딥러닝 올인원

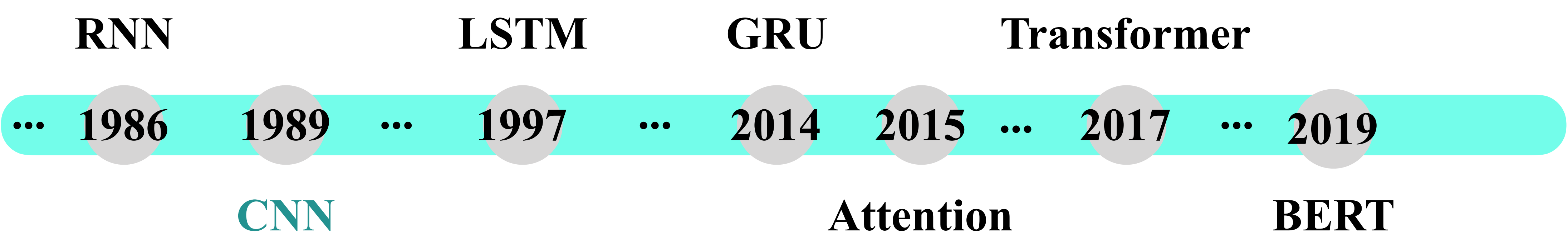
순환 신경망의 발전

18강

딥러닝호형

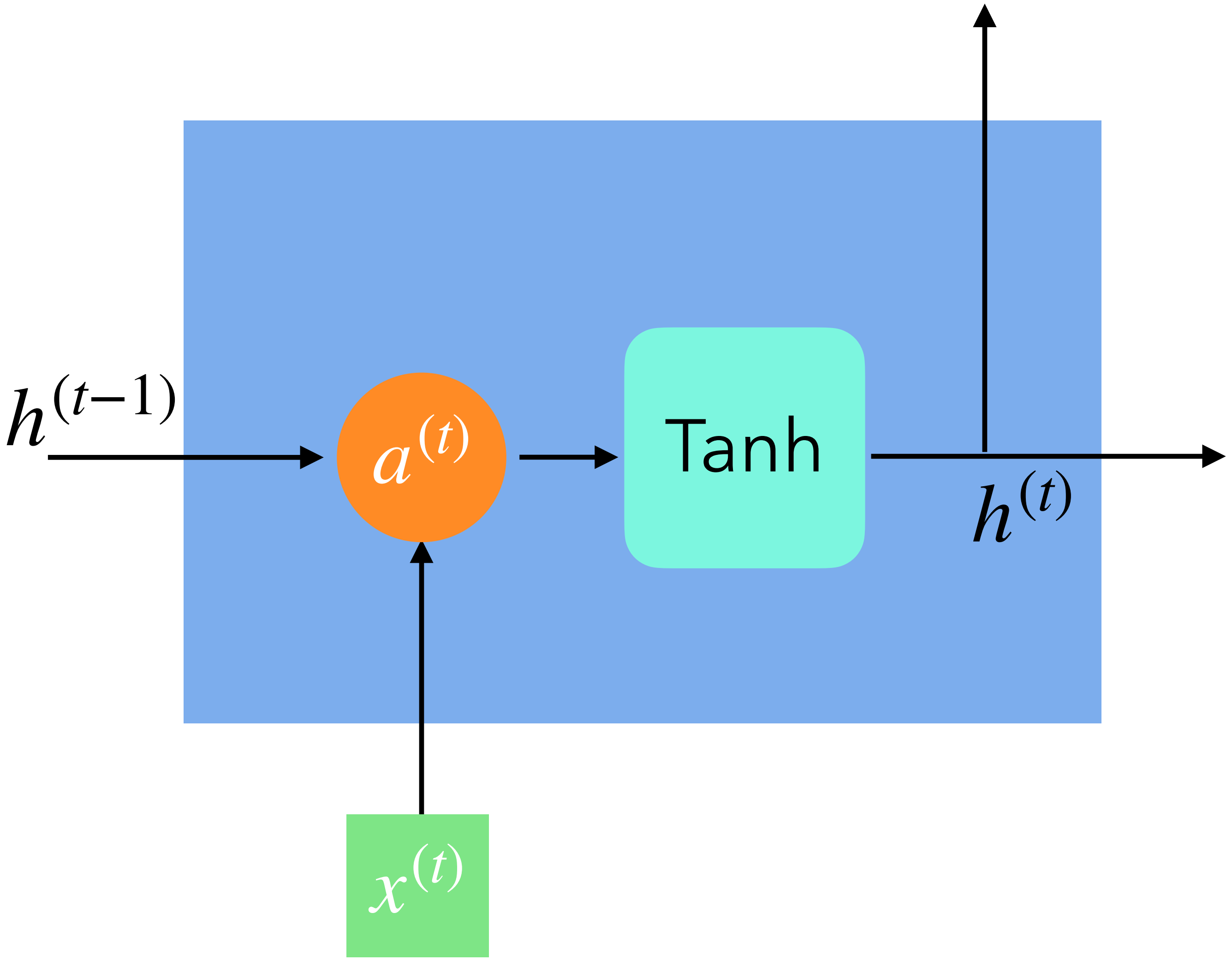
순환 신경망의 발전

RNN의 타임라인



순환 신경망의 발전

Vanilla RNN



