상블로부터 best result를 얻었다.

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



- BLEU(Bilingual Evaluation Understudy)

기계 번역 결과와 사람이 직접 번역한 결과가 얼마나 유사한지 비교하여 번역에 대한 성능을 측정하는 방법

- 대규모 **deep LSTM**이 대규모 기계 번역 작업에 있어 무제한의 어휘록을 가진 **standard SMT(Statistical Machine Translation)** 기반 시스템 보다 더 높은 성능을 발휘함
- **source sentences**의 단어를 역순으로 배치하는 것이 더 높은 성능을 보임
- **LSTM**은 매우 긴 문장도 거의 올바르게 번역하였다.
(**but**, 아주 긴 문장을 역순으로 배치하여 학습할 때는 아직 한계가 보임)

- **SMT(Statistical Machine Translation)**

: 통계적 기법을 사용하여 한 언어에서 다른 언어로 텍스트를 번역하는
기계 번역의 한 형태