RNN 이론

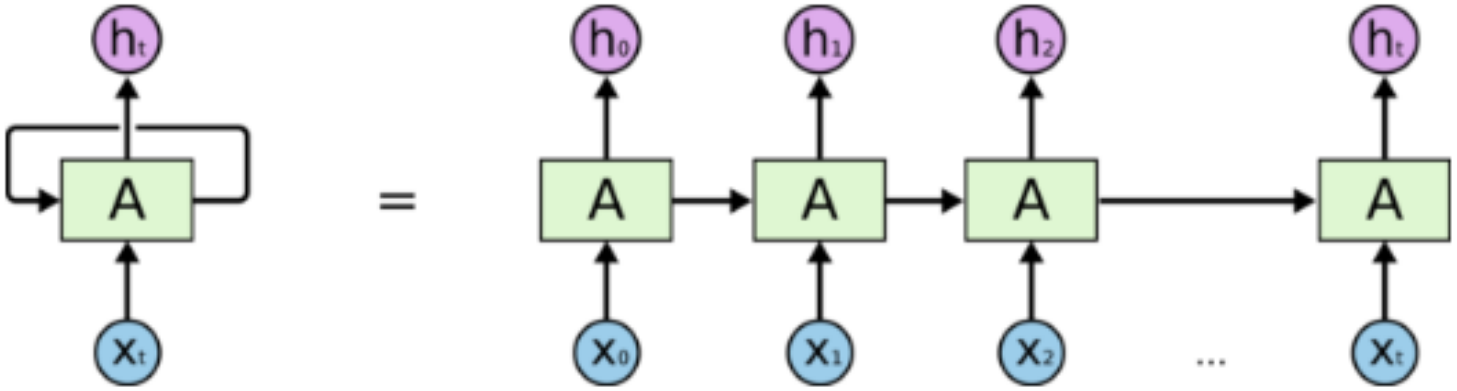
https://youtu.be/Y-uSwtFgnQ8

Contents

CryptoCraft LAB

RNN 이란

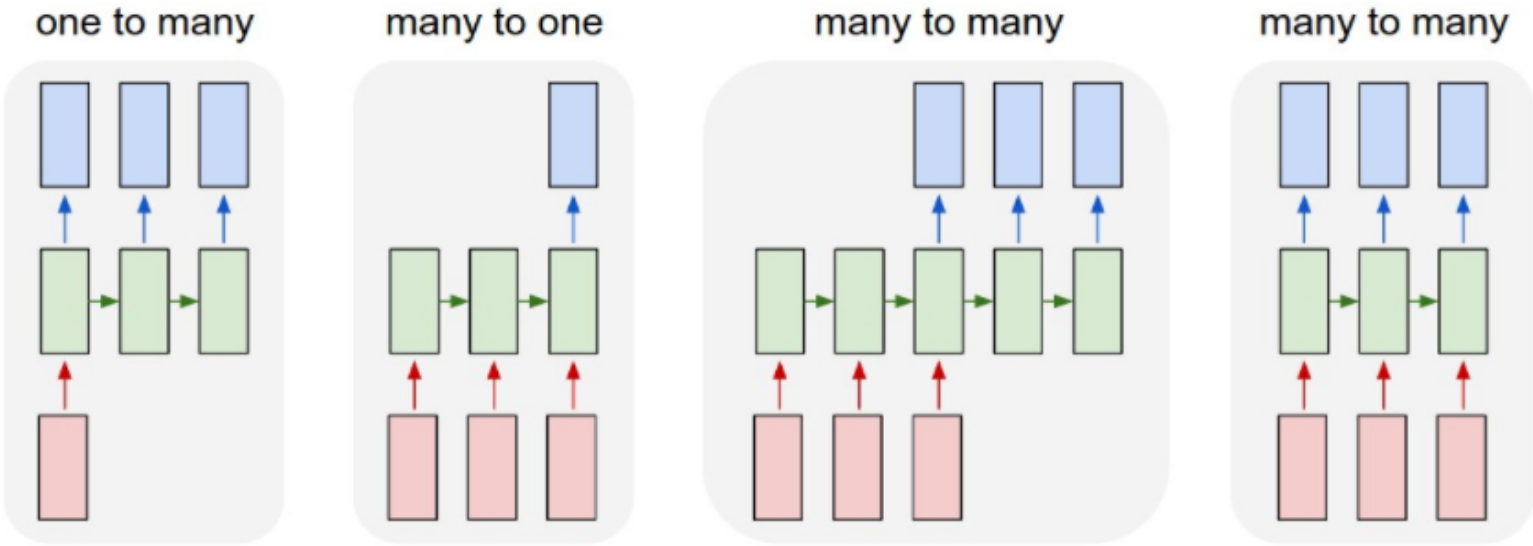
- 순환신경망 (Recurrent Neural Network)
  - 순서를 가지고 있는 데이터를 위한 모델



An unrolled recurrent neural network.

