

# sequence-to-sequence

강사 : 백병인

[pi.paek@modulabs.co.kr](mailto:pi.paek@modulabs.co.kr)

모두의연구소 Research Scientist



2019 모두의연구소

# Language Model revisited

- Language Model은 이전에 나온 word sequence를 맥락으로 해서 그 다음 나올 sequence를 추론하는 sequence model이다..

$$P(w_1, w_2, \dots, w_m) = P(w_1) \prod_{i=2}^m P(w_i | w_{i-1}, \dots, w_{i-n})$$

- 만약 이것을 정확하게 모델링할 수 있다면, 거꾸로 Language Generation도 가능하지 않을까?

$$P(w_i | w_{i-1}, \dots, w_{i-n})$$

# Language Model은 Context가 필요

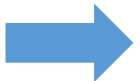
- 빈 칸에 알맞은 말을 규정하는 것은 Context입니다.

상황 / 맥락

나는 지금 배고프다.

선배가 후배에게

엄마가 부엌에서



언어 표현

나는 밥을 먹을래.

나는 밥을 사줄께.

나는 밥을 짓는다.

# Conditional Language Model

- 그렇다면 맥락  $x$ 를 Language Model에 반영해서 Text를 생성해 보면 어떨까?

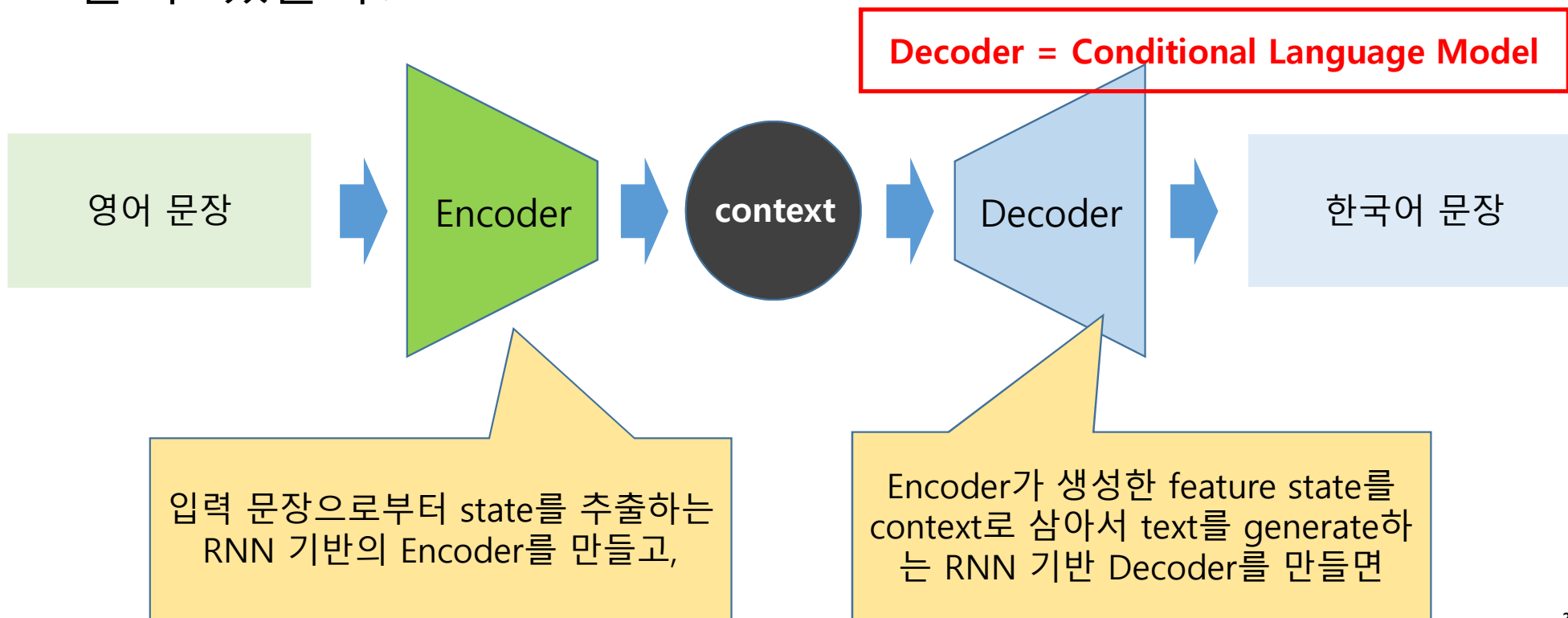
A **conditional language model** assigns probabilities to sequences of words,  $\mathbf{w} = (w_1, w_2, \dots, w_\ell)$ , given some conditioning context,  $\mathbf{x}$ .

