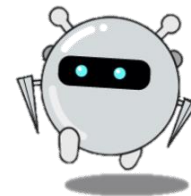


RNN

멘토 현시은



배경

- 배경

RNN

LSTM

GRU

배경지식 - 통계적 언어 모델

언어 모델 - 통계적 언어 모델 / 인공 신경망 언어 모델

- 배경

RNN

LSTM

GRU

배경지식 - 통계적 언어 모델

언어 모델 - 통계적 언어 모델 / 인공 신경망 언어 모델

SLM

$$P(x_1, x_2, \dots, x_n) = \prod_{n=1}^n P(x_n | x_1, x_2, \dots, x_{n-1})$$

$$P(\text{밥을} \mid \text{나는 배가 고파서}) = \frac{\text{count}(\text{나는 배가 고파서 밥을})}{\text{count}(\text{나는 배가 고파서})}$$

- 배경

RNN

LSTM

GRU

배경지식 - 통계적 언어 모델

언어 모델 - 통계적 언어 모델 / 인공 신경망 언어 모델

SLM

$$P(x_1, x_2, \dots, x_n) = \prod_{n=1}^n P(x_n | x_1, x_2, \dots, x_{n-1})$$

$$P(\text{밥을} \mid \text{나는 배가 고파서}) = \frac{\text{count}(\text{나는 배가 고파서 밥을})}{\text{count}(\text{나는 배가 고파서})}$$

= 0번
= 0번
...?

- 배경

RNN

LSTM

GRU

배경지식 - 통계적 언어 모델

언어 모델 - 통계적 언어 모델 / 인공 신경망 언어 모델

SLM

$$P(x_1, x_2, \dots, x_n) = \prod_{n=1}^n P(x_n | x_1, x_2, \dots, x_{n-1})$$

$$P(\text{밥을} \mid \text{나는 배가 고파서}) = \frac{\text{count}(\text{나는 배가 고파서 밥을})}{\text{count}(\text{나는 배가 고파서})}$$

= 0번
= 0번
...?

N - gram

전체 문장을 참고하는게 아닌 N개의 단어만 참고한다

Unigram

Bigram

Trigram 등등..

- 배경

RNN

LSTM

GRU

배경지식 - 통계적 언어 모델

언어 모델 - 통계적 언어 모델 / 인공 신경망 언어 모델

SLM

$$P(x_1, x_2, \dots, x_n) = \prod_{n=1}^n P(x_n | x_1, x_2, \dots, x_{n-1})$$

$$P(\text{밥을} \mid \text{나는 배가 고파서}) = \frac{\text{count}(\text{나는 배가 고파서 밥을})}{\text{count}(\text{나는 배가 고파서})}$$

= 0번
= 0번
...?

N - gram

전체 문장을 참고하는게 아닌 N개의 단어만 참고한다

Unigram

Bigram

Trigram 등등..

희소 문제 발생!!

- 배경

RNN

LSTM

GRU

배경지식 - 통계적 언어 모델

언어 모델 - 통계적 언어 모델 / 인공 신경망 언어 모델

SLM

$$P(x_1, x_2, \dots, x_n) = \prod_{n=1}^n P(x_n | x_1, x_2, \dots, x_{n-1})$$

$$P(\text{밥을} \mid \text{나는 배가 고파서}) = \frac{\text{count}(\text{나는 배가 고파서 밥을})}{\text{count}(\text{나는 배가 고파서})}$$

= 0번
= 0번
...?

N - gram

전체 문장을 참고하는게 아닌 N개의 단어만 참고한다

Unigram

Bigram

Trigram 등등..

희소 문제 발생!! → 인공 신경망 언어 모델

- 배경

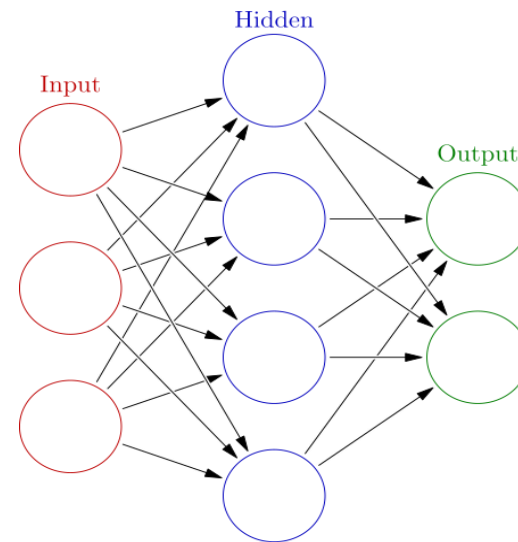
RNN

LSTM

GRU

배경지식 - 인공 신경망

- ANN (Artificial Neural Network)
(NN)



- 배경

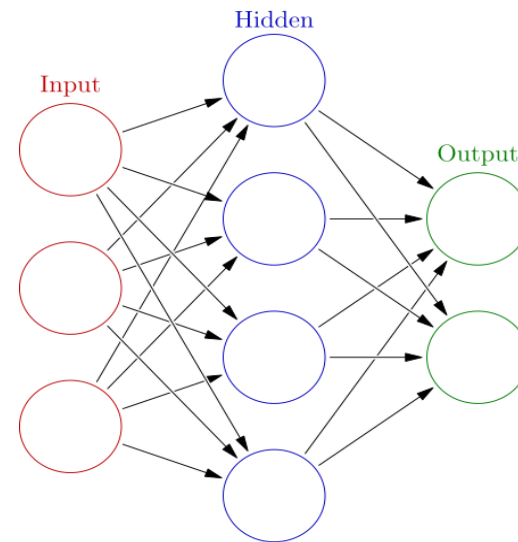
RNN

LSTM

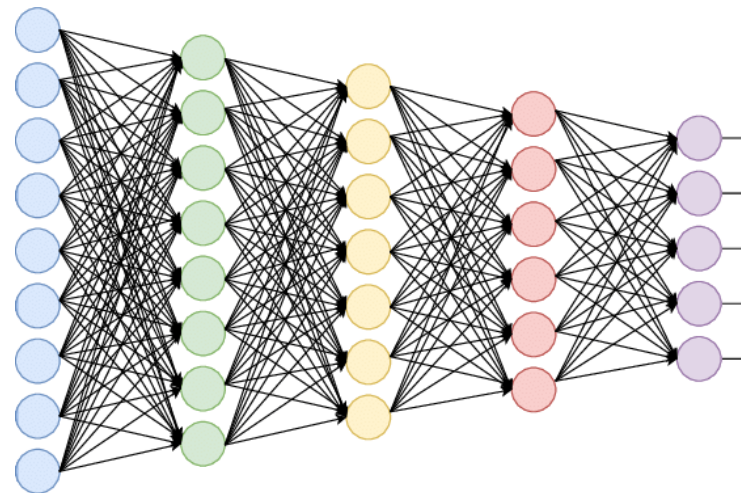
GRU

배경지식 - 인공 신경망

- ANN (Artificial Neural Network)
(NN)



- DNN (Deep Neural Network)



- 배경

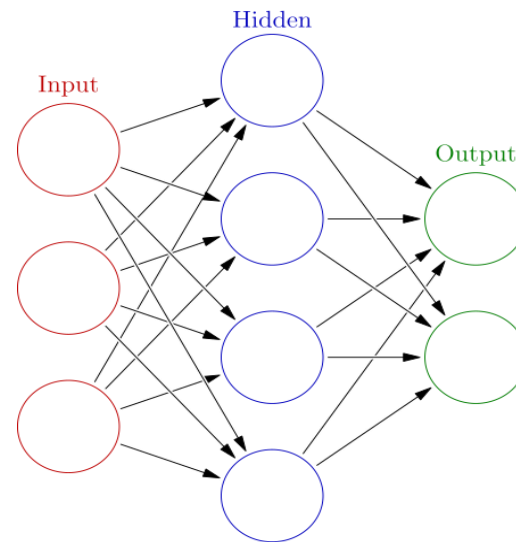
RNN

LSTM

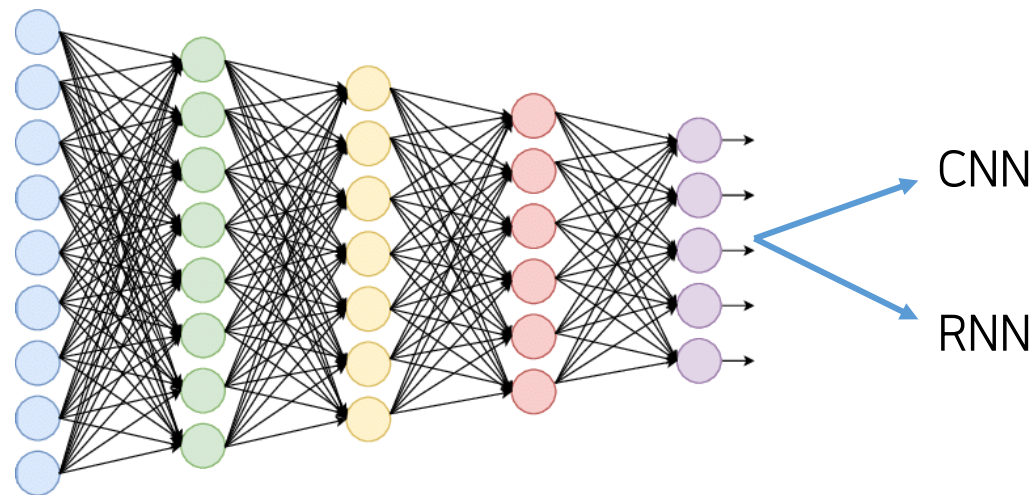
GRU

배경지식 - 인공 신경망

- ANN (Artificial Neural Network)
(NN)



- DNN (Deep Neural Network)





RNN

배경

- RNN

LSTM

GRU

RNN (Recurrent Neural Network)

