Recurrent Neural Network - 1
(순환 신경망)

조윤상

DMQA

목차

❖ 순환 신경망 개요

❖ 순환 신경망 학습(Backpropagation Trough Time)

❖ 순환 신경망 모델링

❖ 순환 신경망 한계

DMQA

목차

❖ **순환 신경망 개요**

❖ 순환 신경망 학습(Backpropagation Trough Time)

❖ 순환 신경망 모델링

❖ 순환 신경망 한계

순환 신경망 개요

❖ 기계학습 모델

$$f(\cdot) = \text{Model}$$

설명 변수(X) $f(X) = y$ 반응 변수(y)

변수 / 관측치	x_1	x_2	$\cdots$	x_{p-1}	x_p
N_1					
N_2					
$\cdots$					
N_{n-1}					
N_n					

학습

y

순환 신경망(recurrent neural network, RNN) 이란?

순환 신경망 개요

❖ 순환 신경망(recurrent neural network, RNN)

$$f(\cdot) = \text{RNN Model}$$

설명 변수(X) $f(X) = y$ 반응 변수(y)

관측치 \ 변수	x_1	x_2	…	x_{p-1}	x_p
N_1					Time
N_2					
…		Sequential data			
N_{n-1}					
N_n	Time				

학습

y

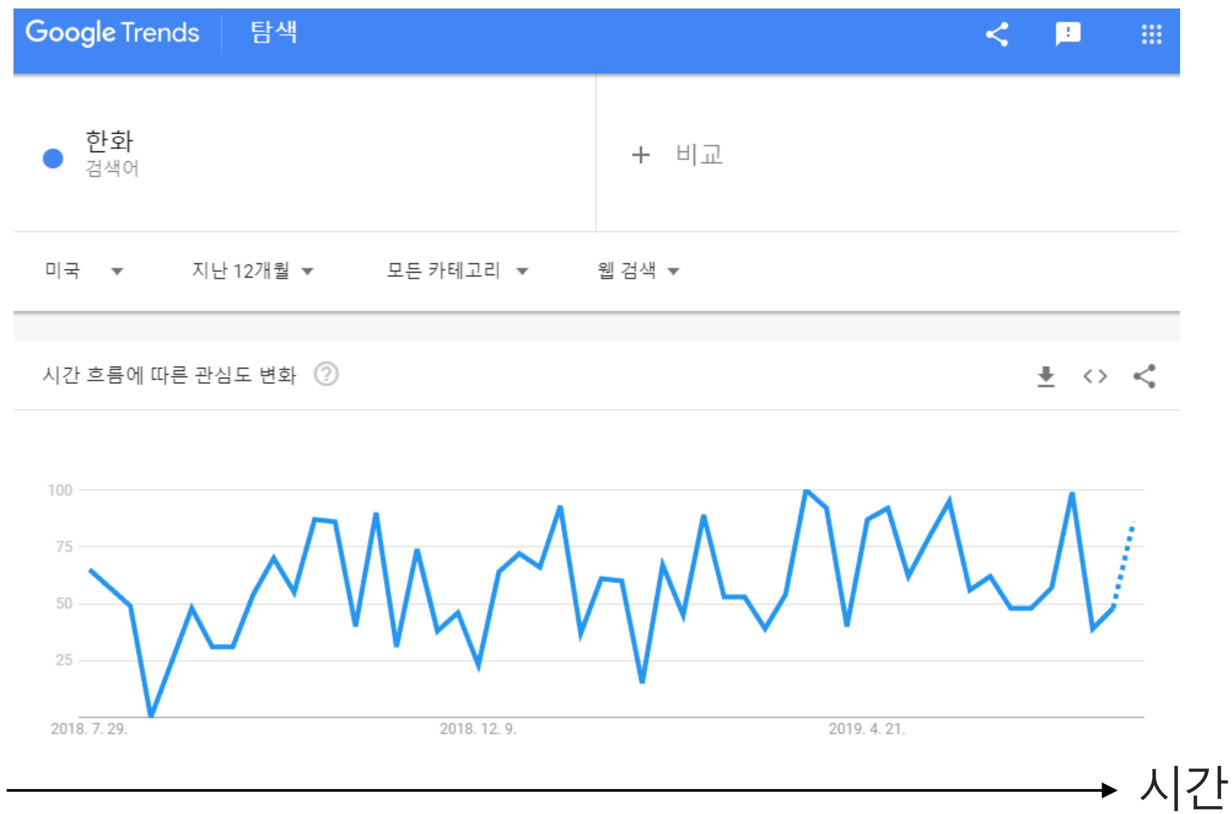
순차 데이터(sequential data) 모델링을 위한 인공 신경망(neural network) 모델

