

[Week 5] 2022.08.04.

Sequence-to-Sequence (Seq2Seq)

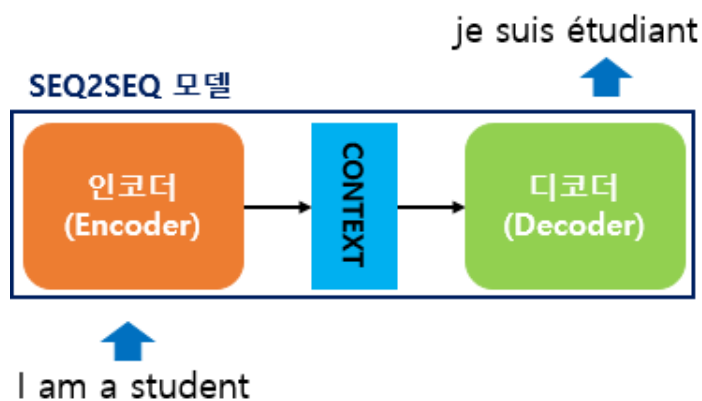
KUBIG 15기 수학과 이승은

1. Seq2Seq란?

입력된 시퀀스로부터 다른 도메인의 시퀀스를 출력

e.g.) Chatbot (input: 질문, output: 대답), Machine Translation (input: 입력 문장, output: 번역 문장),
내용 요약 (Text Summarization), STT (Speech to Text)

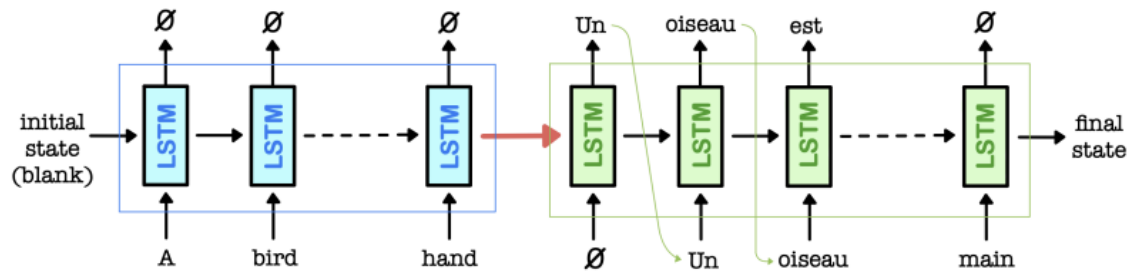
2. Seq2Seq의 전체적인 구조



Encoder: 입력 문장의 단어들을 순차적으로 입력 받은 후 이 단어들의 정보를 압축해서 하나의 "컨텍스트 벡터" (context vector)로 만든다.

Decoder: context vector를 받아서 번역된 단어를 한 개씩 순서대로 출력한다.

3. Seq2Seq 내부 구조



- Encoder, Decoder는 각각 RNN 구조 (실제로는 성능 문제로 LSTM, GRU 사용)

1) Test

- Decoder는 기본적으로 RNNLM (RNN Language Model)

- 테스트 과정 동안의 Decoder 작동방식:

초기 입력으로 문장의 시작을 의미하는 심볼 <sos>가 들어감.

디코더는 <sos>가 입력되면, 다음에 등장할 확률이 높은 단어를 예측.

첫번째 time step의 디코더 RNN 셀은 다음에 등장할 단어로 je를 예측.

이후 첫번째 시점의 디코더 RNN 셀은 예측된 단어 je를 다음 시점의 RNN 셀로 입력.

두번째 시점의 디코더 RNN 셀은 입력된 단어 je로부터 다시 다음에 올 단어인 suis를 예측.

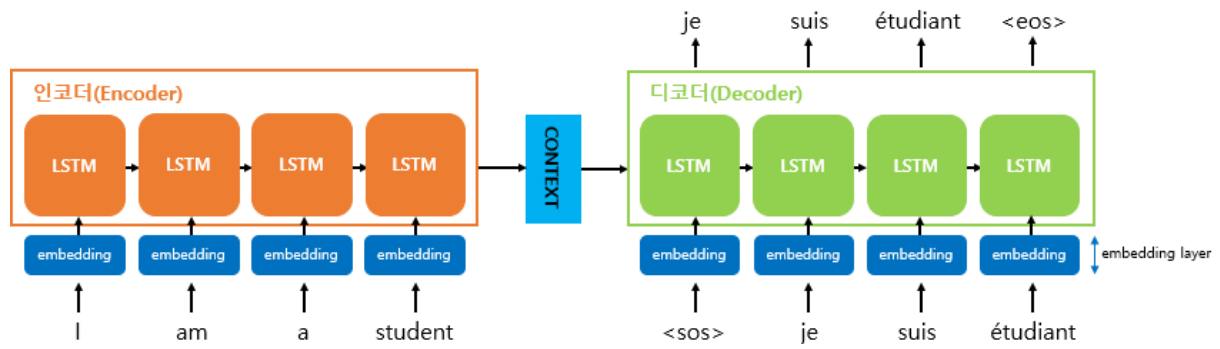
다시 다음 시점의 RNN 셀로 입력.

- 정리: 기본적으로 다음에 올 단어를 예측하고, 그 예측한 단어를 다음 시점의 RNN 셀의 입력으로 넣는 행위를 반복. 문장의 끝을 의미하는 심볼인 <eos>가 다음 단어로 예측될 때까지 반복.