RNN - '순환 신경망' 이름 그대로 내부에서 순환되는 구조를 갖는 특징이 있다

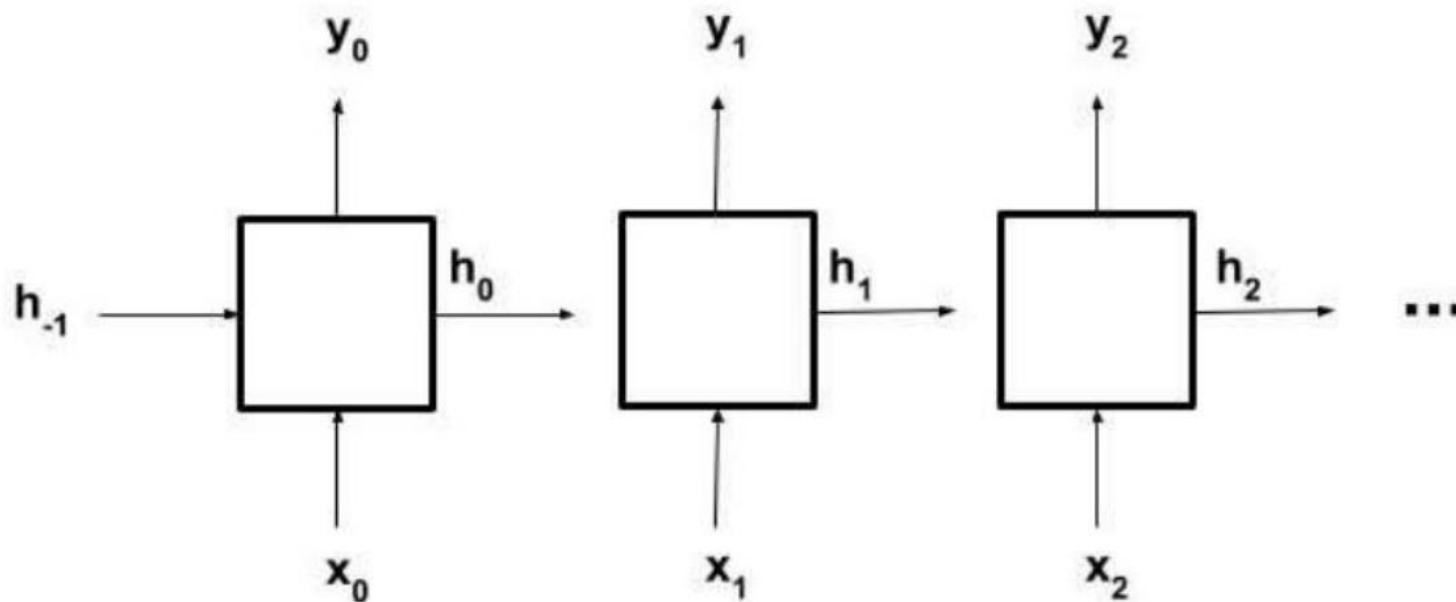
배경

• RNN

LSTM

GRU

Vanilla RNN (가장 기본적인 RNN 구조)



- x_t : 입력 값 / y_t : 출력 값 / h_t : Hidden State
- 은닉층의 계산 값이 출력층(y)과 은닉층의 다음 계산 과정(h)에 전달
- 각 층을 하나의 step이라고 함

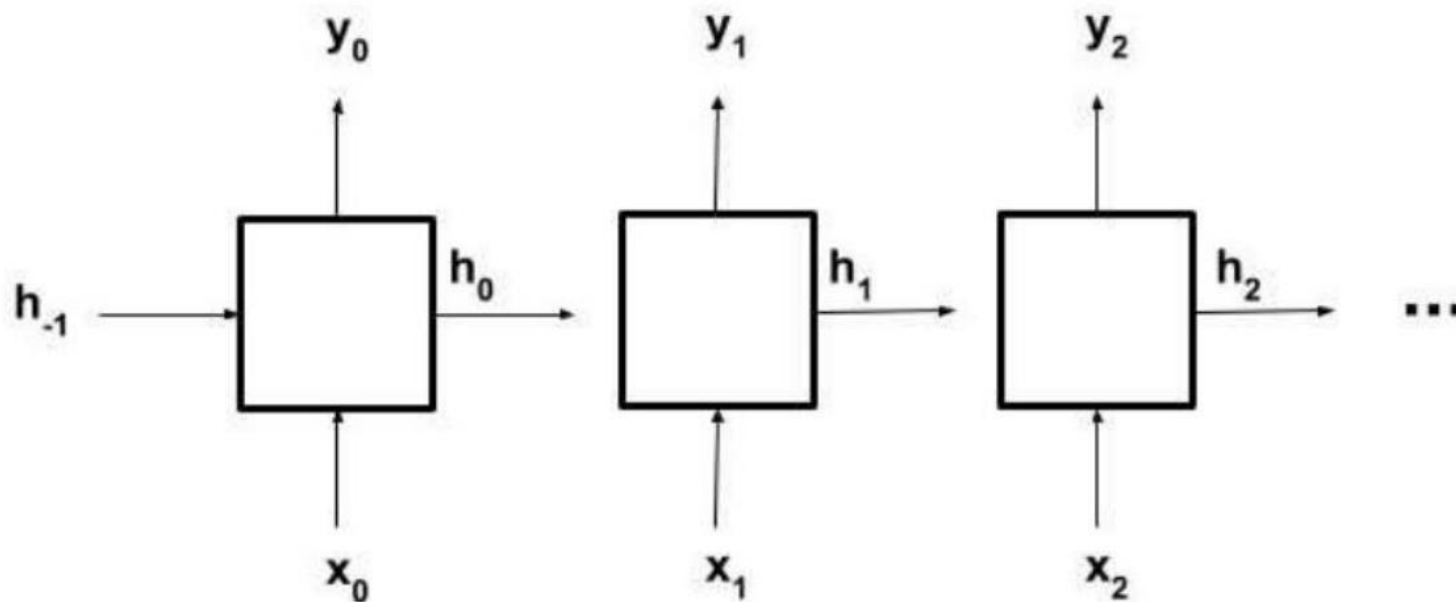
배경

• RNN

LSTM

GRU

Vanilla RNN (가장 기본적인 RNN 구조)



- x_t : 입력 값 / y_t : 출력 값 / h_t : Hidden State
- 은닉층의 계산 값이 출력층(y)과 은닉층의 다음 계산 과정(h)에 전달
- 각 층을 하나의 step이라고 함

배경

- RNN

LSTM

GRU

RNN 층별 연산 수식

h_t : Step t에서의 Hidden State

x_t : Sequence의 t+1 번째 원소

f_h / f_y : 은닉층과 출력층 활성화 함수

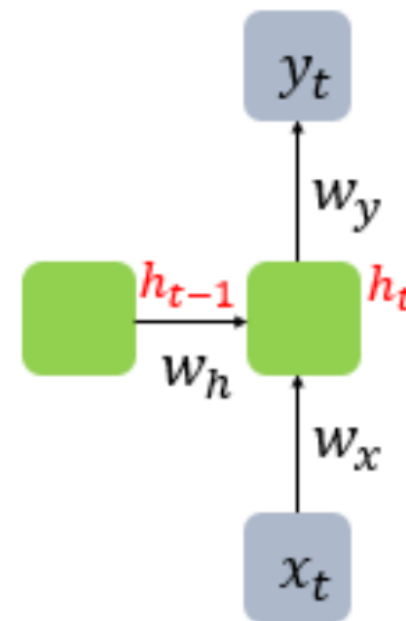
y_t : t+1 번째 원소까지의 출력 값

W, b : 각 층의 가중치와 편향, 모델이 학습하게 되는 값

y_{T-1} : 최종 output (T는 입력 데이터의 길이)