순환 신경망 개요

❖ 순차 데이터(sequential data)

- **변수** 간 or **관측치** 간 순서가 있는 데이터

구글 트렌드 데이터



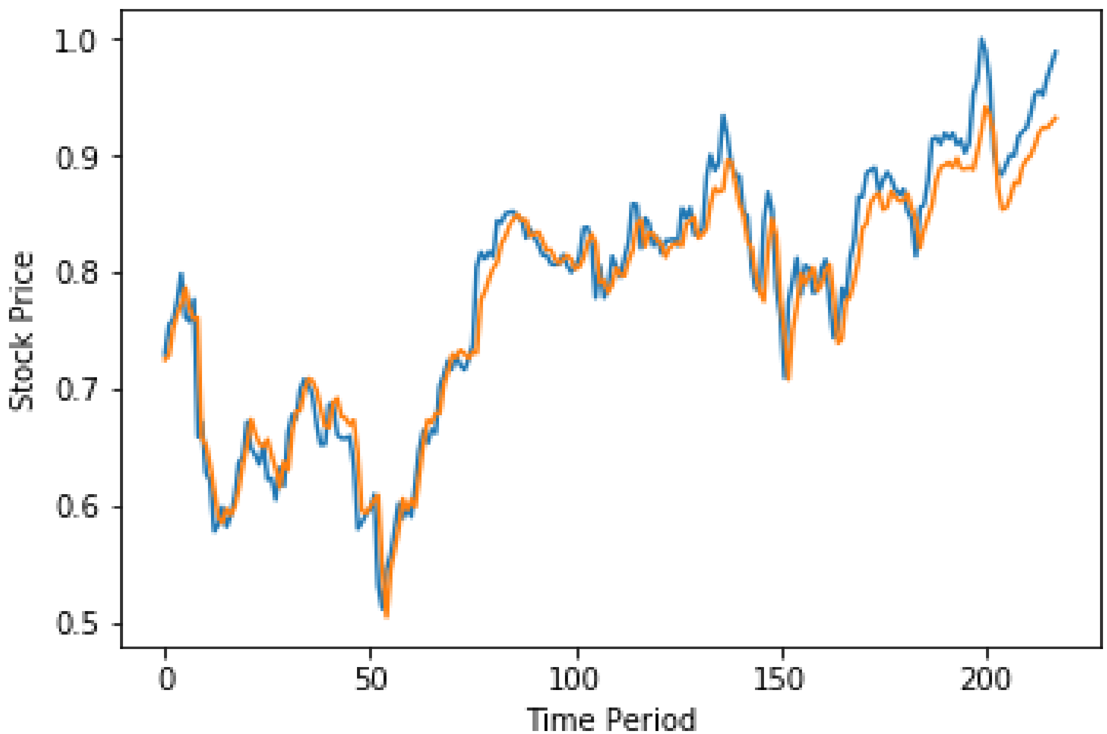
DMQA

순환 신경망 개요

❖ 순차 데이터(sequential data)

- **변수** 간 or **관측치** 간 순서가 있는 데이터

주가 데이터



순환 신경망 개요

❖ 순차 데이터(sequential data)