As with unconditional models, it is again helpful to use the chain rule to decompose this probability:

$$p(\mathbf{w} \mid \mathbf{x}) = \prod_{t=1}^{\ell} p(w_t \mid \mathbf{x}, w_1, w_2, \dots, w_{t-1})$$

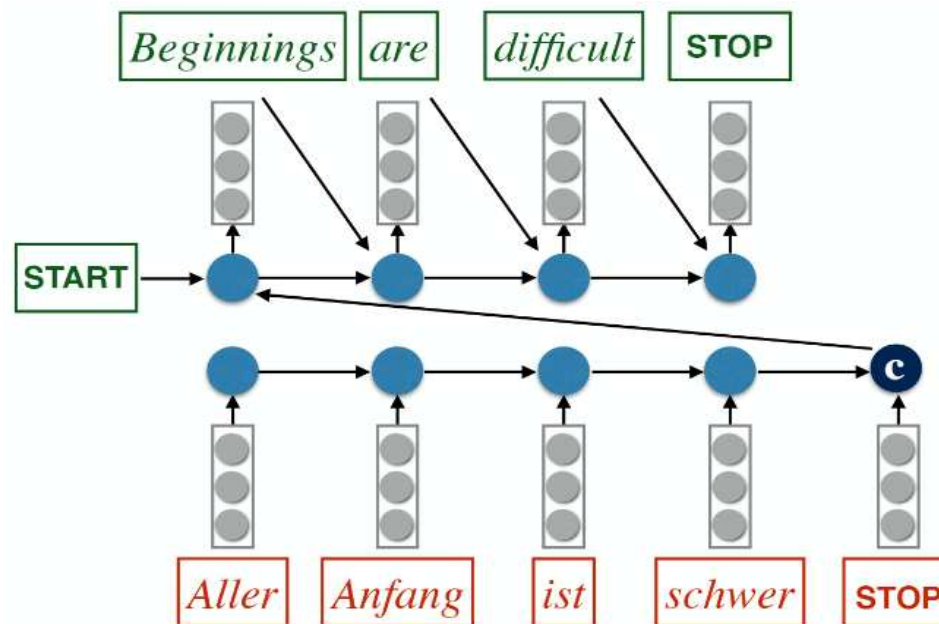
# Neural Translation

- 영어 문장을 한국어로 번역하는 task를 딥러닝으로 어떻게 구성할 수 있을까?