$$h^{(t)} = f_W(h^{(t-1)}, x^{(t)})$$

$$a^{(t)} = b + Wh^{(t-1)} + Ux^{(t)}$$

$$a^{(t)}, h^{(t)} : N \times 1 \quad x^{(t)} : M \times 1$$

$$W : N \times N, U : N \times M, V : L \times N$$

$$h^{(t)} = tanh(a^{(t)})$$

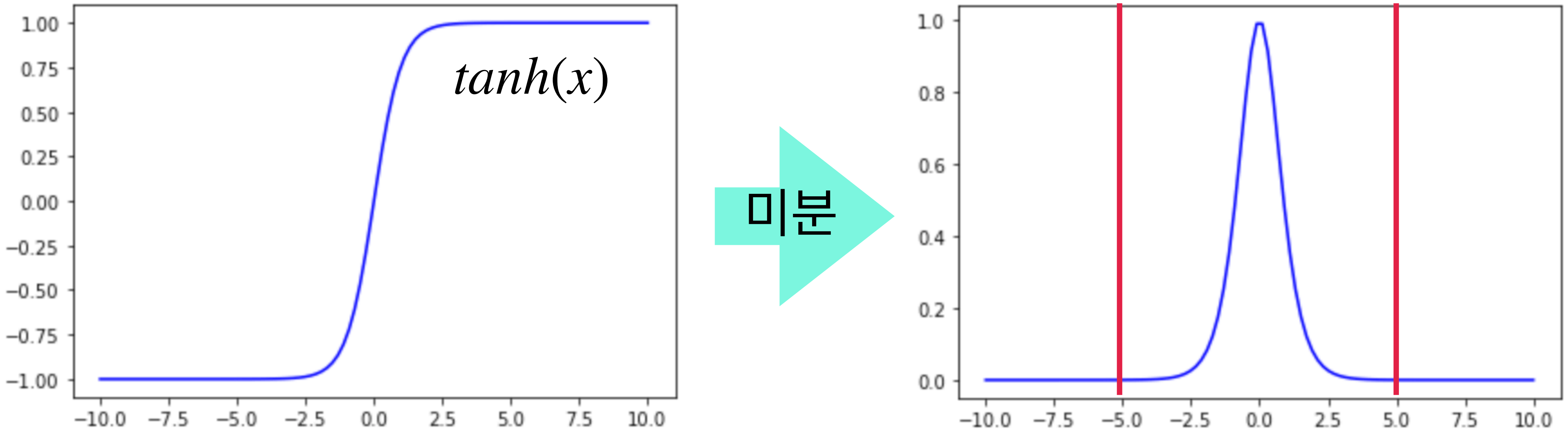
$$o^{(t)} = c + Vh^{(t)}$$

$$y^{(t)} = softmax(o^{(t)})$$

순환 신경망의 발전

Vanilla RNN의 문제점

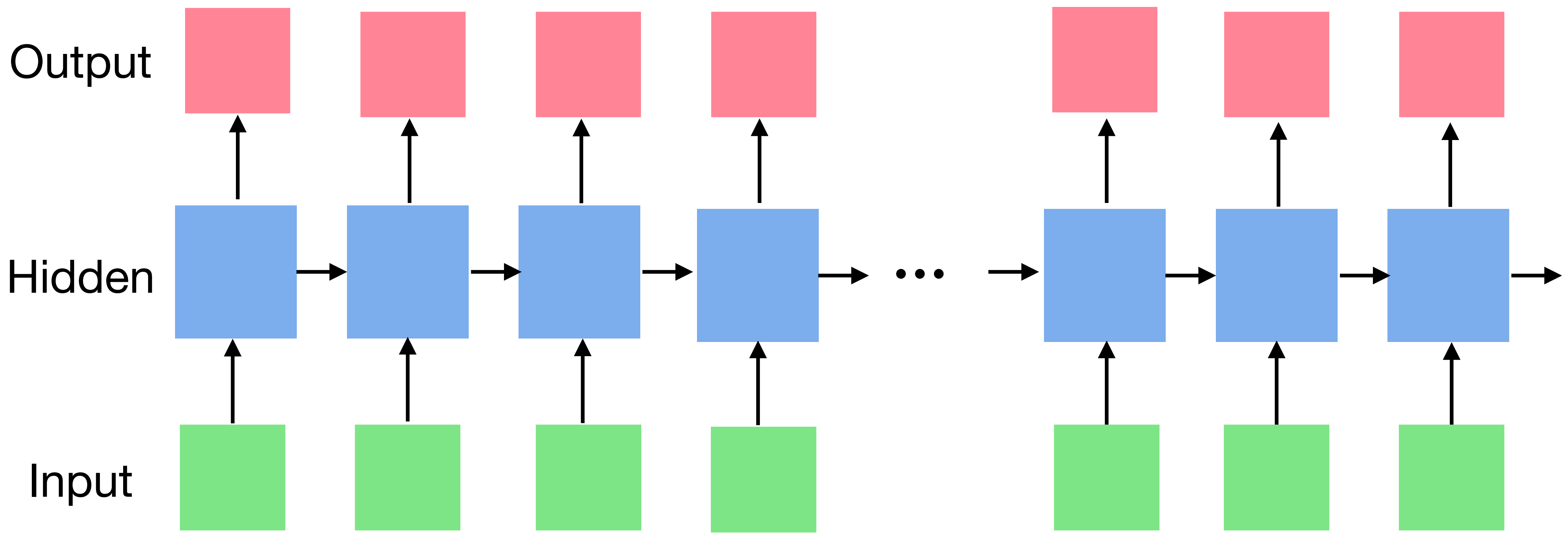