$$h_t = f_h(x_t W_{xh} + h_{t-1} W_{hh} + b_h)$$

$$y_t = f_y(h_t W_{hy} + b_y)$$



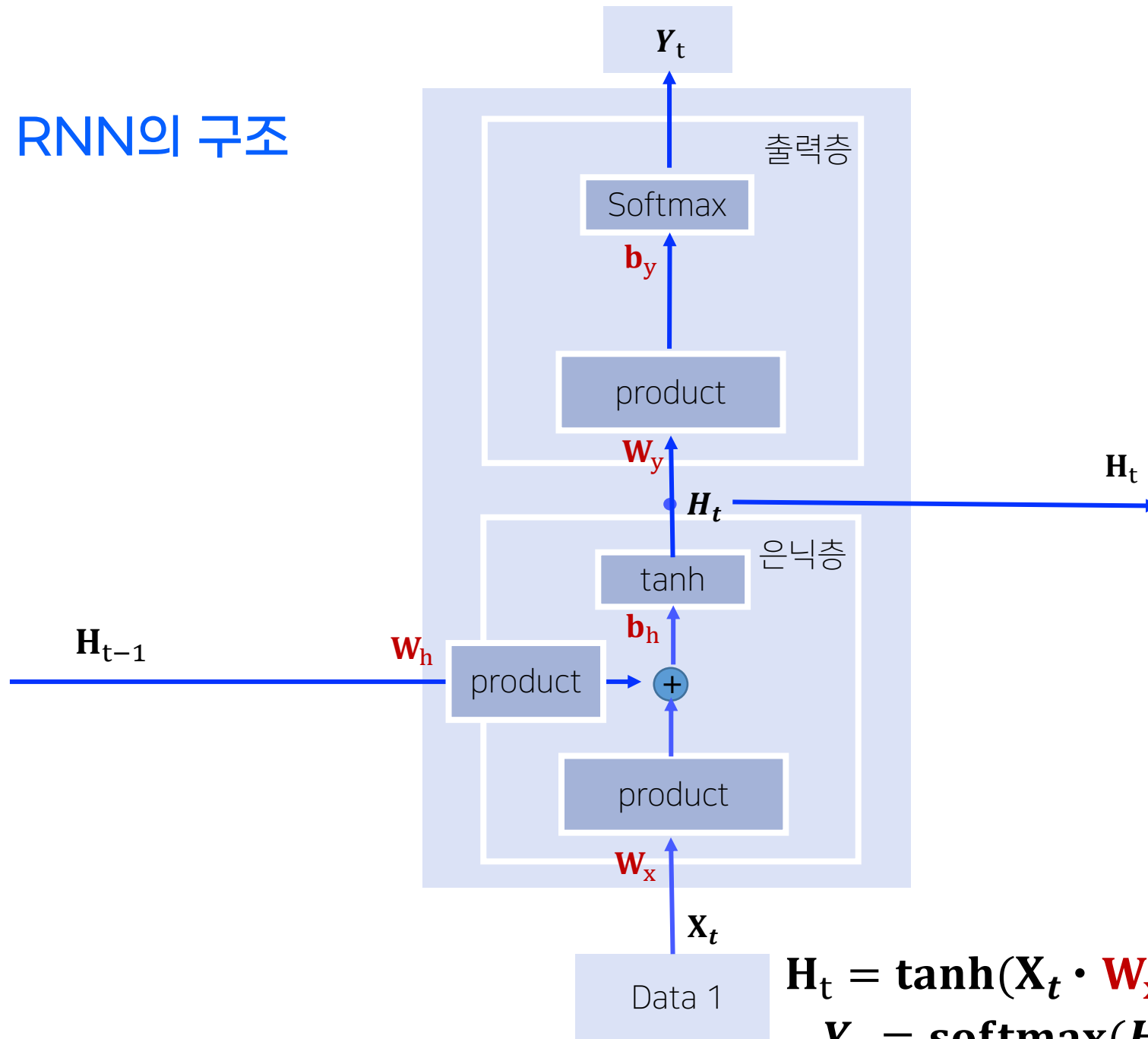
배경

- RNN

LSTM

GRU

RNN의 구조



배경

• RNN

LSTM

GRU

RNN 층별 연산 수식

* 수식을 구성하는 요소들의 shape

$$x_t : (d \times 1)$$

$$W_x : (D_h \times d)$$

$$W_h : (D_h \times D_h)$$

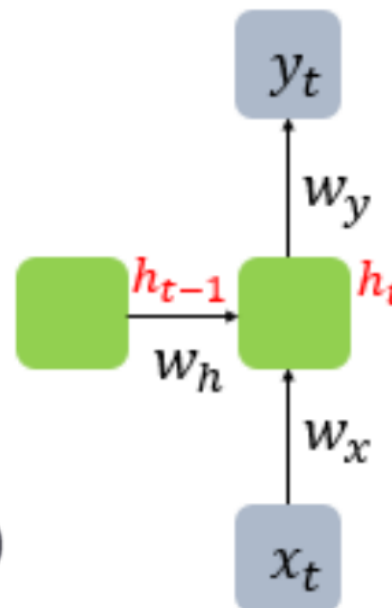
$$h_{t-1} : (D_h \times 1)$$

$$b : (D_h \times 1)$$

$$W_y : (D_h \times D_y)$$

$$h_t = f_h(x_t W_x + h_{t-1} W_h + b_h)$$

$$y_t = f_y(h_t W_y + b_y)$$



- 입력 값 x_t 는 단어 벡터이고 이 단어 벡터의 차원이 d
- 은닉 상태(hidden state)의 크기는 D_h

배경

• RNN

LSTM

GRU

RNN 층별 연산 수식

단어 벡터의 차원과 hidden state의 크기를 모두 4라고 했을 때,
은닉층 연산은 다음과 같음

$$\tanh \left(\begin{matrix} W_h \\ D_h \times D_h \end{matrix} \times \begin{matrix} h_{t-1} \\ D_h \times 1 \end{matrix} + \begin{matrix} W_x \\ D_h \times d \end{matrix} \times \begin{matrix} x_t \\ d \times 1 \end{matrix} + \begin{matrix} b \\ D_h \times 1 \end{matrix} \right) = \begin{matrix} h_t \\ D_h \times 1 \end{matrix}$$

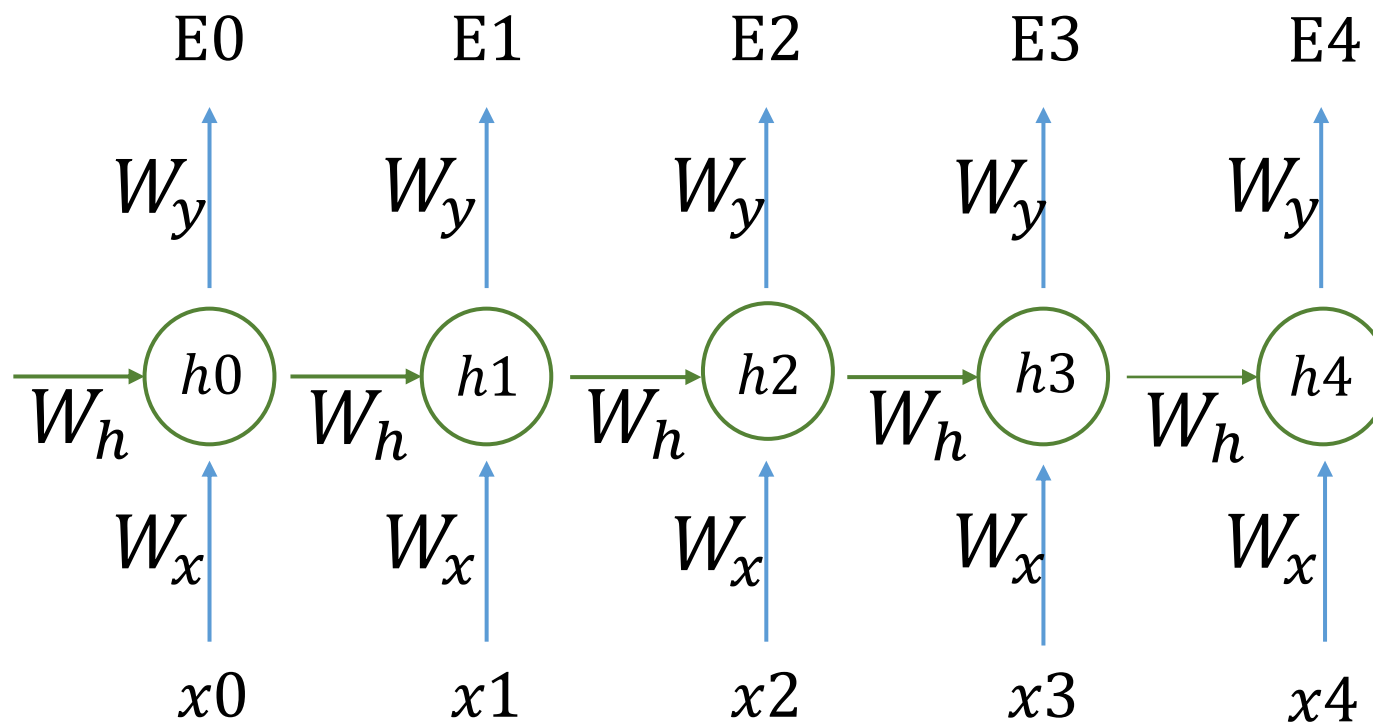
배경

- RNN

LSTM

GRU

RNN - BPTT (BackPropagation Through Time)



$$E = y - \hat{y}$$

$$h_t = \tanh(W_x x_t + W_h h_{t-1})$$
$$\hat{y}_t = \text{softmax}(W_y h_t)$$

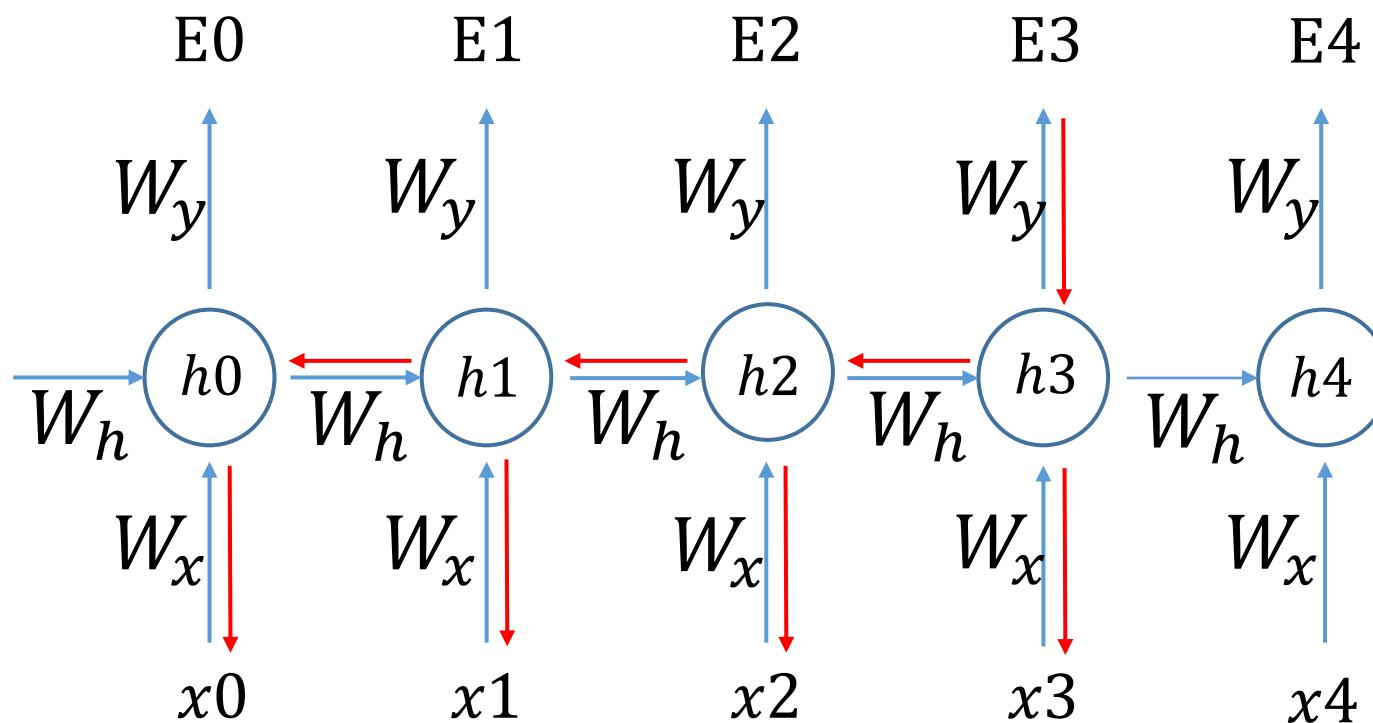
배경

- RNN

LSTM

GRU

RNN - BPTT (BackPropagation Through Time)