- **변수** 간 or **관측치** 간 순서가 있는 데이터

텍스트 데이터

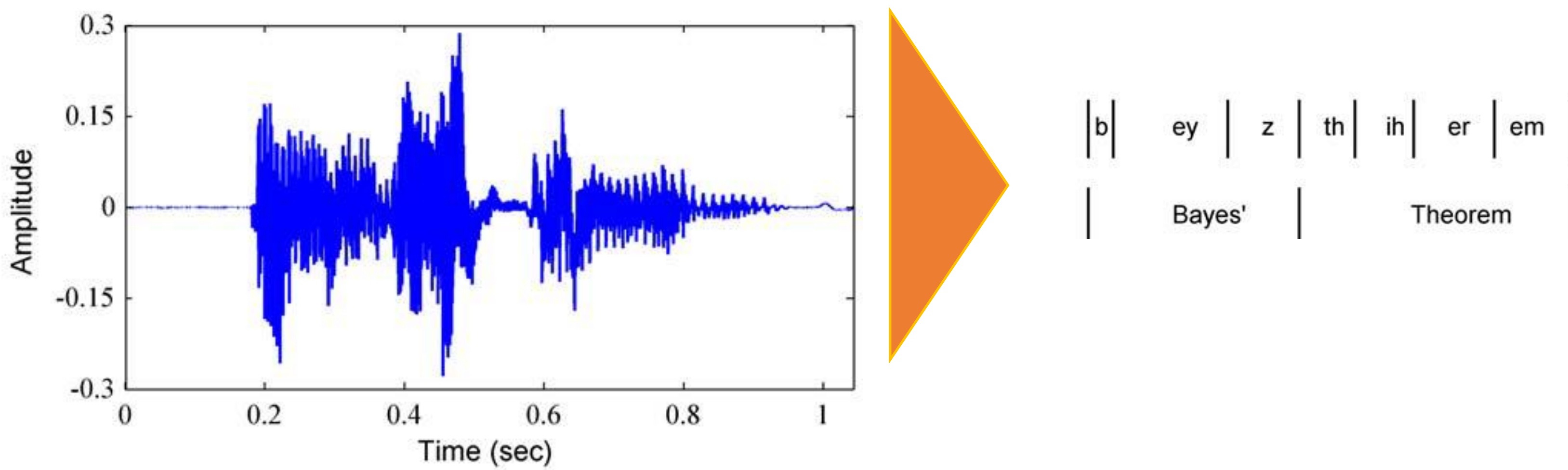
	Review	Rating
0	excellent service, great price, an...	5
1	Comfortable seating with adequate ...	5
2	I like Flix Bus. It seems like an ...	4
3	Very good service, clean bus, big,...	4
4	Great first trip! Great price, cle...	5
5	It was a very bad trip. The bus de...	1
6	Im first time by flixbus, it's nic...	4
7	Good way to travel	4
8	Very fast, very cheap and rather c...	4
9	Booked bus from Bayreuth in German...	1

순환 신경망 개요

❖ 순차 데이터(sequential data)

- **변수** 간 or **관측치** 간 순서가 있는 데이터

음성 데이터



순환 신경망 개요

❖ 순환 신경망(recurrent neural network, RNN) 이란?

- 순차 데이터(sequential data) 모델링을 위한 인공 신경망(neural network) 모델
- 순차 데이터(sequential data): **변수** 간 or **관측치** 간 순서가 있는 데이터

인공신경망

- 1943: Neural networks
- 1957: Perceptron
- 1974-86: Backpropagation, RBM, RNN
- 1989-98: CNN, MNIST, LSTM, Bidirectional RNN
- 2006: "Deep Learning", DBN
- 2009: ImageNet
- 2012: AlexNet, Dropout
- 2014: GANs
- 2014: DeepFace
- 2016: AlphaGo
- 2017: AlphaZero, Capsule Networks
- 2018: BERT

* Dates are for perspective and not as definitive historical record of invention or credit