- 기울기 사라짐



순환 신경망의 발전

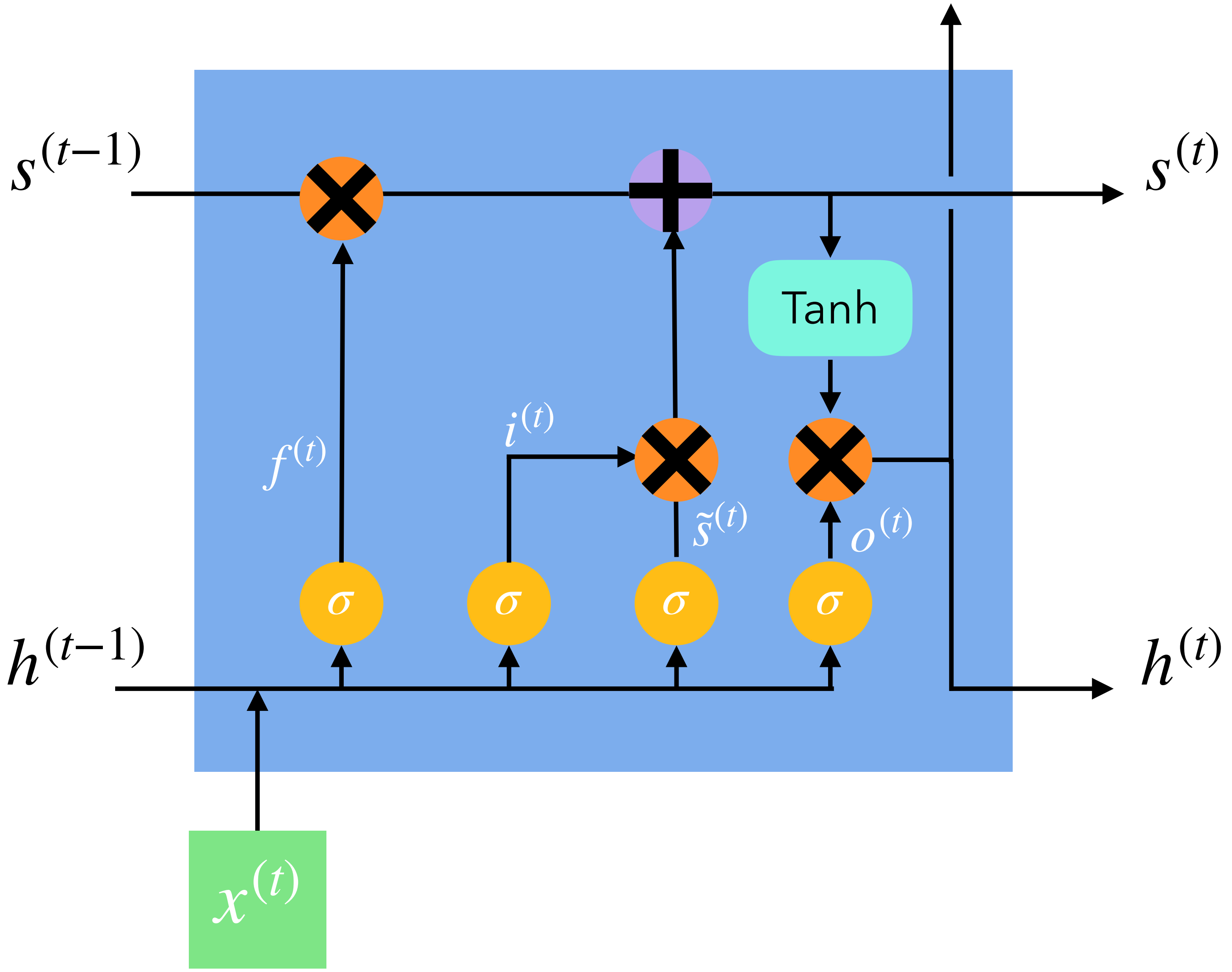
Vanilla RNN의 문제점

- 장기 의존성(Long term dependency)



순환 신경망의 발전

LSTM(Long Short-Term Memory models)