$$E = y - \hat{y}$$

$$h_t = \tanh(W_x x_t + W_h h_{t-1})$$
$$\hat{y}_t = \text{softmax}(W_y h_t)$$

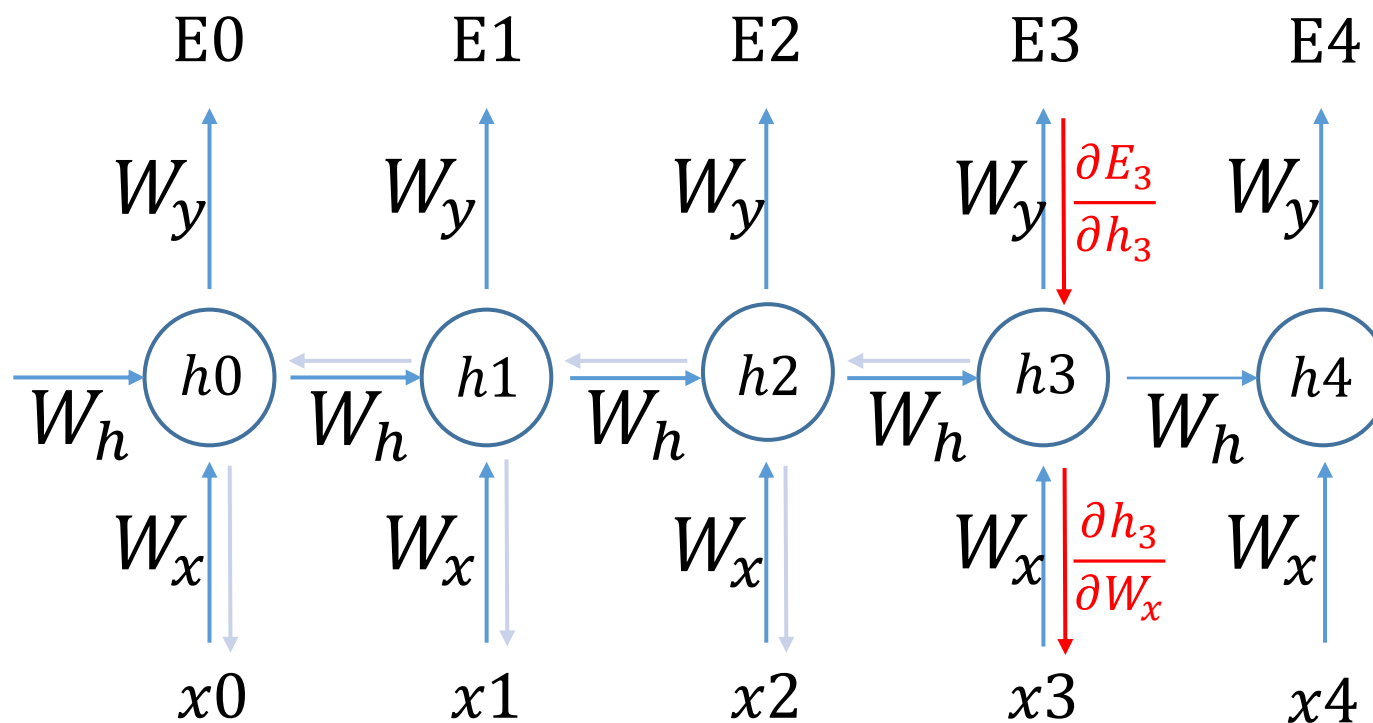
배경

- RNN

LSTM

GRU

RNN - BPTT (BackPropagation Through Time)



$$\frac{\partial E_3}{\partial W_x} = \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial h_3} \frac{\partial h_3}{\partial W_x}$$

$$E = y - \hat{y}$$

$$h_t = \tanh(W_x x_t + W_h h_{t-1})$$
$$\hat{y}_t = \text{softmax}(W_y h_t)$$

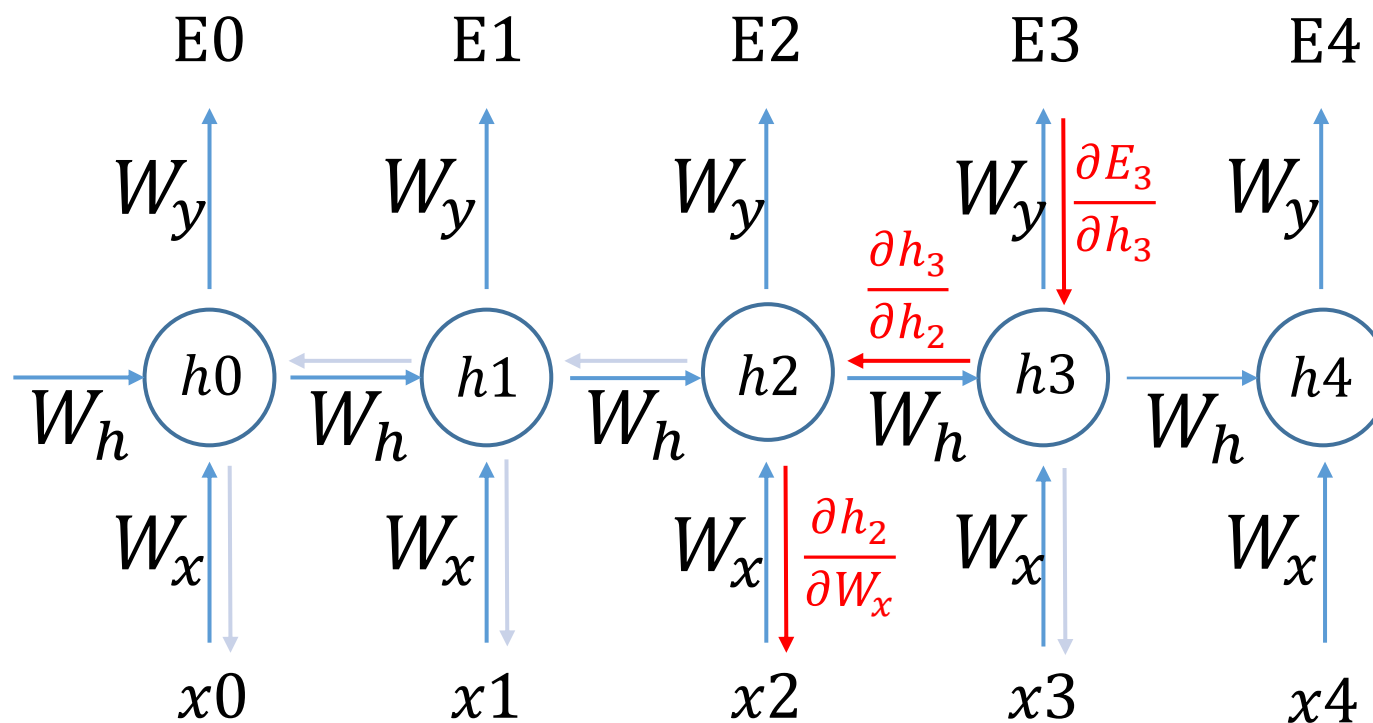
배경

- RNN

LSTM

GRU

RNN - BPTT (BackPropagation Through Time)



$$\frac{\partial E_3}{\partial W_x} = \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial h_3} \frac{\partial h_3}{\partial W_x} + \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial h_3} \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial W_x}$$