순환 신경망 구조

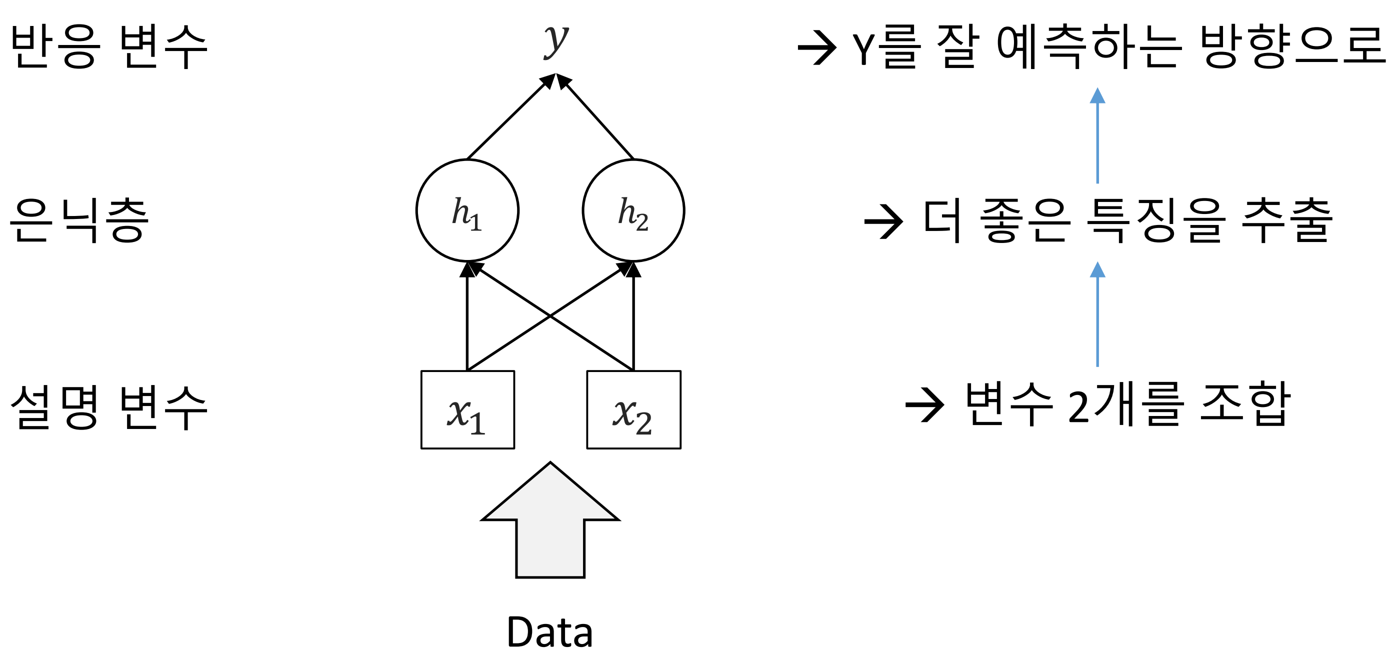
No	RNN Architectures
1	Fully Recurrent Neural Network
2	Recursive Neural Network
3	Hopfield Network
4	Elman Networks And Jordan Networks or Simple Recurrent Network (SRN)
5	Echo State Network
6	Neural History Compressor
7	Long Short-Term Memory (LSTM)
8	Gated Recurrent Unit
9	Bi-Directional Recurrent Neural Network
10	Continuous-Time Recurrent Neural Network (CTRNN)
11	Hierarchical Recurrent Neural Network
12	Recurrent Multilayer Perceptron Network
13	Multiple Timescales Model
14	Neural Turing Machines (NTM)
15	Differentiable Neural Computer (DNC)
16	Neural Network Pushdown Automata (NNPDA)

https://www.rsisinternational.org/journals/ijrsi/digital-library/volume-5-issue-3/124-129.pdf

순환 신경망 개요

❖ 신경망(NN) vs 순환 신경망(RNN)

$$f(\cdot) = \text{Neural Network}$$



DMQA

순환 신경망 개요

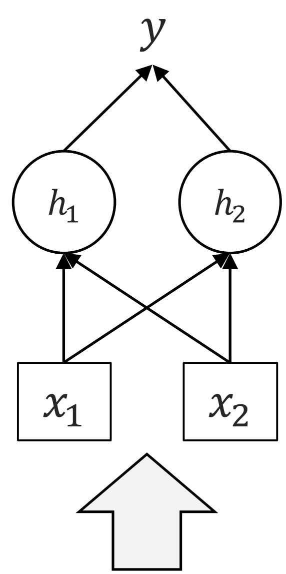
❖ 신경망(NN) vs 순환 신경망(RNN)

$$f(\cdot) = \text{Recurrent Neural Network}$$

반응 변수

은닉층

설명 변수



Data(t=1)

→ Y를 잘 예측하는 방향으로.