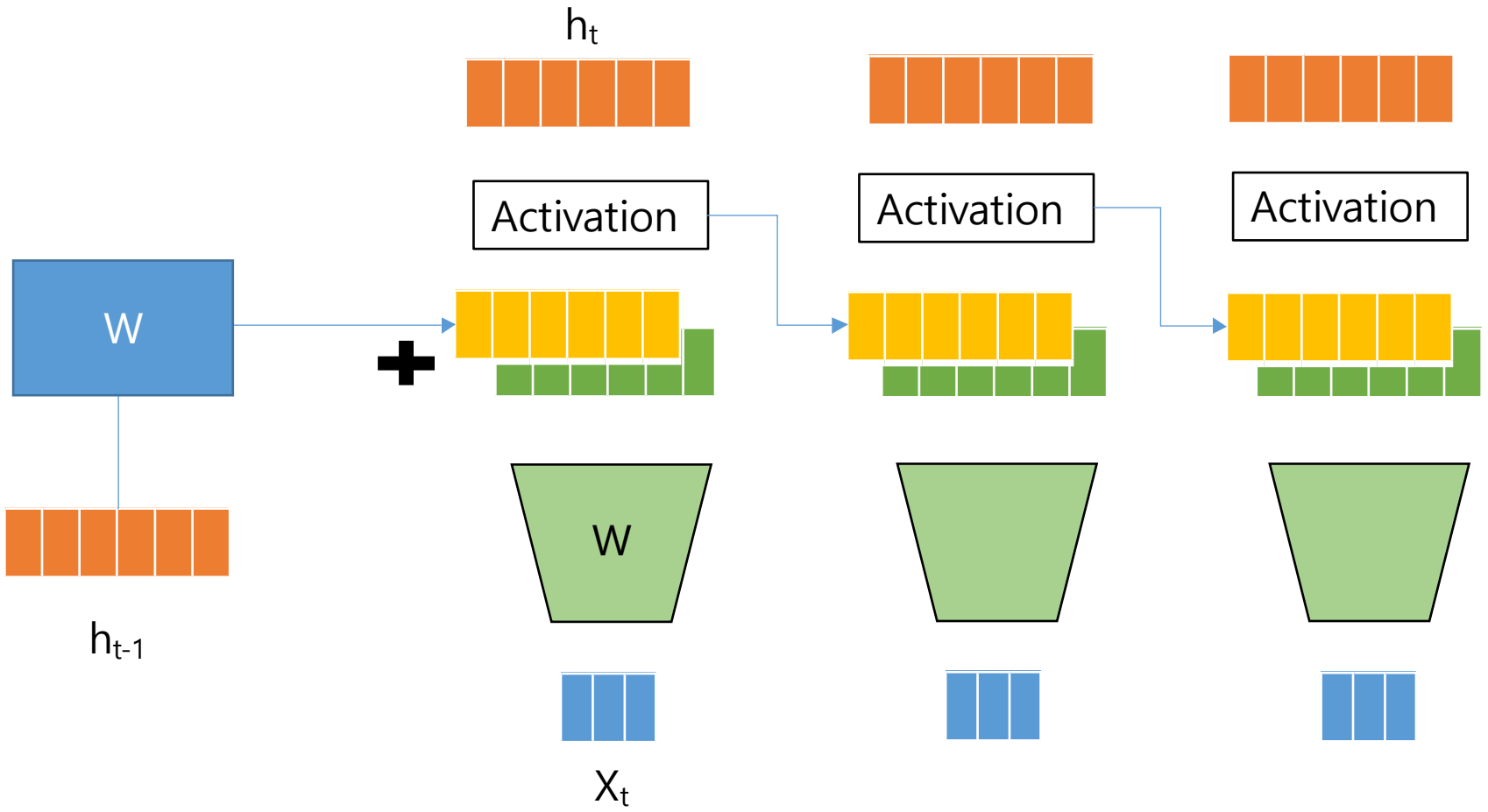
RNN 종류

- One-to-many
  - 이미지를 설명하는 문장 생성
- Many-to-one
  - 회사의 주가를 보고 흥망 예측
  - 감성 분석
- Many-to-many
  - 문장을 다른 언어로 번역
  - 문장의 각 품사를 분류