$$E = y - \hat{y}$$

$$h_t = \tanh(W_x x_t + W_h h_{t-1})$$
$$\hat{y}_t = \text{softmax}(W_y h_t)$$

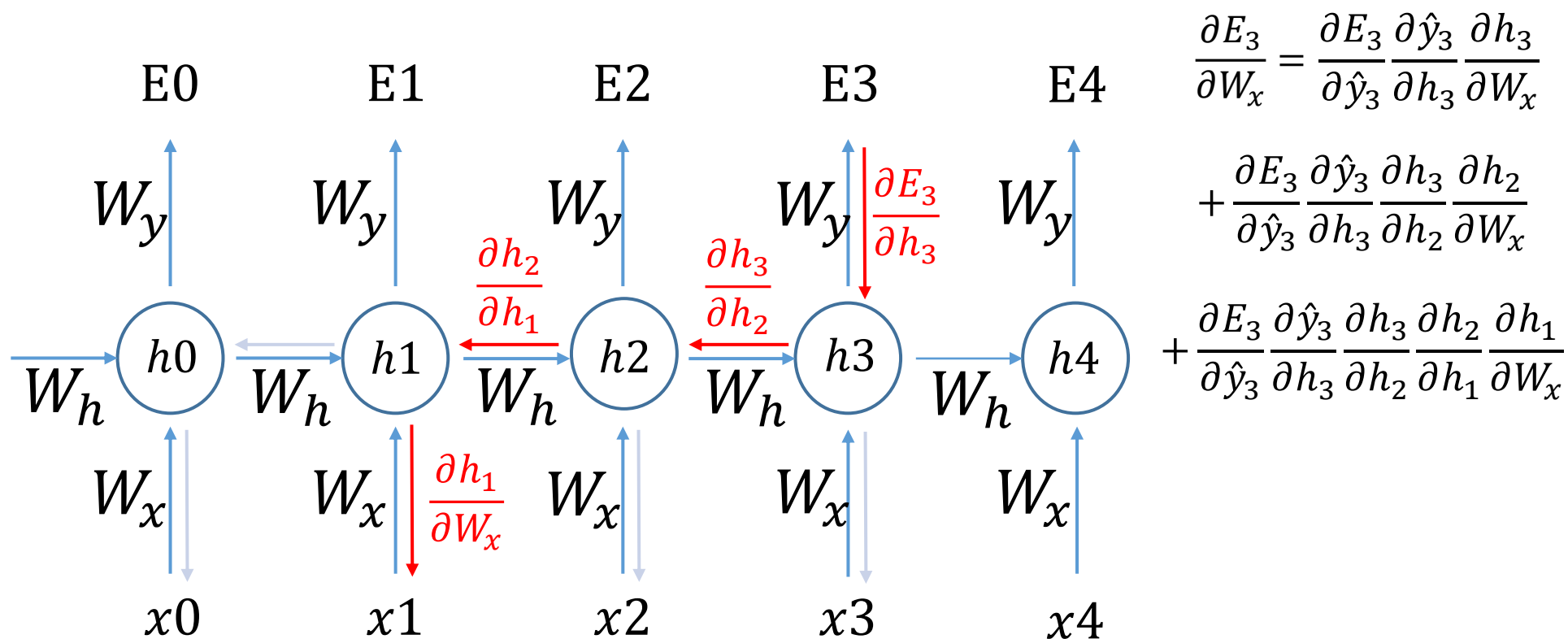
배경

- RNN

LSTM

GRU

RNN - BPTT (BackPropagation Through Time)



$$E = y - \hat{y}$$

$$\begin{aligned} h_t &= \tanh(W_x x_t + W_h h_{t-1}) \\ \hat{y}_t &= \text{softmax}(W_y h_t) \end{aligned}$$

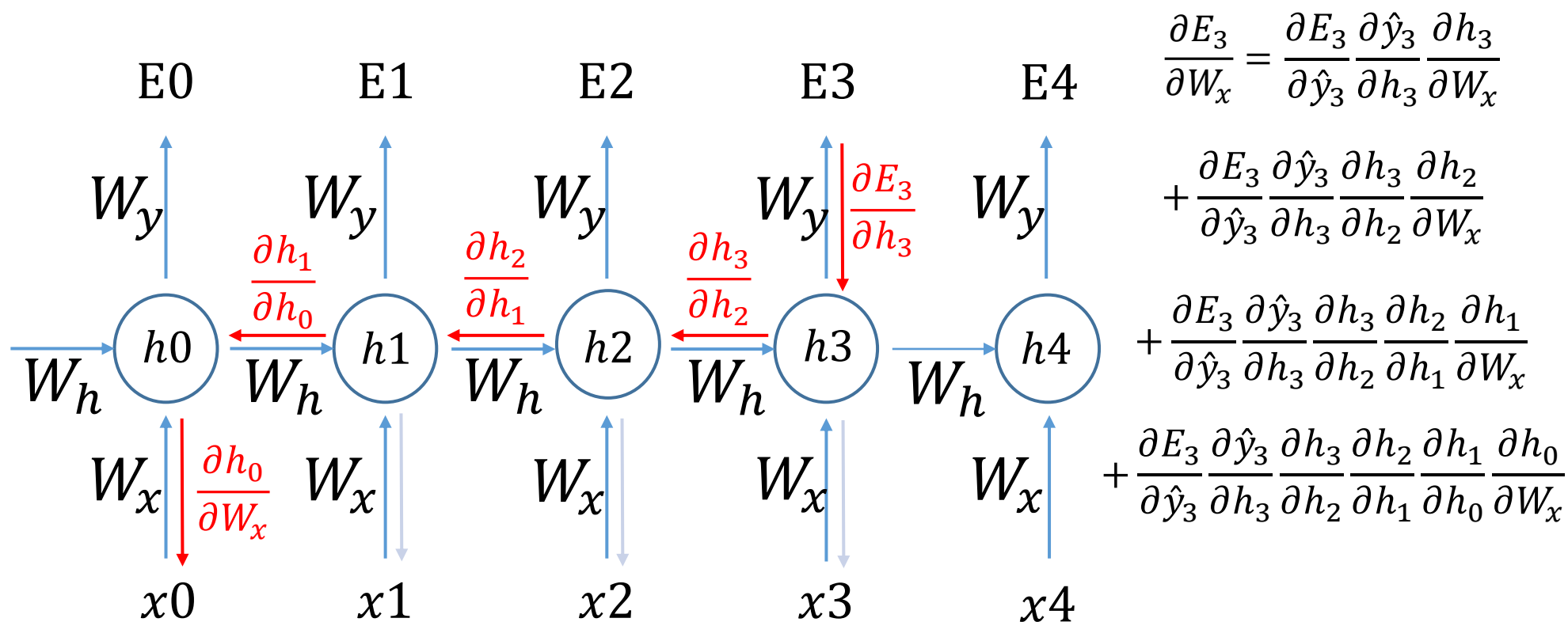
배경

- RNN

LSTM

GRU

RNN - BPTT (BackPropagation Through Time)



$$E = y - \hat{y}$$

$$h_t = \tanh(W_x x_t + W_h h_{t-1})$$
$$\hat{y}_t = \text{softmax}(W_y h_t)$$

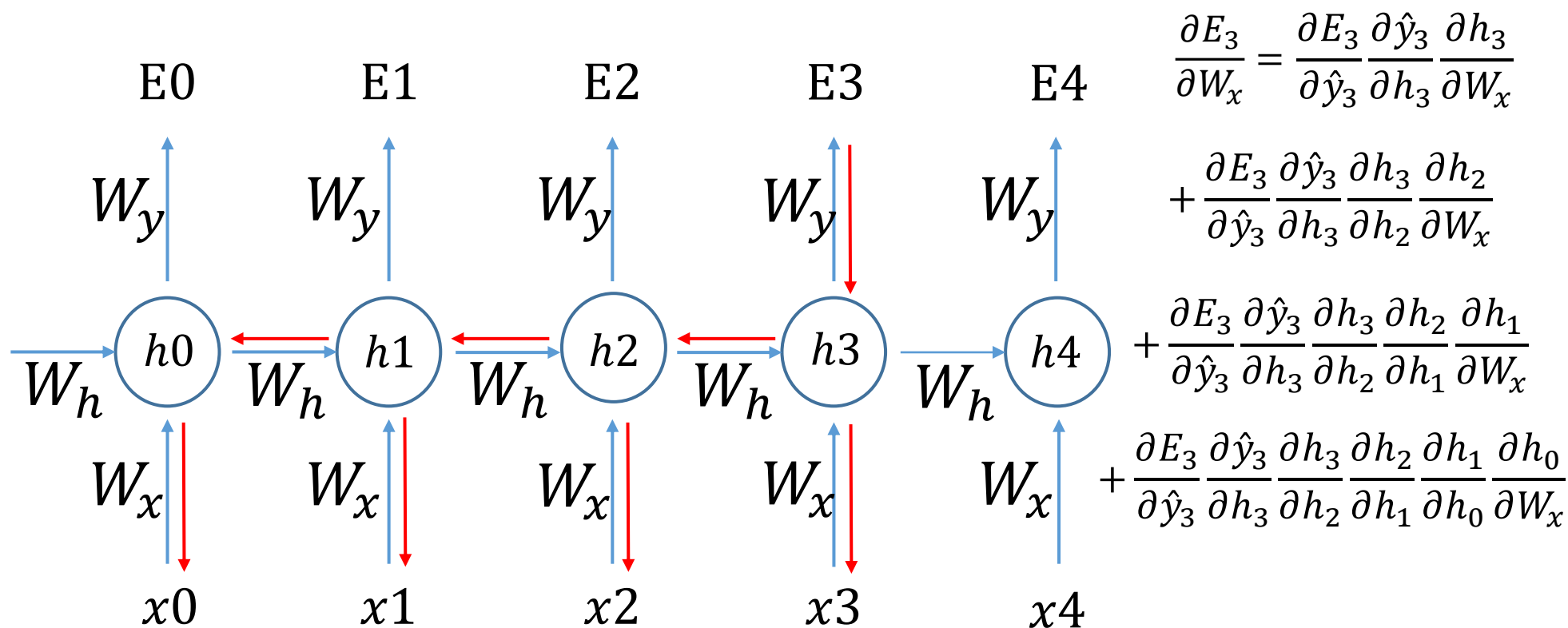
배경

• RNN

LSTM

GRU

RNN - BPTT (BackPropagation Through Time)