→ 더 좋은 특징을 추출

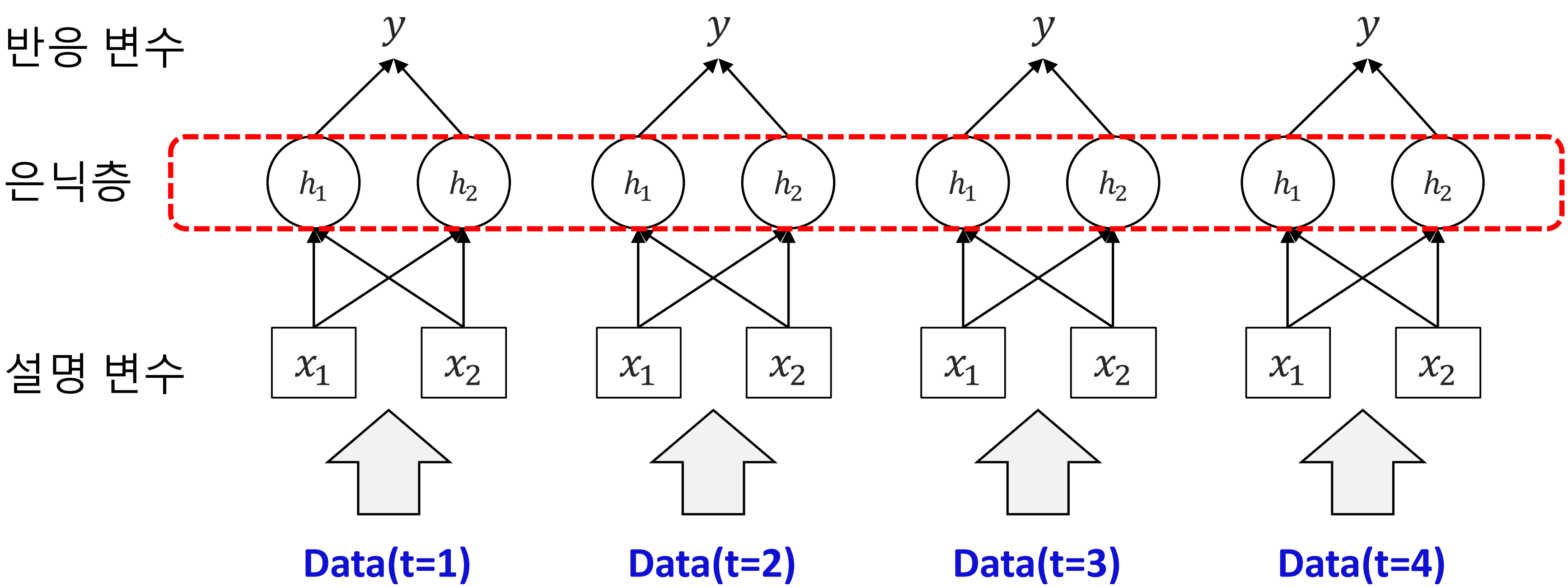
→ 변수 2개를 조합

순환 신경망 개요

❖ 신경망(NN) vs 순환 신경망(RNN)

$$f(\cdot) = \text{Recurrent Neural Network}$$

시간을 반영하는 것은 Hidden layer, 어떻게?