RNN

- 구조



RNN

- 활성화 함수에 비선형 함수를 사용하는 이유

Ex) h(x) = cx 일 때의 3층 네트워크

= y(x) = c * c * c * x

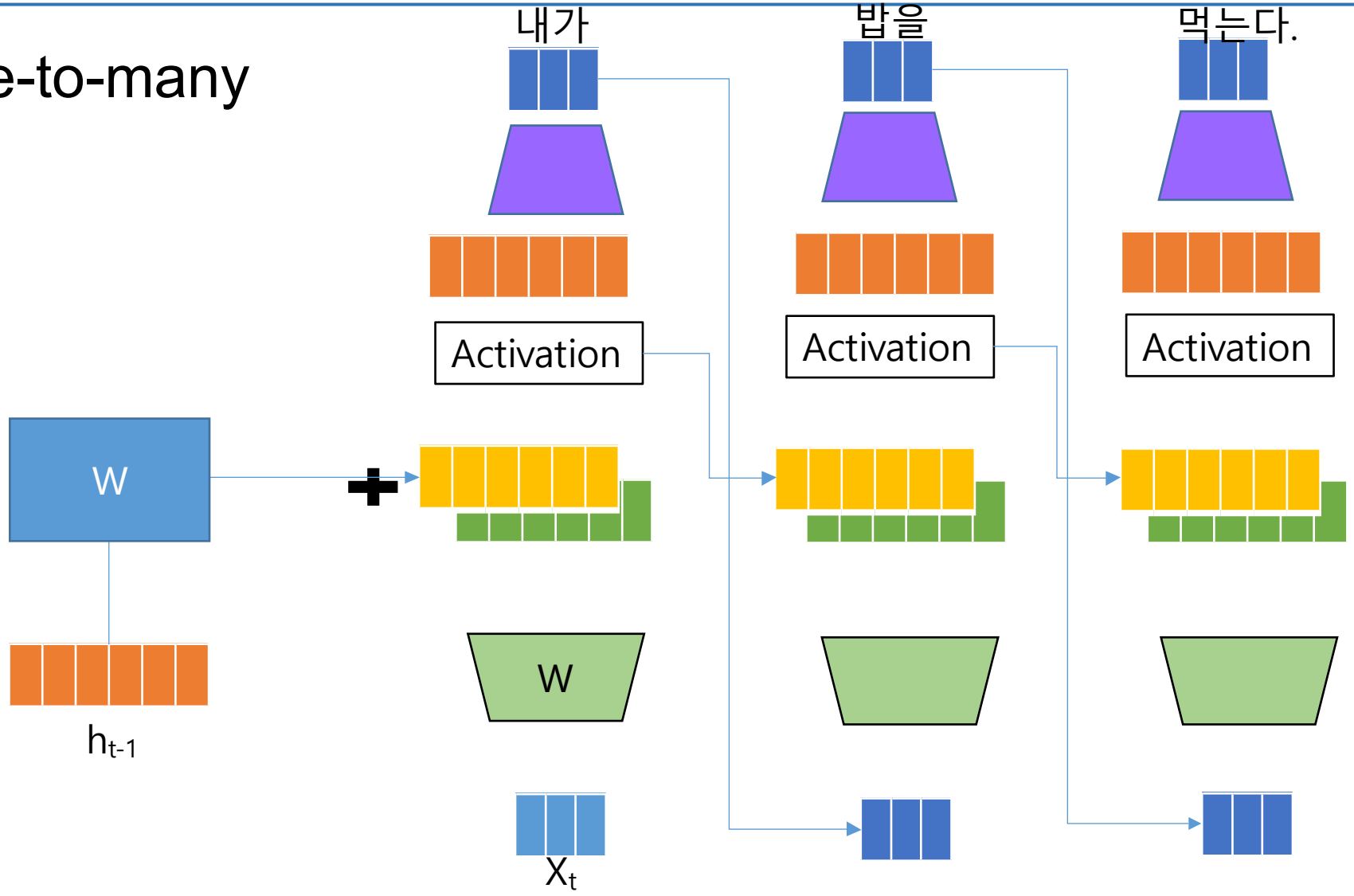
= c^3x

= ax

=> 층을 쌓는 이유가 없다!