$$W = W - \text{learning rate} * \frac{\partial E}{\partial W}$$

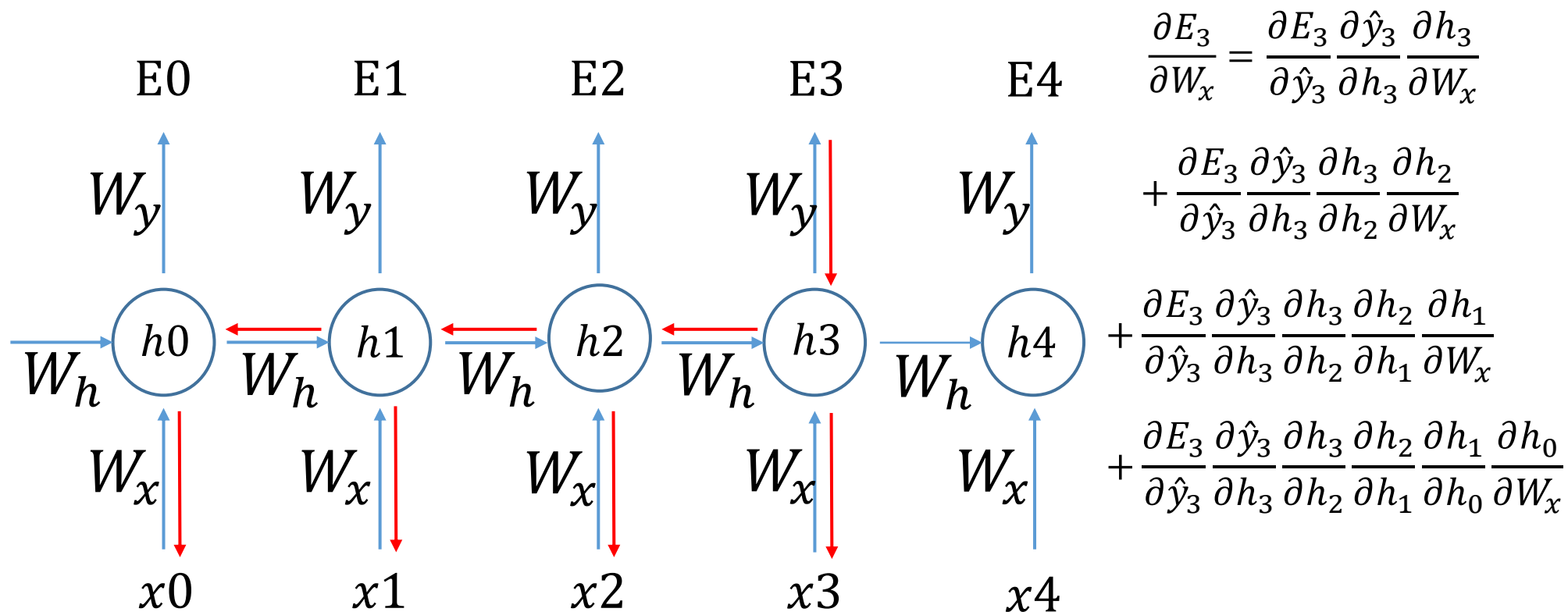
배경

- RNN

LSTM

GRU

RNN - 한계



$$W = W - \text{learning rate} * \frac{\partial E}{\partial W}$$

만약 문장의 길이가 엄청 길어지면..?