$$f^{(t)} = \sigma(W^f h^{(t-1)} + U^f x^{(t)})$$

$$i^{(t)} = \sigma(W^i h^{(t-1)} + U^i x^{(t)})$$

$$\tilde{s}^{(t)} = \sigma(W^g h^{(t-1)} + U^g x^{(t)})$$

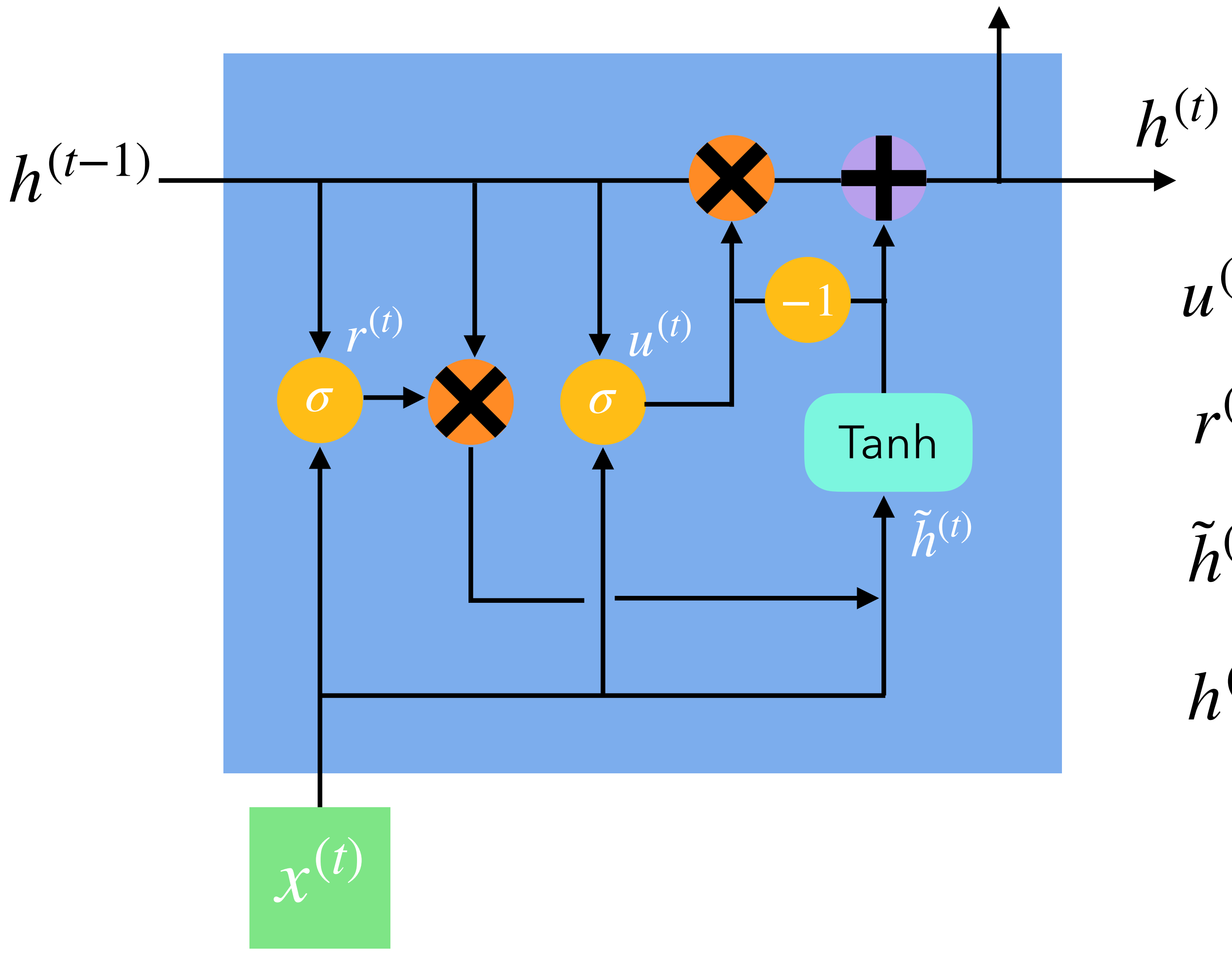
$$s^{(t)} = f^{(t)} s^{(t-1)} + i^{(t)} \tilde{s}^{(t)}$$

$$o^{(t)} = \sigma(W^o h^{(t-1)} + U^o x^{(t)})$$

$$h^{(t)} = o^{(t)} tanh(s^{(t)})$$

순환 신경망의 발전

GRU(Gated Recurrent Unit)