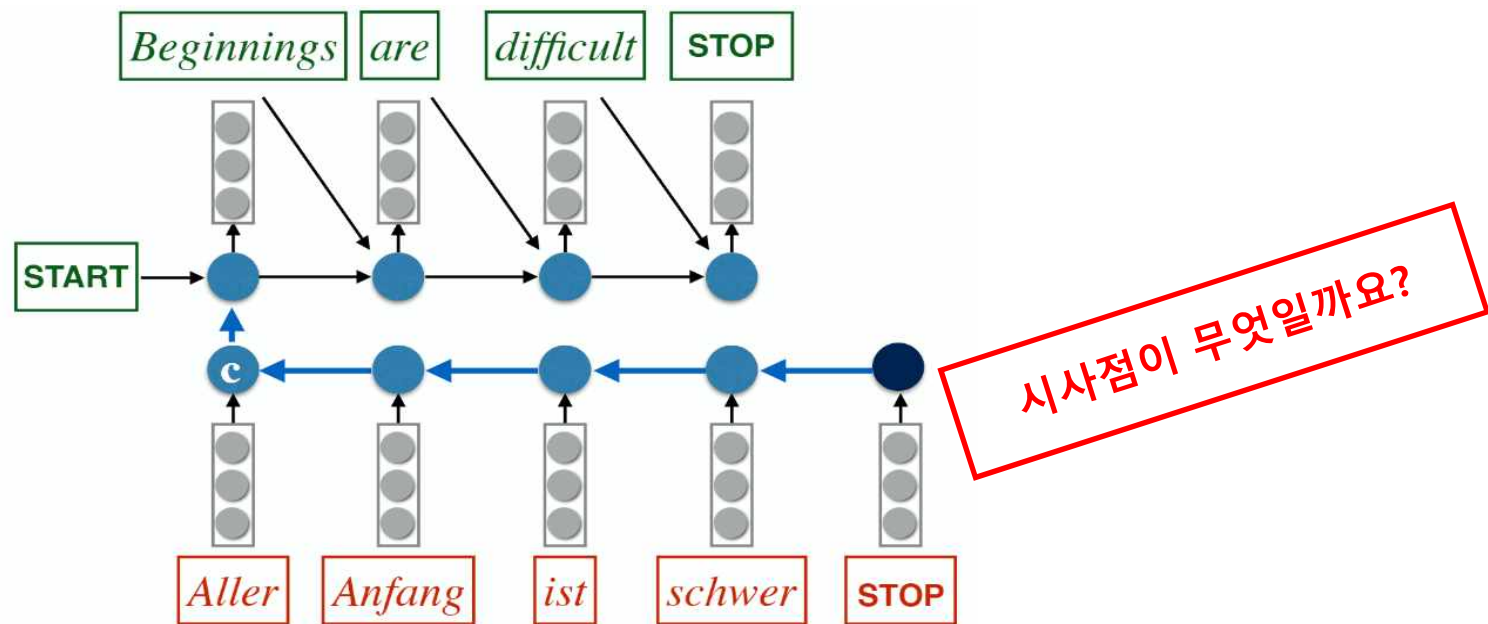
# seq2seq

- Sutskever et al. (2014) Sequence to Sequence Learning with Neural Networks



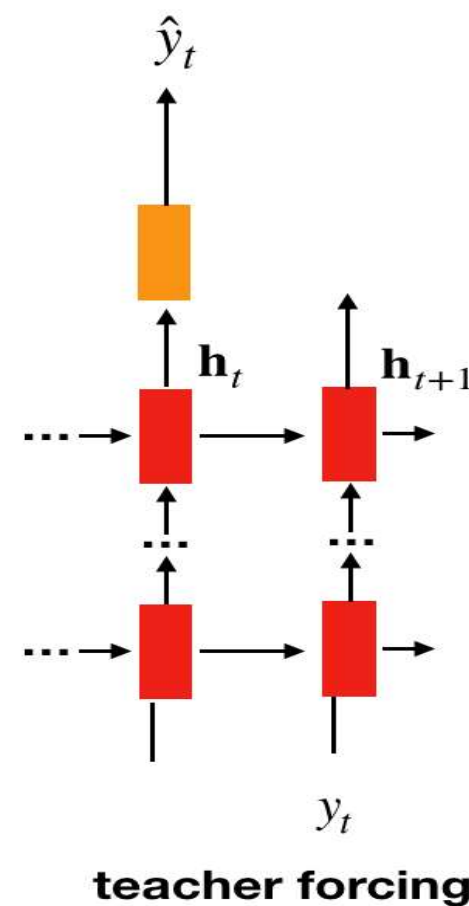
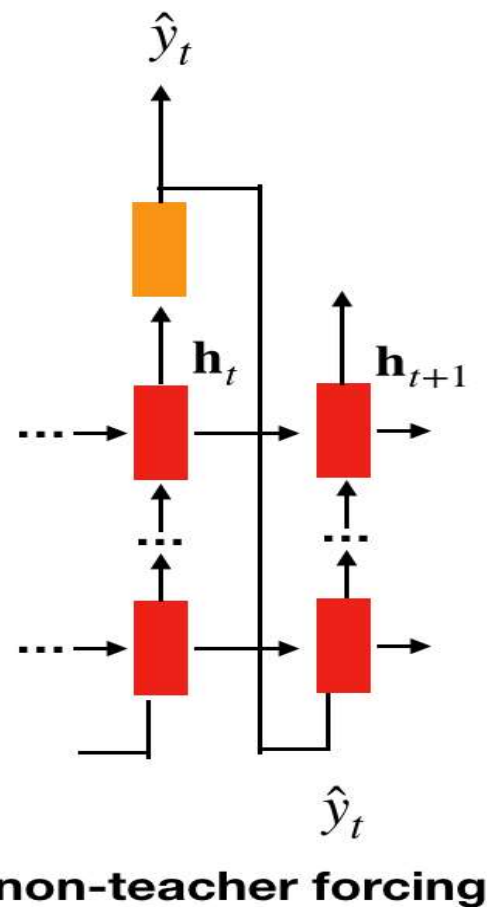
# seq2seq reversed input

- 입력문장의 어순을 거꾸로 처리했더니 번역 품질이 훨씬 좋아졌다. (BLEU 기준 +4)



# Teacher Forcing

- seq2seq decoder의 train 시점에 decode의 input 을 뭘로 할 것인가?
- Teacher Forcing
  - model의 추론 결과가 아니라 ground truth를 다음 time-step의 입력으로 대신 넣어주자.