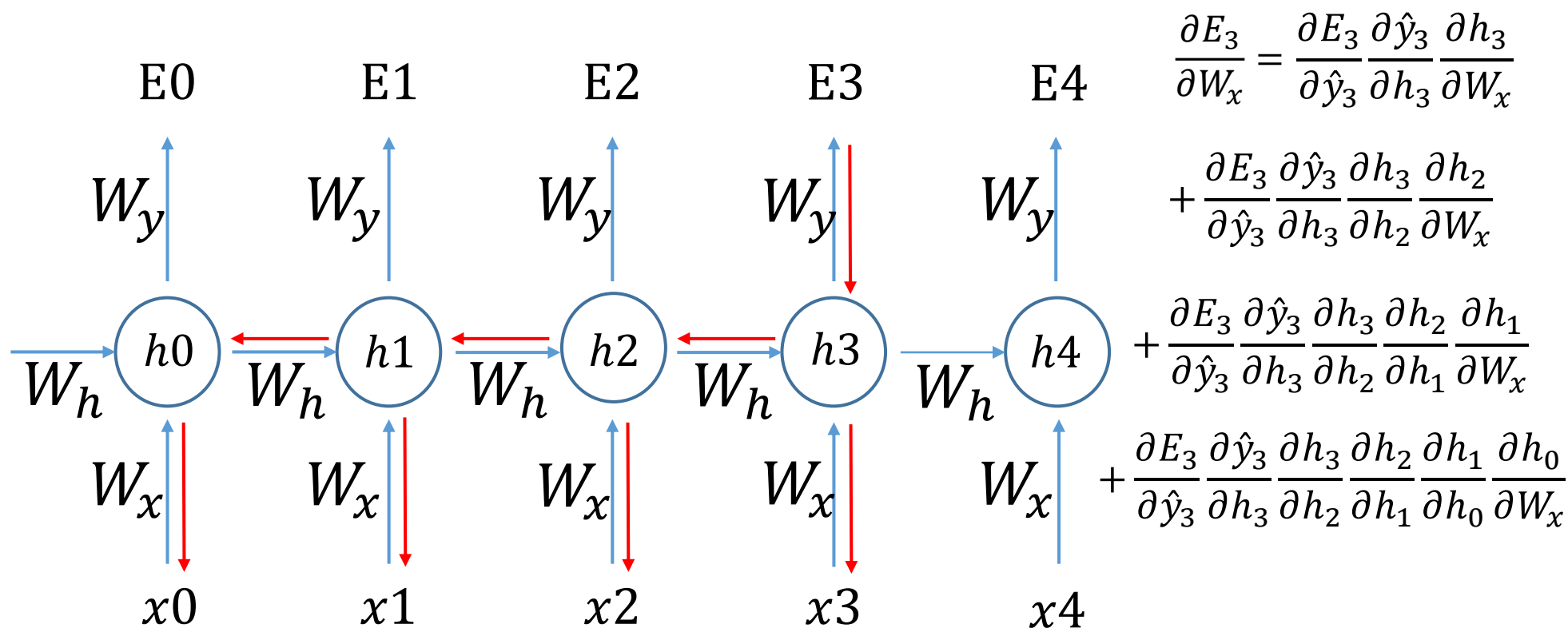
배경

• RNN

LSTM

GRU

RNN - 한계



미분 값이 작을 경우
Gradient Vanishing

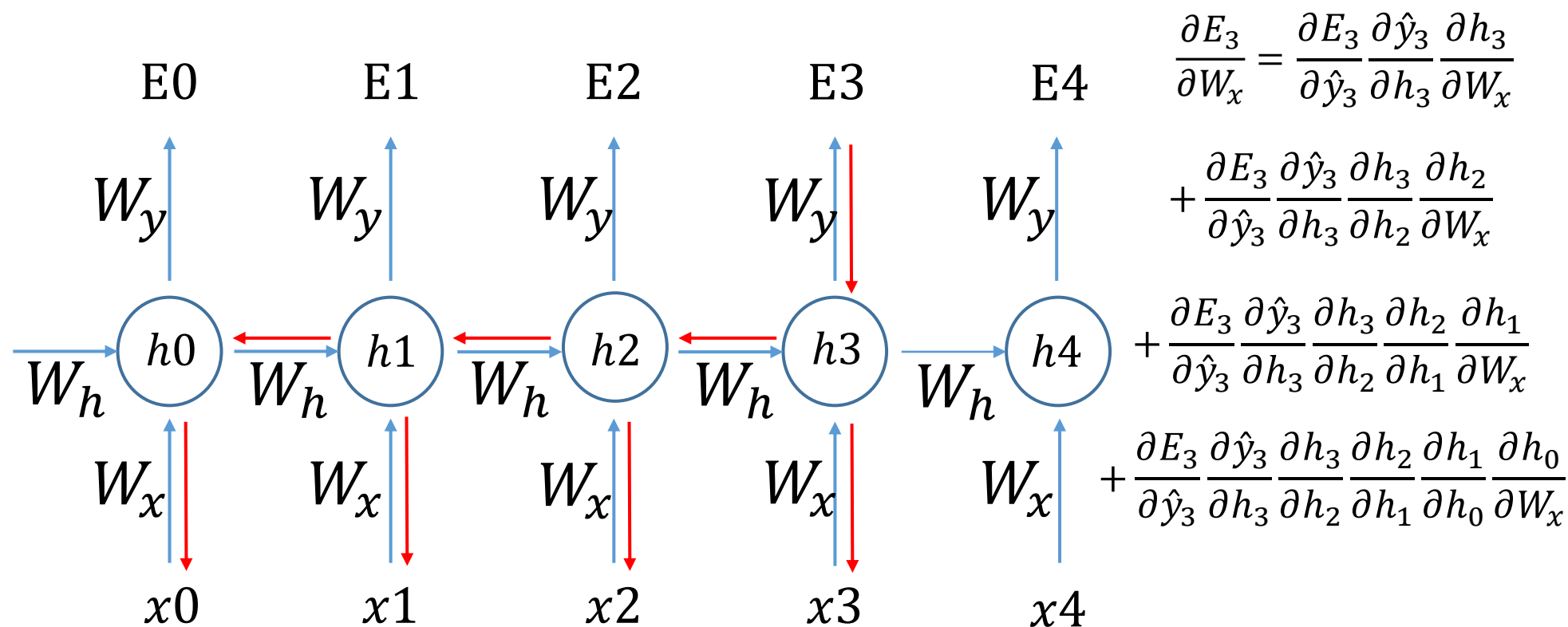
배경

• RNN

LSTM

GRU

RNN - 한계



미분 값이 클 경우
Gradient Exploding

→ Gradient clipping

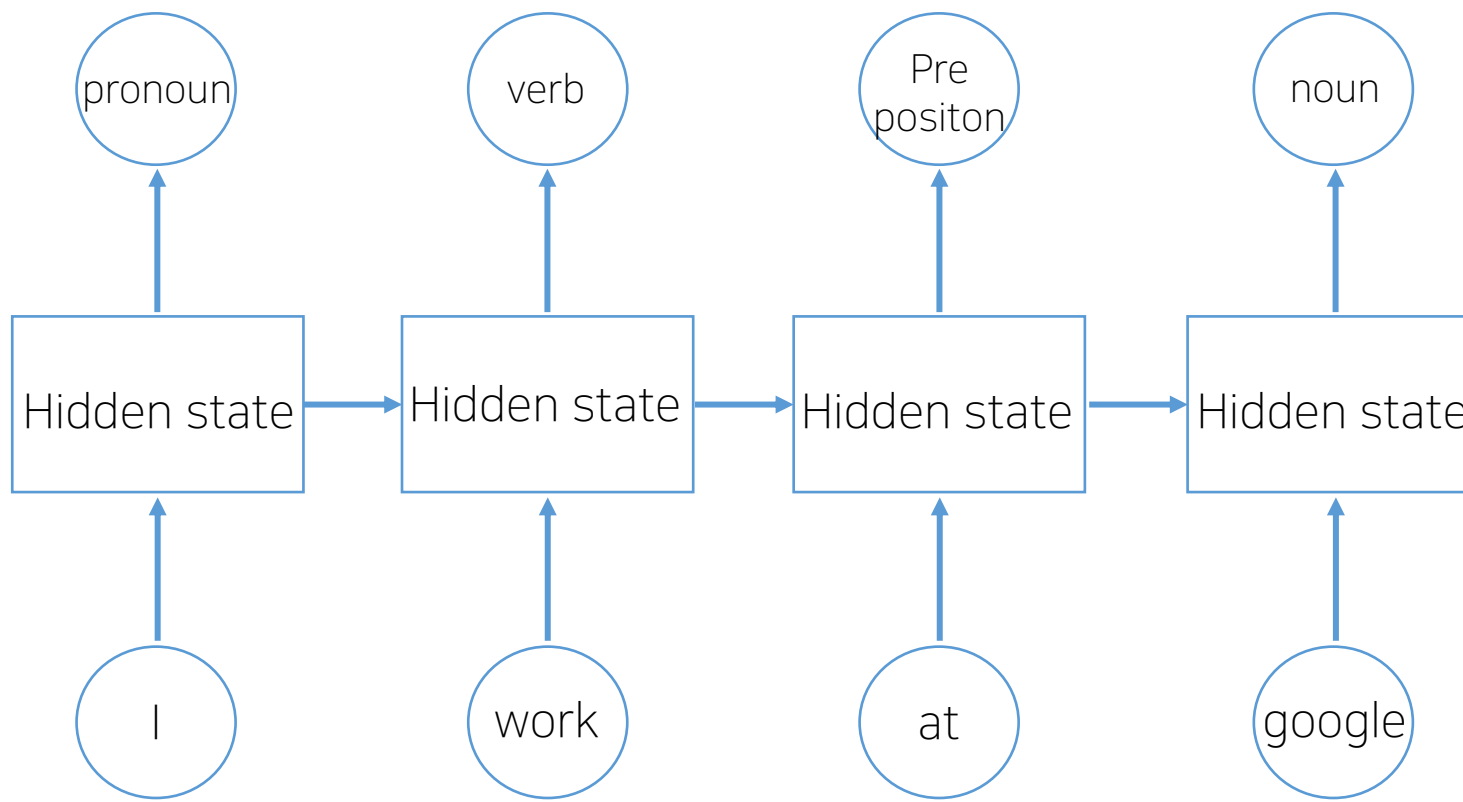
배경

- RNN

LSTM

GRU

RNN



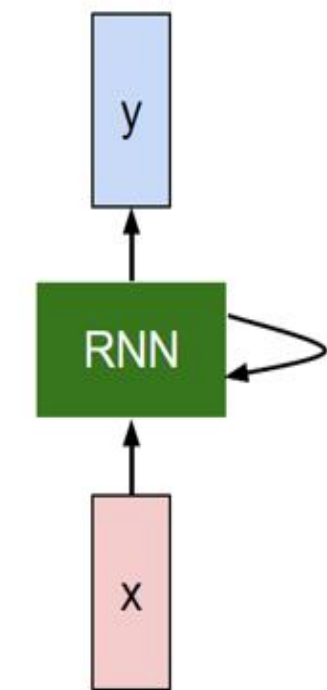
배경

- RNN

LSTM

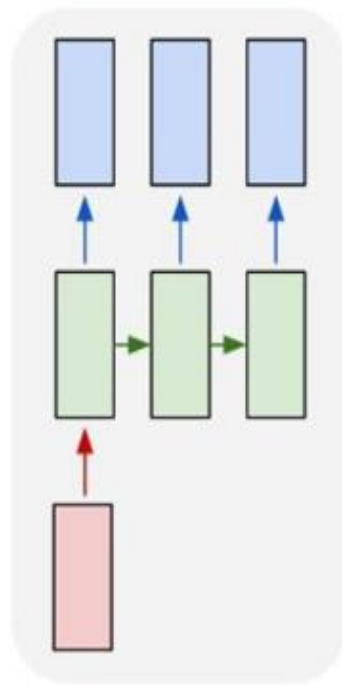
GRU

RNN



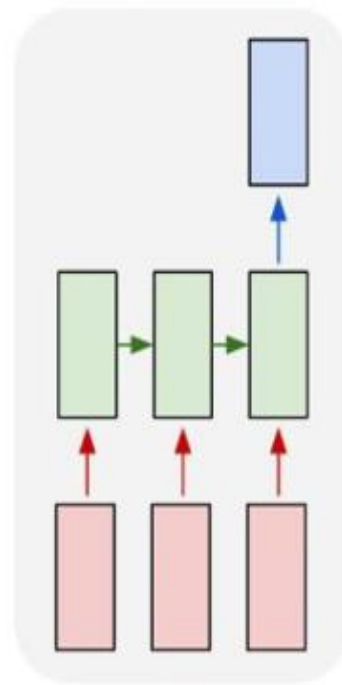
<기본구조>

one to many



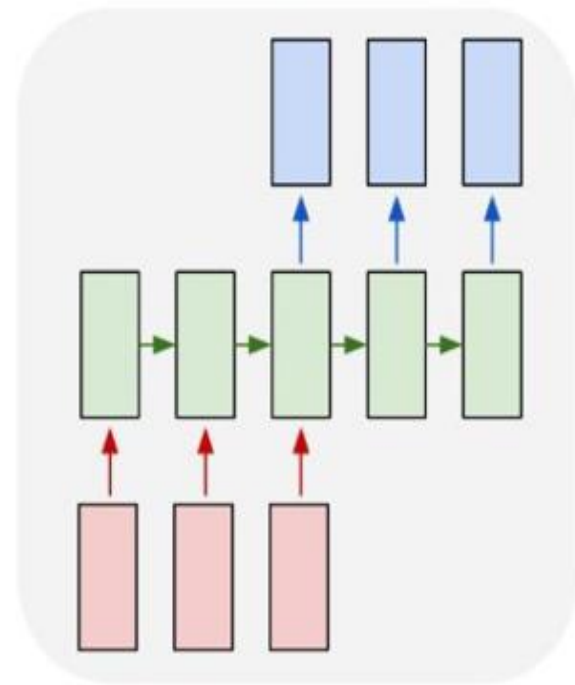
<사진설명 붙이기>
사진 → 단어들

many to one



<감성분석>
단어들 → 감성점수

many to many



<번역>
단어들 → 단어들



LSTM

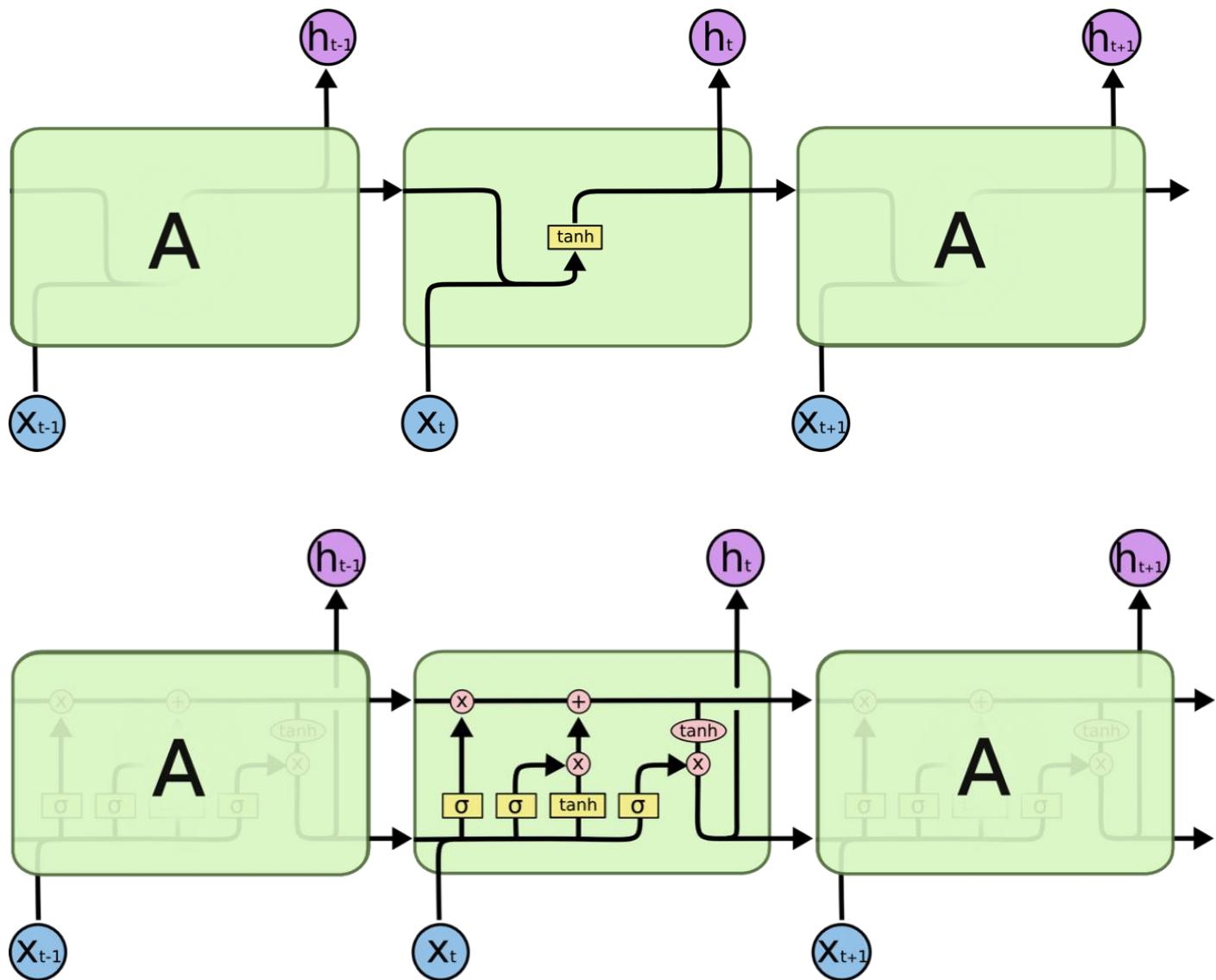
배경

RNN

• LSTM

GRU

LSTM



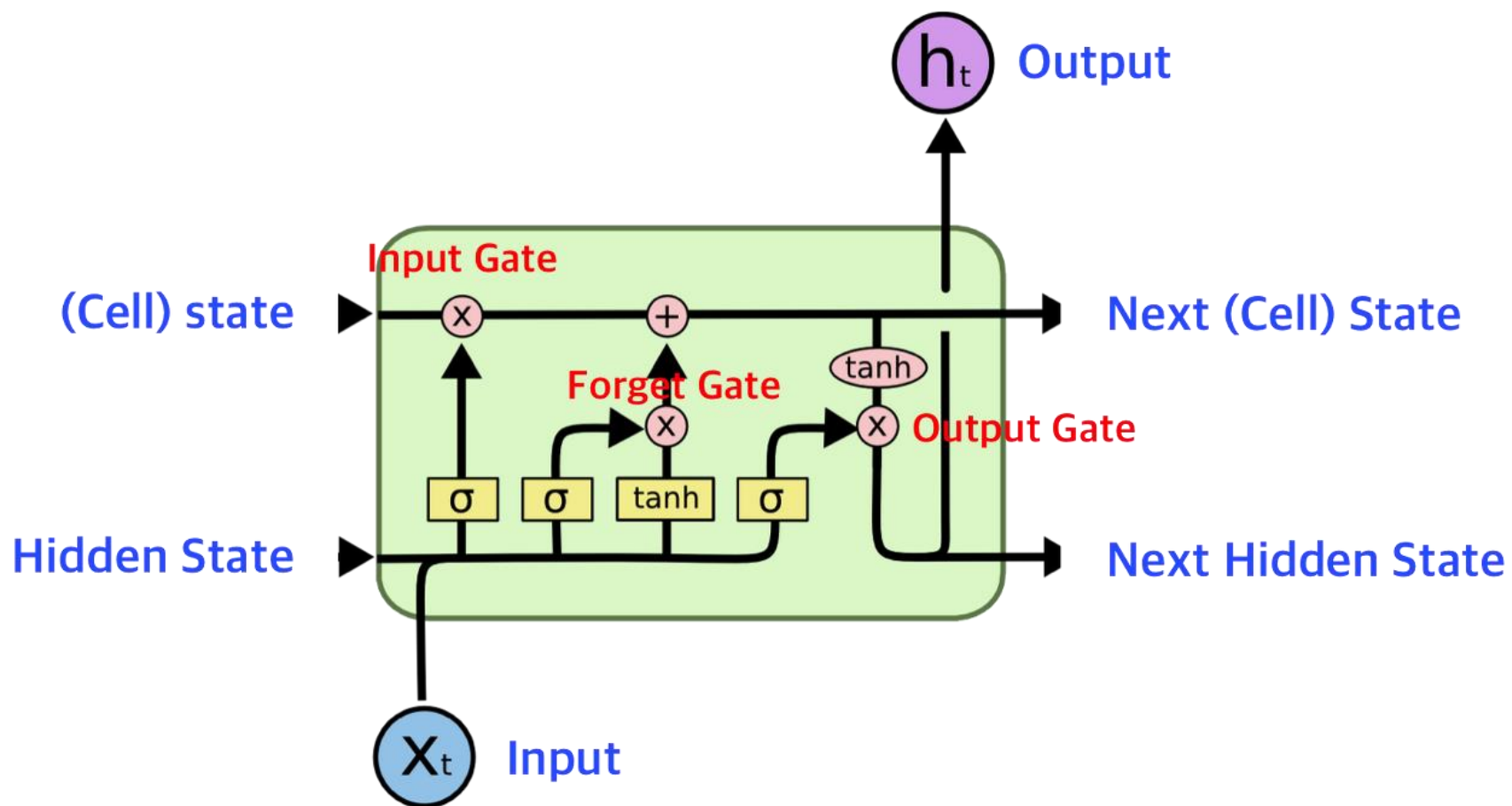
배경

RNN

• LSTM

GRU

LSTM - 모델 구조