$$u^{(t)} = \sigma(W^u h^{(t-1)} + U^u x^{(t)})$$

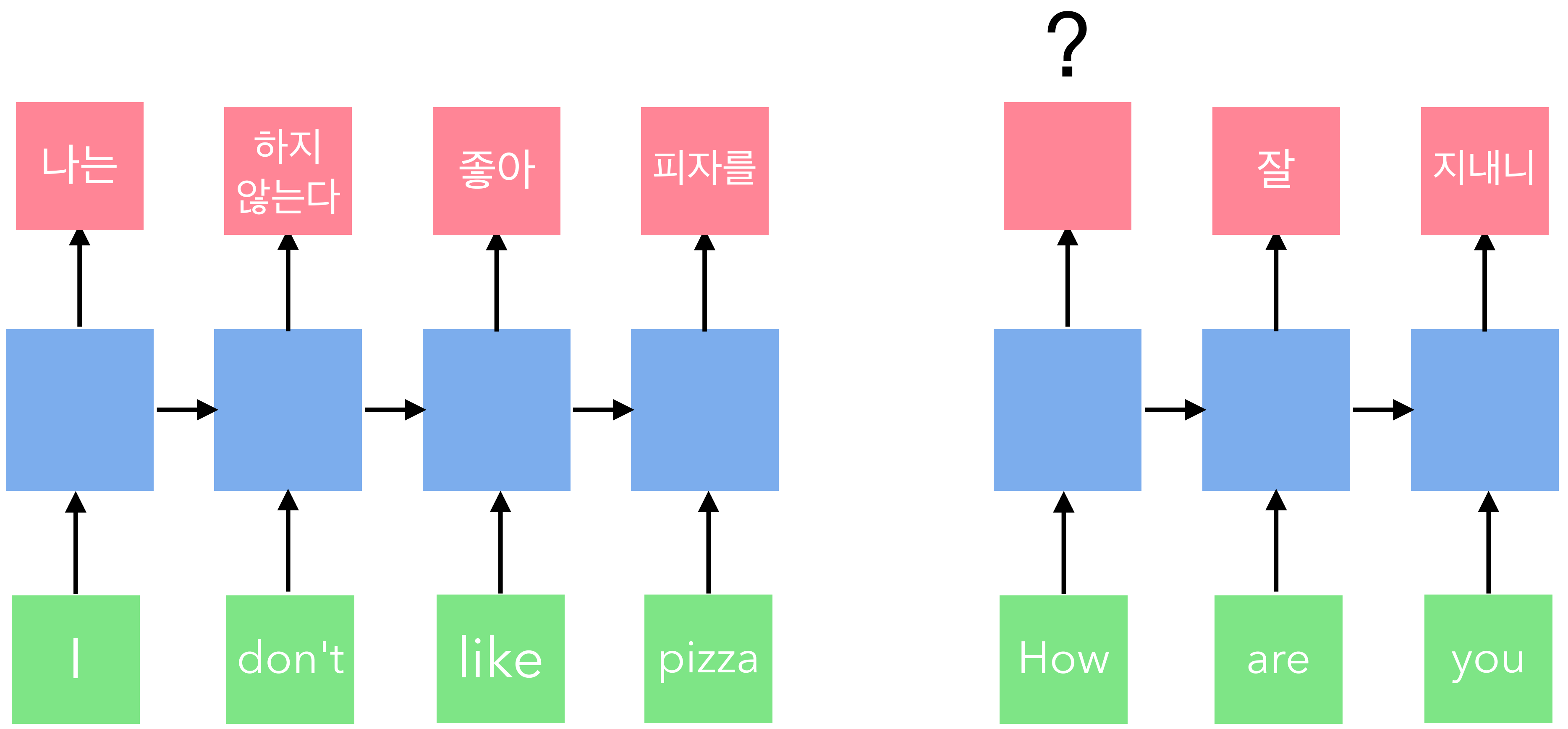
$$r^{(t)} = \sigma(W^r h^{(t-1)} + U^r x^{(t)})$$