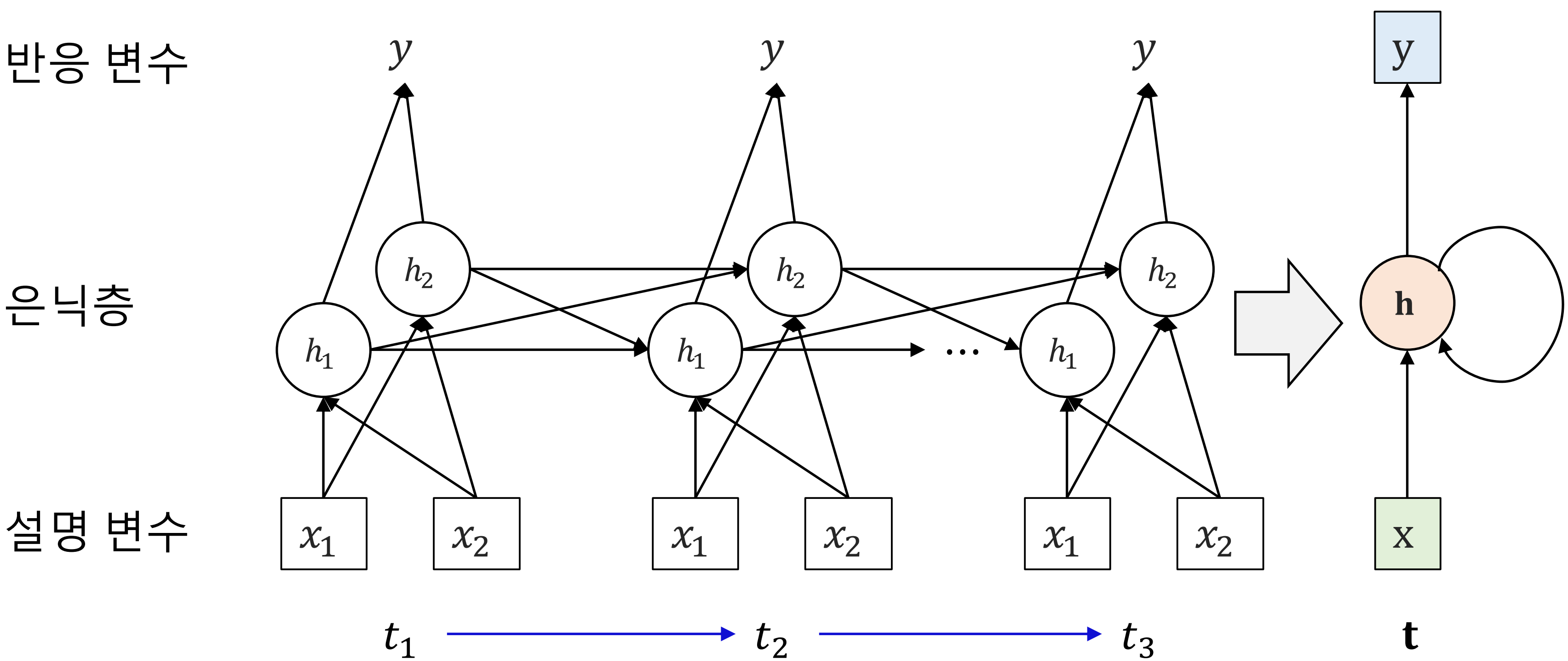


DMQA

순환 신경망 개요

❖ 신경망(NN) vs 순환 신경망(RNN)

$$f(\cdot) = \text{Recurrent Neural Network}$$



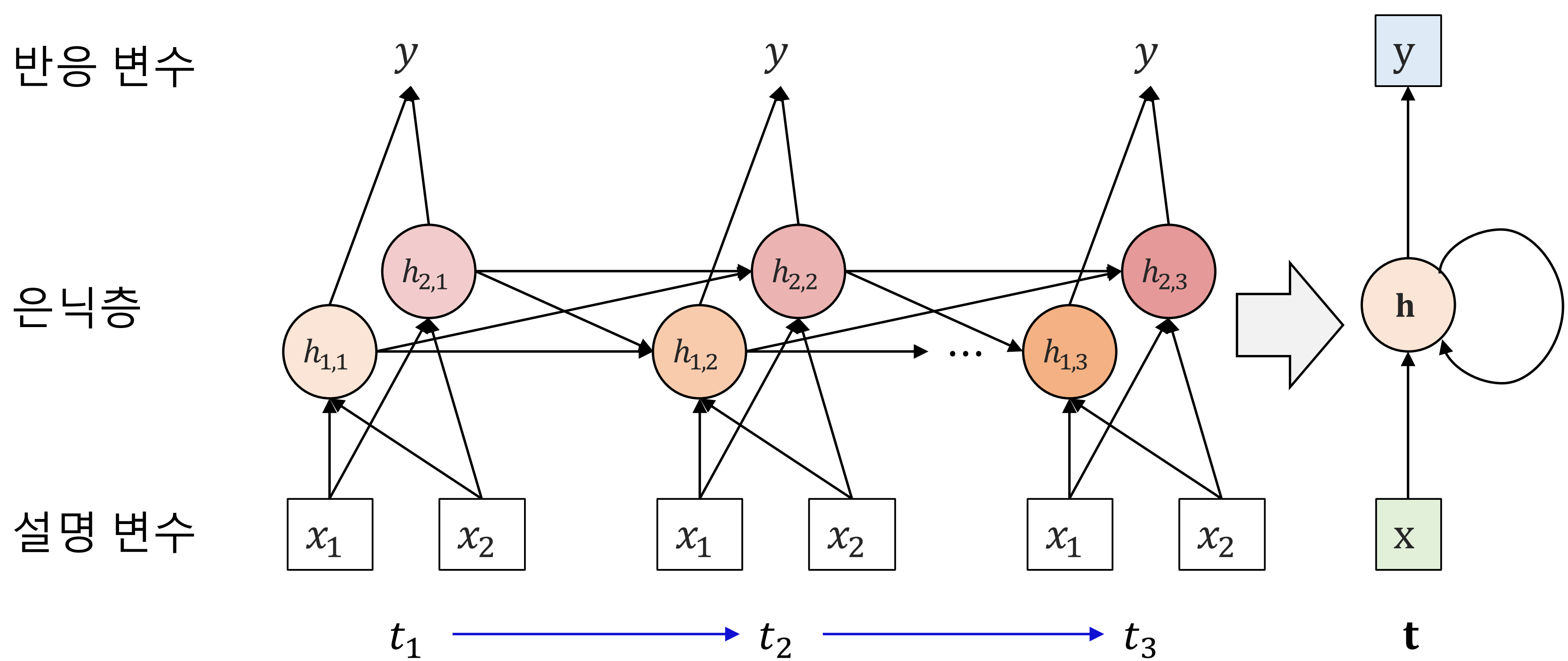
DMQA

순환 신경망 개요

❖ 순환 신경망(RNN)