<https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>



GRU

배경

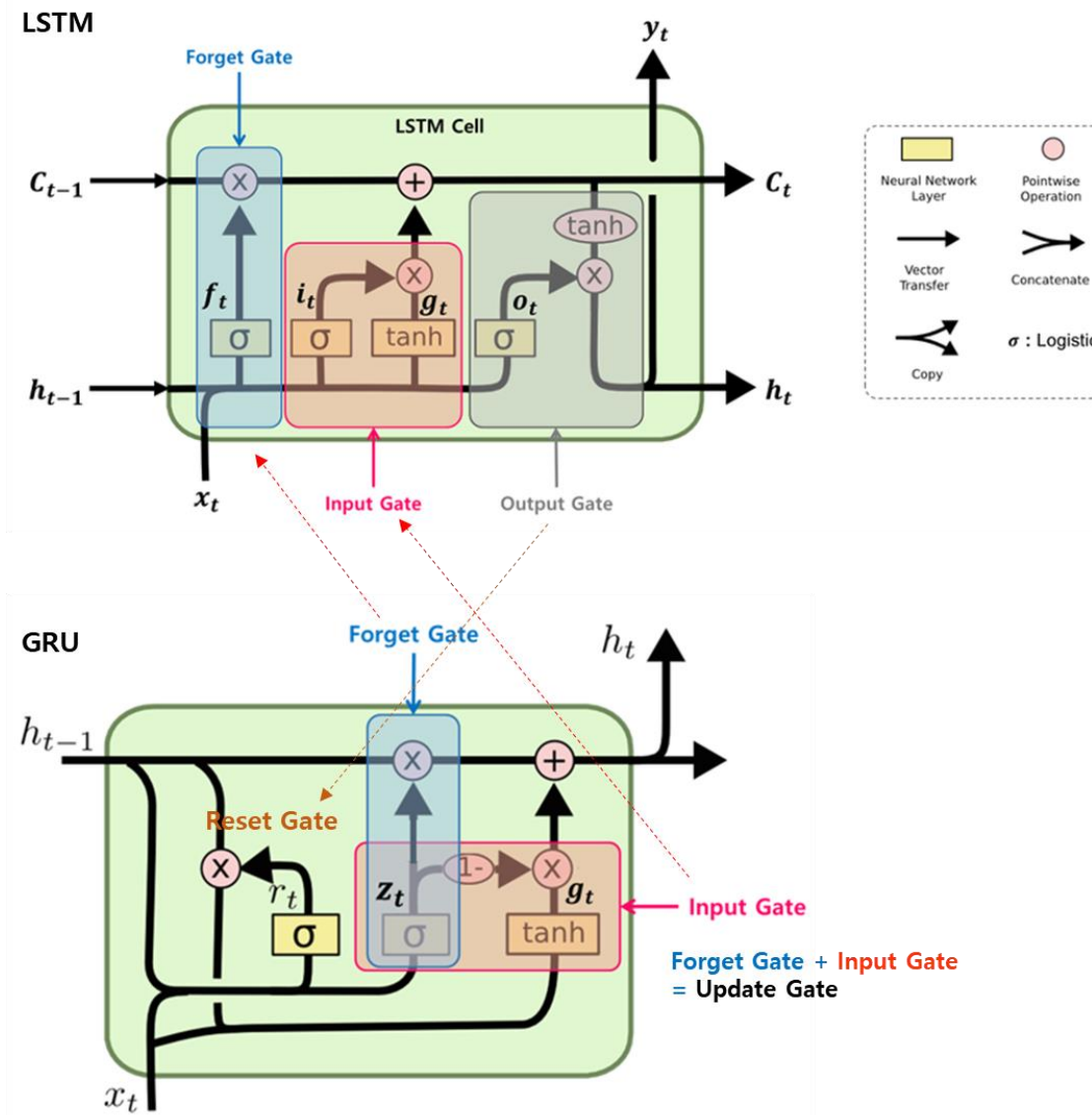
RNN

LSTM

- GRU

GRU - 모델 구조

LSTM과 비교



- <https://velog.io/@lighthouse97/Gated-Recurrent-UnitGRU%EC%9D%98-%EC%9D%B4%ED%95%B4>

감사합니다

배경

RNN

LSTM

GRU

출처

- <https://dgkim5360.tistory.com/entry/understanding-long-short-term-memory-lstm-kr>
- <https://ehpub.co.kr/keras-ann-rnn-handwriting-deeplearning/>
- https://www.researchgate.net/figure/A-DNN-with-three-hidden-layers-L3-Input-and-output-layers-are-counted-separately_fig8_309488185
- <https://www.youtube.com/watch?v=rDkDyGJpp7E&t=17s>
- <https://ratsgo.github.io/natural%20language%20processing/2017/03/09/rnnlstm/>
- <https://coding-yoon.tistory.com/132>
- <https://velog.io/@lighthouse97/Gated-Recurrent-UnitGRU%EC%9D%98-%EC%9D%B4%ED%95%B4>