- Scheduled Sampling
  - 그래도 일정 비율로는 non-teacher forcing도 병행하자.

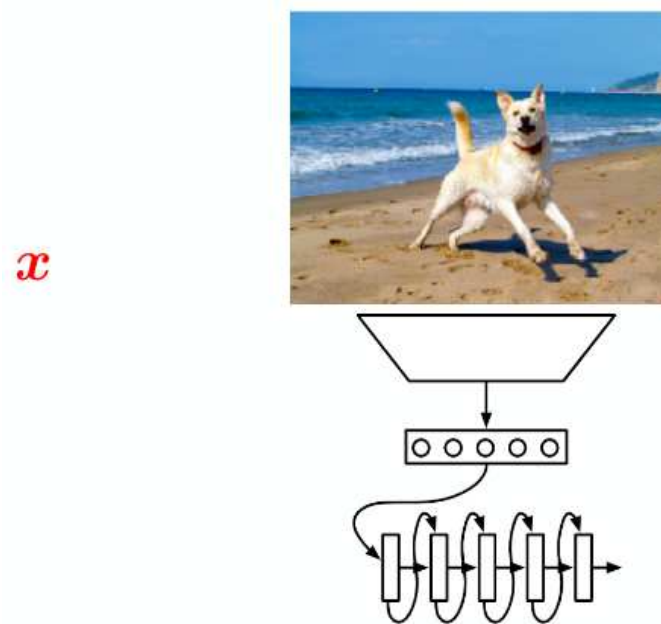




# X-to-Y Translation Pattern

- seq2seq는 NMT가 SMT보다 훨씬 저비용 고성능임을 입증
- Encoder-Decoder 구조를 통해 X를 Y로 변환하는 모델 대유행
- sentence-to-sentence
  - translation, Q&A, ...
- image-to-sentence
  - image captioning
- graph-to-graph, set-to-set, video-to-video, ....

Image Captioning



***A dog is playing on the beach.***