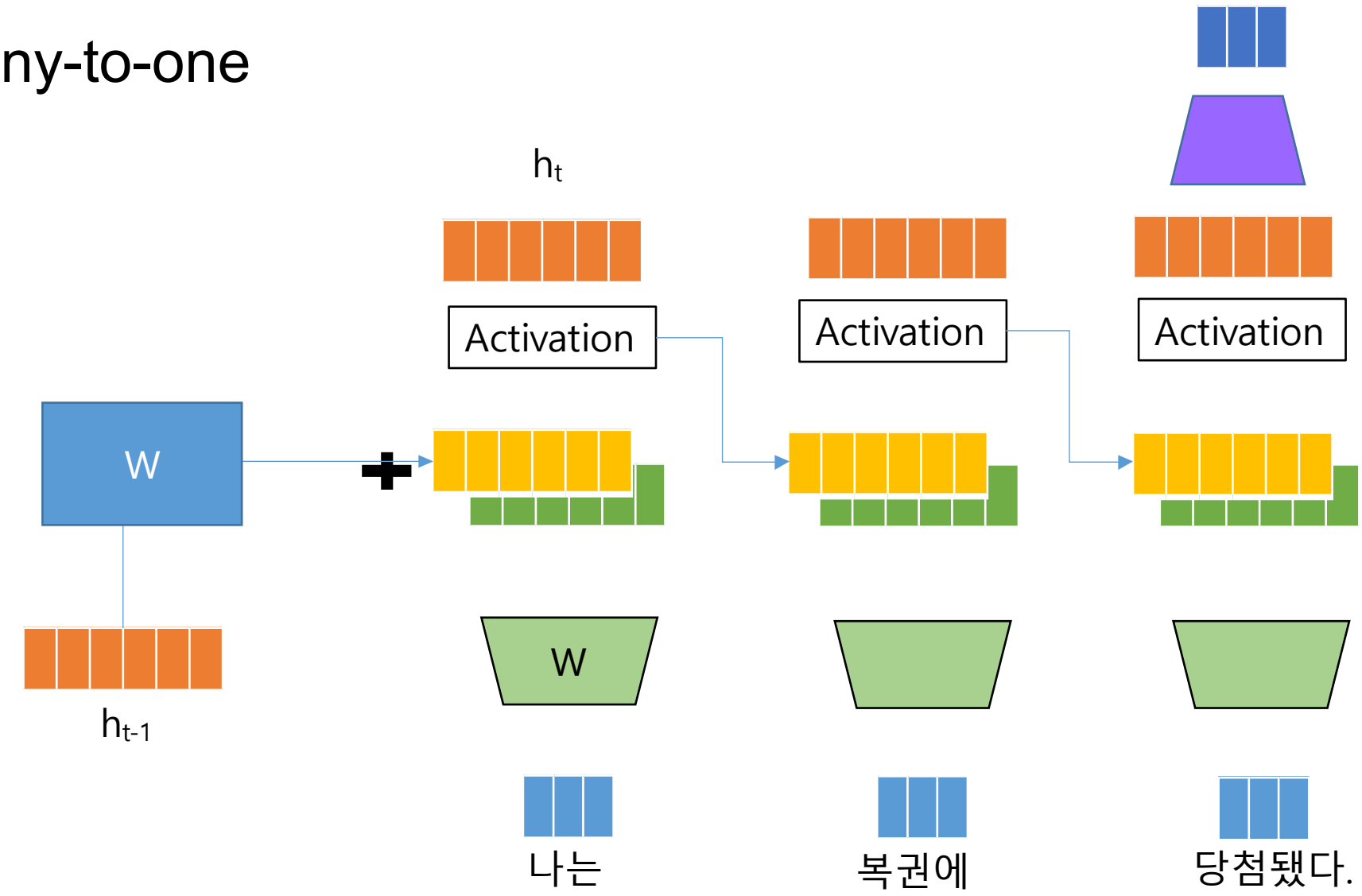
RNN

- One-to-many



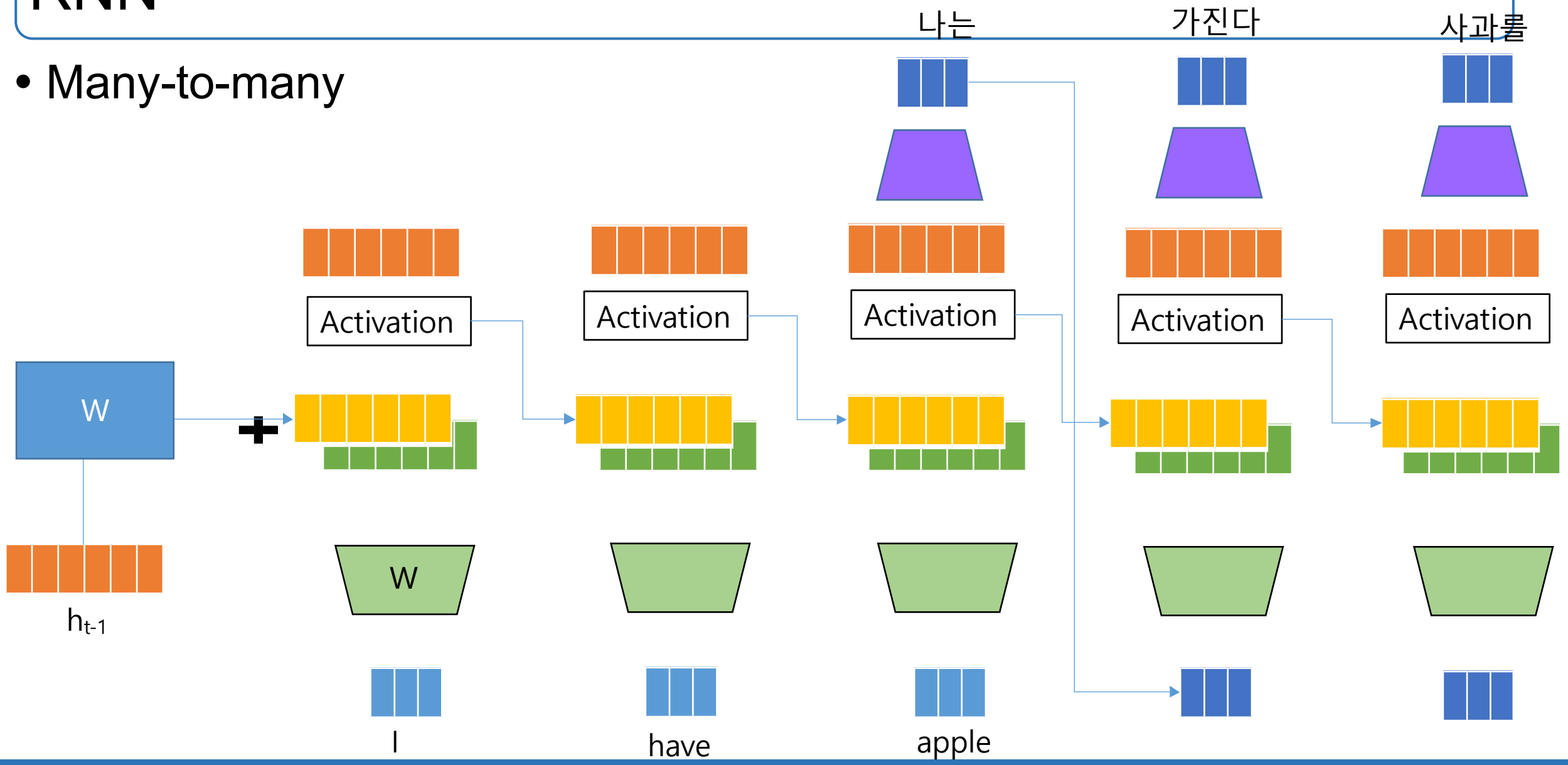
RNN

- Many-to-one



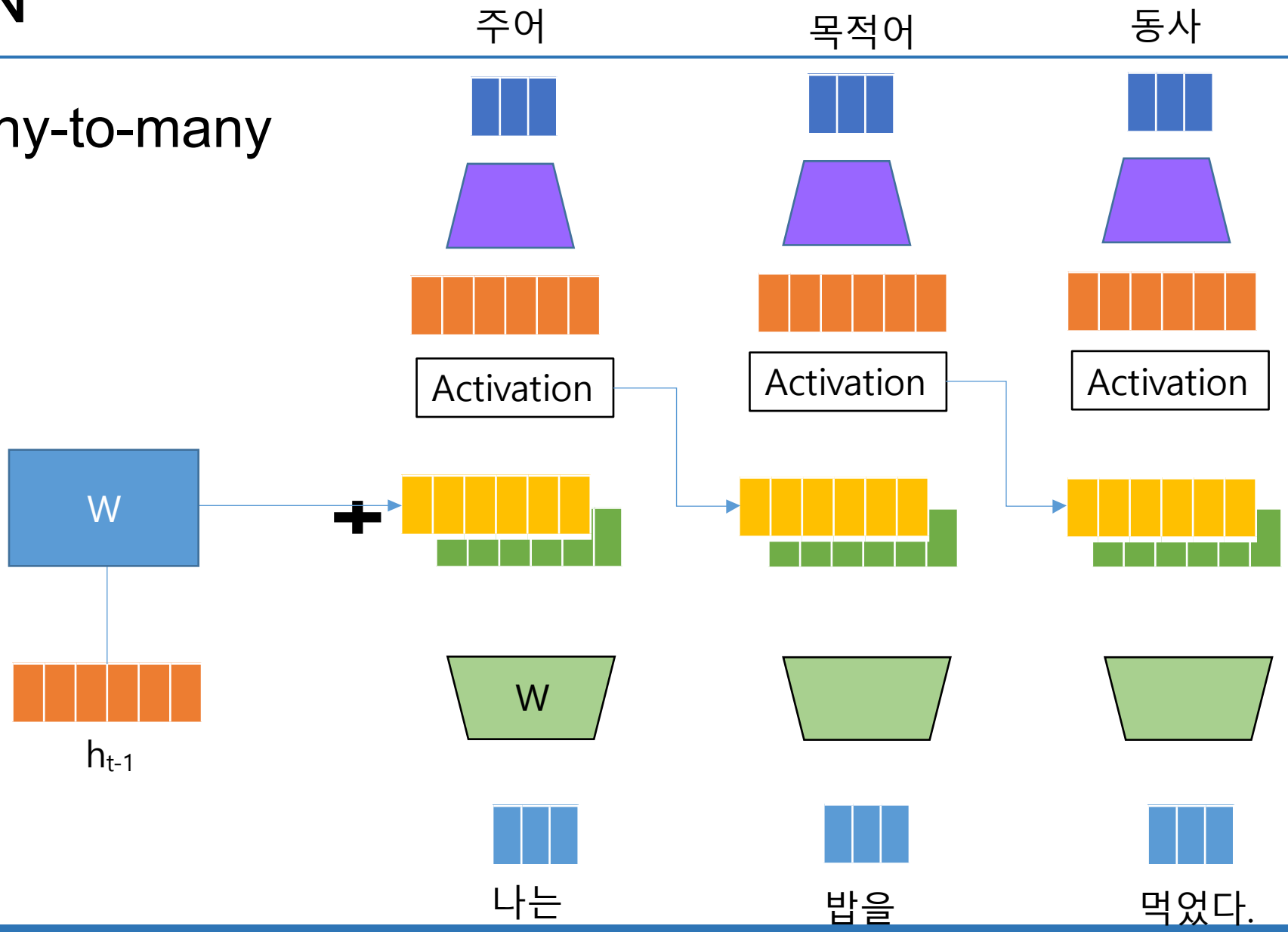
RNN

- Many-to-many