$$\tilde{h}^{(t)} = tanh(W^h h^{(t-1)} \circ r^{(t)} + U^h x^{(t)})$$

$$h^{(t)} = u^{(t)} h^{(t-1)} + (1 - u^{(t)}) \tilde{h}^{(t)}$$

순환 신경망의 발전

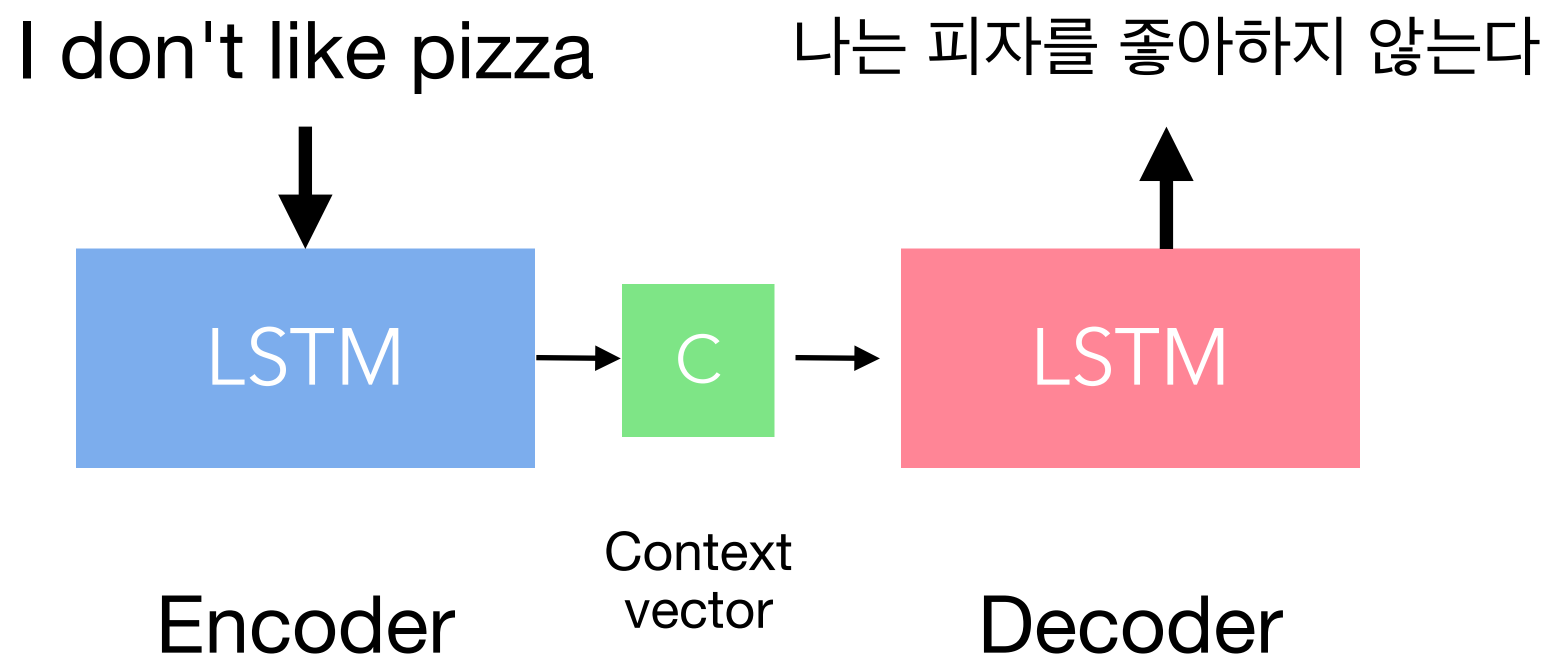
seq2seq(sqeuence-to-sqeuence)



입출력의 시퀀스 길이가 다르고 순서가 뒤 섞인다.

순환 신경망의 발전

seq2seq(sqeuence-to-sqeuence)



순환 신경망의 발전

Attention Mechanism - 어순의 차이를 극복