t 시점 y를 예측

$$f(\cdot) = \text{Recurrent Neural Network}$$



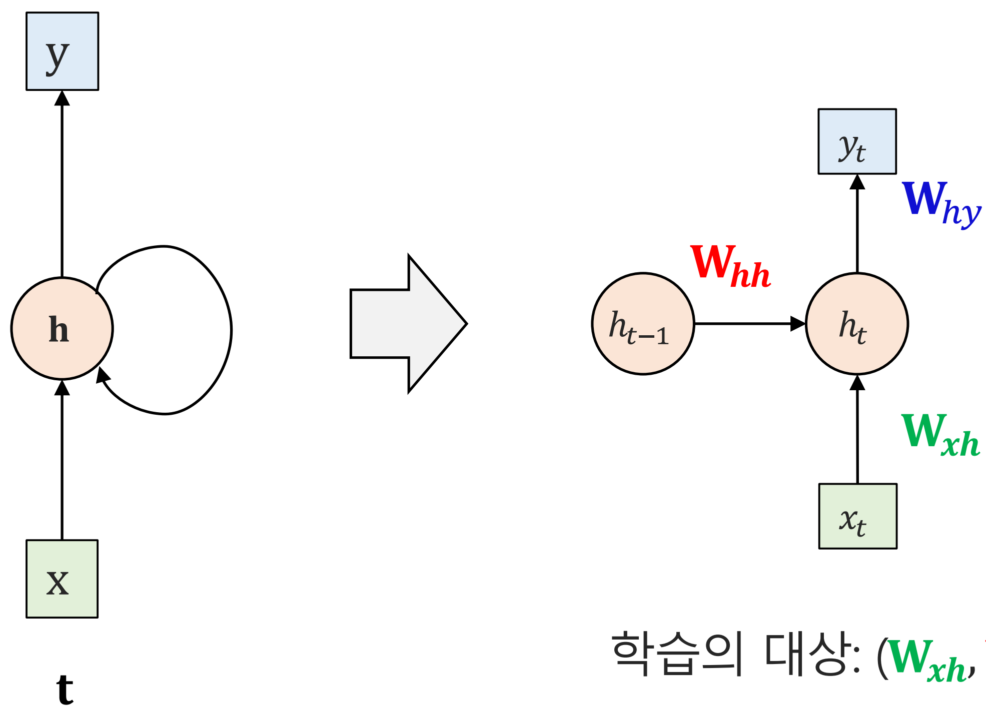
순환 신경망 개요

- t 시점 이전 정보와, → **hidden layder로부터 전달되는 weight(W_{hh})**
- t 시점 input 데이터 정보를 가지고 → W_{xh}
- t 시점의 y를 예측 → W_{hy}