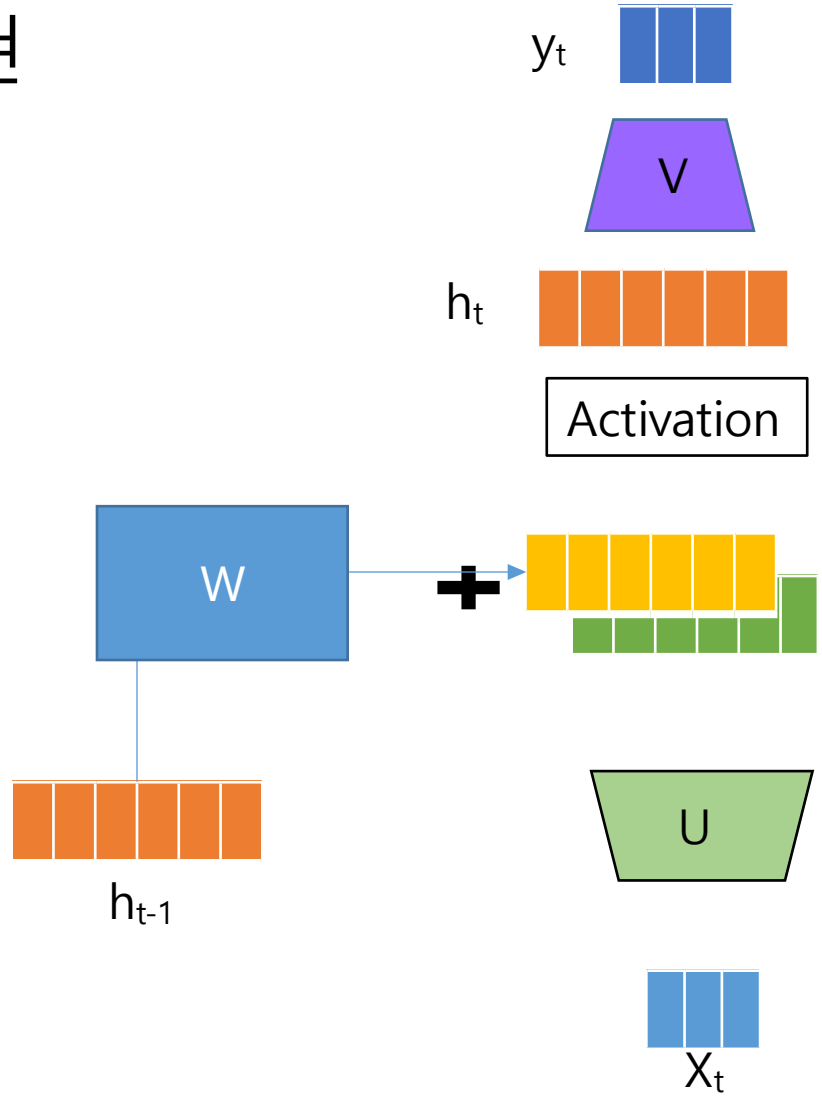
RNN

- Many-to-many



RNN

- 수학적 표현



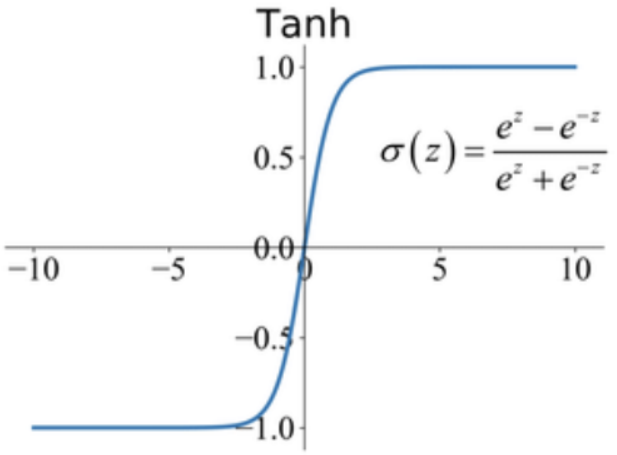
tanh

$h_t = f(Ux_t + Wh_{t-1})$

$y_t = f(Vh_t)$

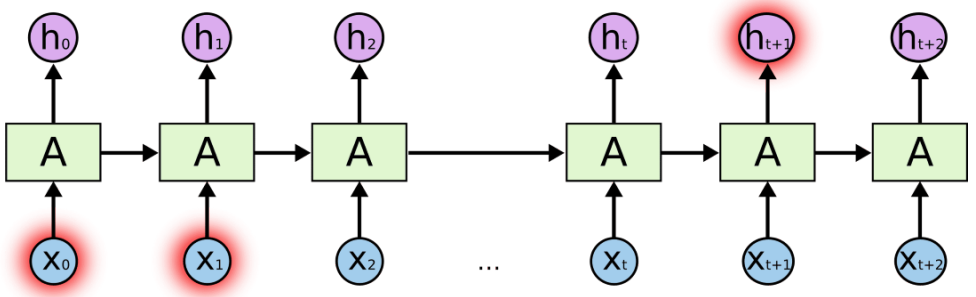
RNN 한계

- 장기 의존성 문제 (Long-Term Dependendy)
  - 기울기 소실(Gradient Vanishing)
    - RNN은 hidden layer의 활성함수로써 tanh를 사용
    - tanh의 출력값은 -1 ~ 1 사이의 값
    - 층을 거듭할수록 기울기가 0에 가까워짐
    - 즉, 과거의 정보가 먼 미래에 영향을 끼치지 못한다.