2) Train

“seq2seq는 훈련 과정과 테스트 과정 (또는 실제 번역기 사용 시)의 작동 방식이 조금 다르다.”

훈련 과정에서 디코더에게 인코더가 보낸 컨텍스트 벡터와 실제 정답인 상황인 <sos> je suis étudiant를 입력 받았을 때, je suis étudiant <eos>가 나와야 된다고 정답을 알려주면서 훈련.

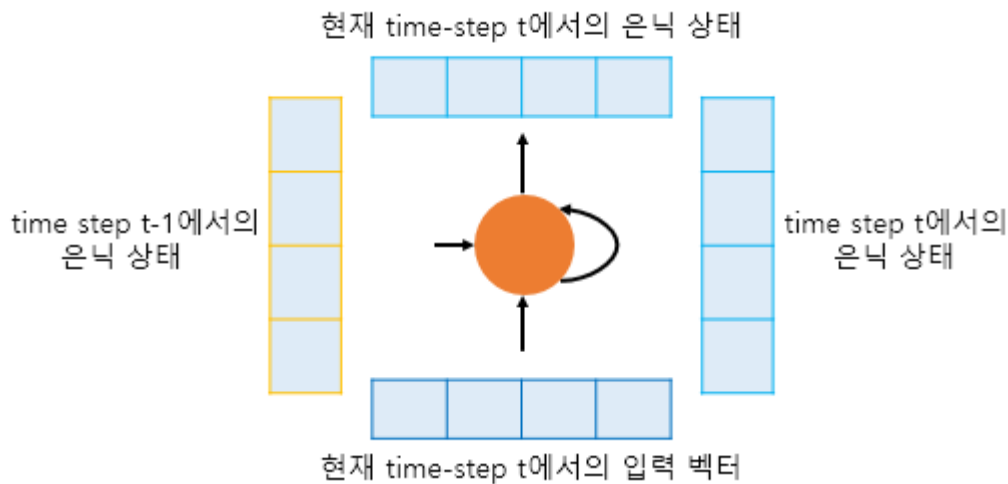


워드 임베딩 (텍스트를 벡터로 바꾸는 방법으로)

seq2seq에서 단어들을 임베딩 벡터로 변환 후 입력으로 사용: 임베딩 층(embedding layer)

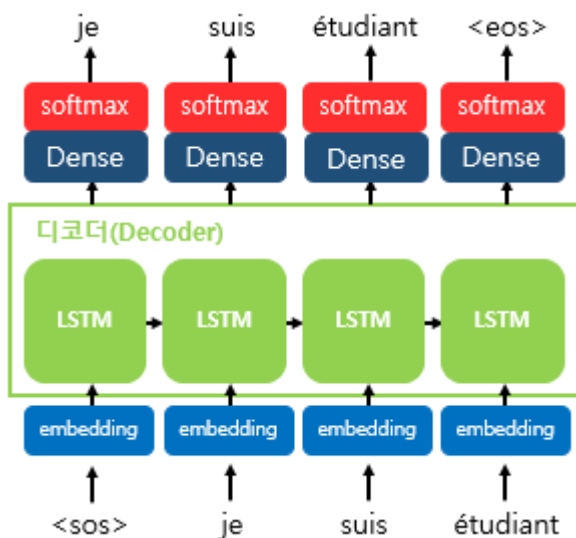
e.g.) 4-dim -> 실제로는 더 큰 차원

I		0.157
		-0.25
		0.478
		-0.78
am		0.78
		0.29
		-0.96
		0.52
a		0.75
		-0.81
		0.96
		0.12
student		0.88
		-0.17
		0.29
		0.48



현재 시점 (time step)을 t 라고 할 때, RNN 셀은 $t-1$ 에서의 은닉 상태와 t 에서의 입력 벡터를 입력으로 받고, t 에서의 은닉 상태를 만듦.

이러한 RNN 구조에서 현재 시점 t 에서의 은닉 상태는 과거 시점의 동일한 RNN 셀에서의 모든 은닉 상태의 값들의 영향을 누적해서 받아온 값이라고 할 수 있음. -> context vector는 인코더에서의 마지막 RNN 셀의 은닉 상태값이고, 입력 문장의 모든 단어 토큰들의 정보를 요약해서 담고 있음.



디코더는 인코더의 마지막 RNN 셀의 은닉 상태인 컨텍스트 벡터를 첫번째 은닉 상태의 값으로 사용. 디코더의 첫번째 RNN 셀은 이 첫번째 은닉 상태의 값과, 현재 t 에서의 입력값인 $< sos >$ 에서 다음에 등장할 단어를 예측. 그리고 이 예측된 단어는 다음 시점인 $t+1$ RNN에서의 입력값이 되고, 이 $t+1$ 에서의 RNN 또한 이 입력값과 t 에서의 은닉 상태로부터 $t+1$ 에서의 출력 벡터. 즉, 또 다시 다음에 등장할 단어를 예측.

이때, Dense (nn.Linear)와 Softmax 함수로 예측할 수 있는 여러 가지 단어 중 하나를 골라 예측.

=> Attention (LSTM)

어텐션 매커니즘: 컨텍스트 벡터보다 더욱 다양한 문맥을 반영할 수 있는 컨텍스트 벡터를 구하여 매 시점마다 하나의 입력으로 사용

4. Seq2Seq 구현

Ref.

<https://wikidocs.net/24996>