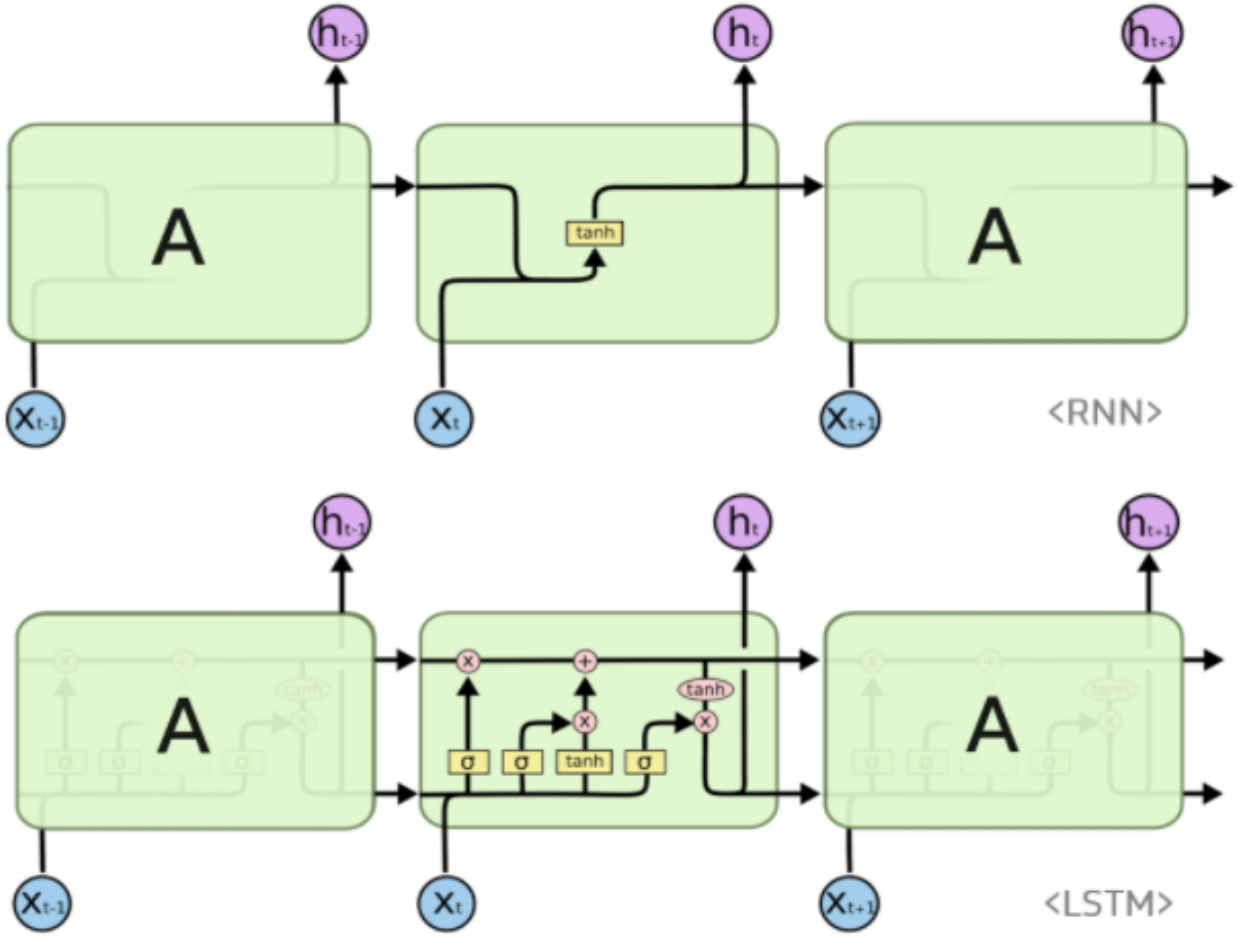
Ex) "부산에 여행을 왔는데 건물도 예쁘고 먹을 것도 맛있었어.
그런데 글쎄 직장 상사한테 전화가 왔어. 어디냐고 묻더라구 그래서 나는 말했지.
저 여행왔는데요. 여기 __"

=> LSTM을 사용하여 해결 가능!



LSTM

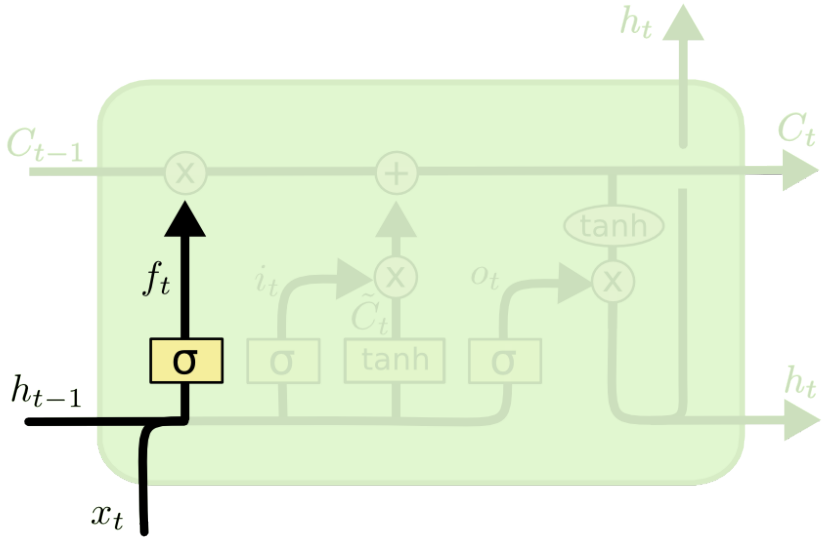
- LSTM 구조



LSTM

- Forget gate

  ex) 기존 주어의 성별정보



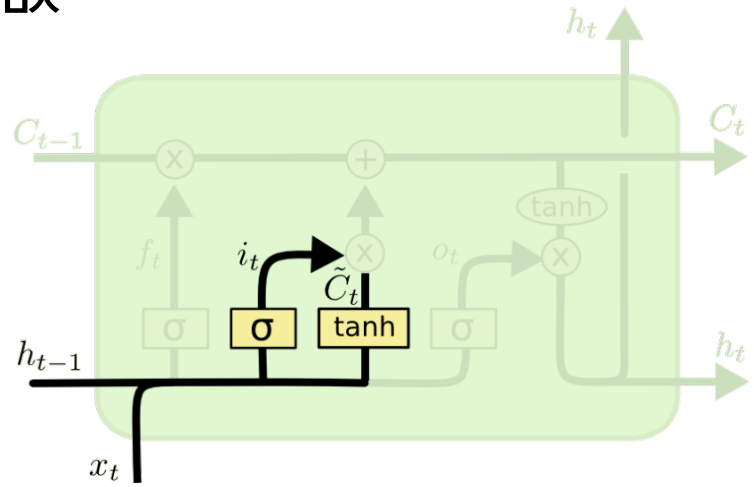
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

LSTM

- Input gate

  Ex) 새로운 주어의 성별정보

- Cell 후보값

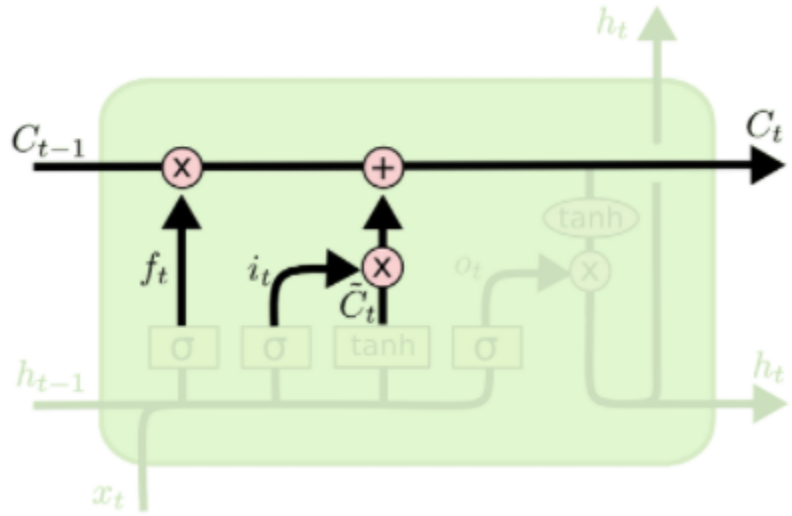


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

LSTM

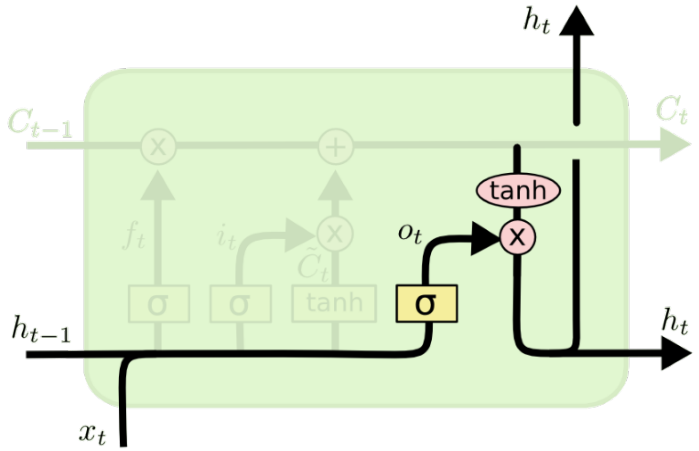
- Cell state update



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

LSTM

- Output gate



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Q & A