반응 변수

은닉층

설명 변수

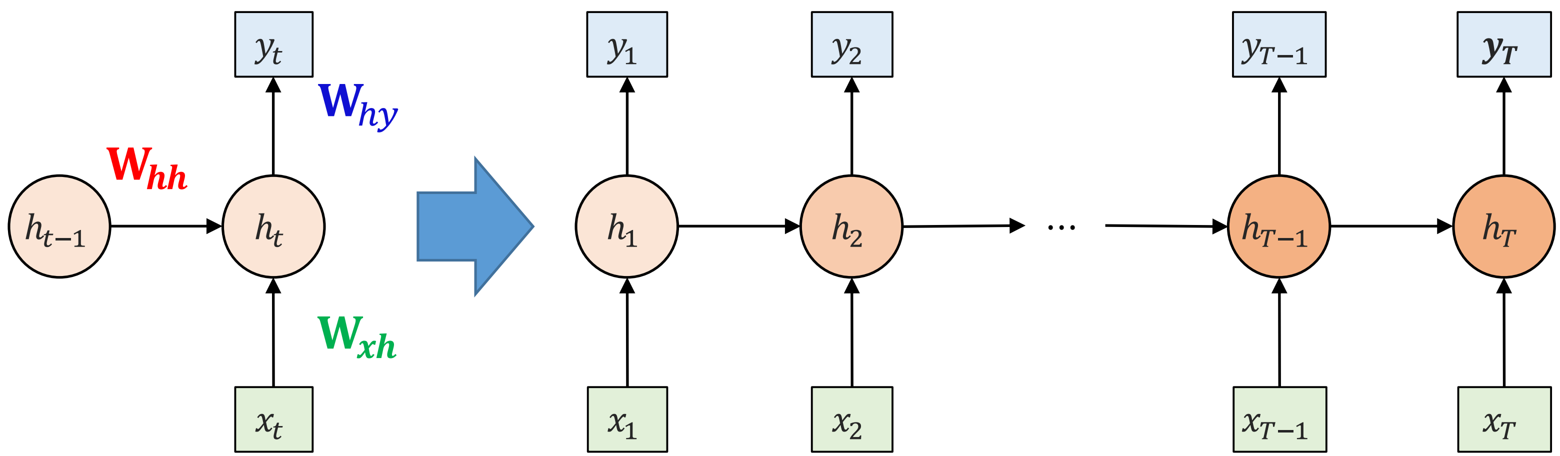


학습의 대상: (W_{xh}, W_{hh}, W_{hy})

순환 신경망 개요

❖ RNN 구조: t 시점 이전 정보, t시점 데이터 X, 를 가지고 t 시점 Y를 예측

- $y_t = g(W_{xh}h_t + b_y)$
- $h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_x)$



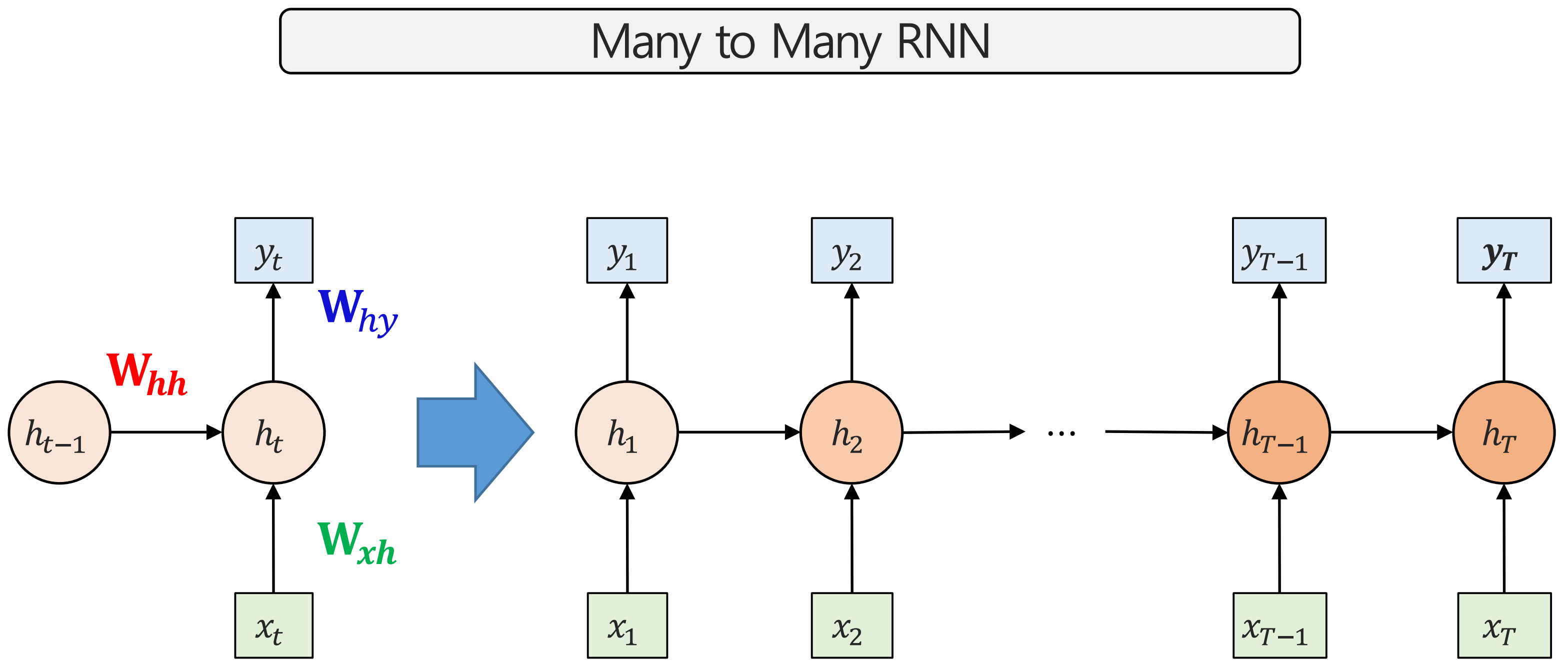
학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)　　학습 방법: Backpropagation **Trough Time**

DMQA

순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조



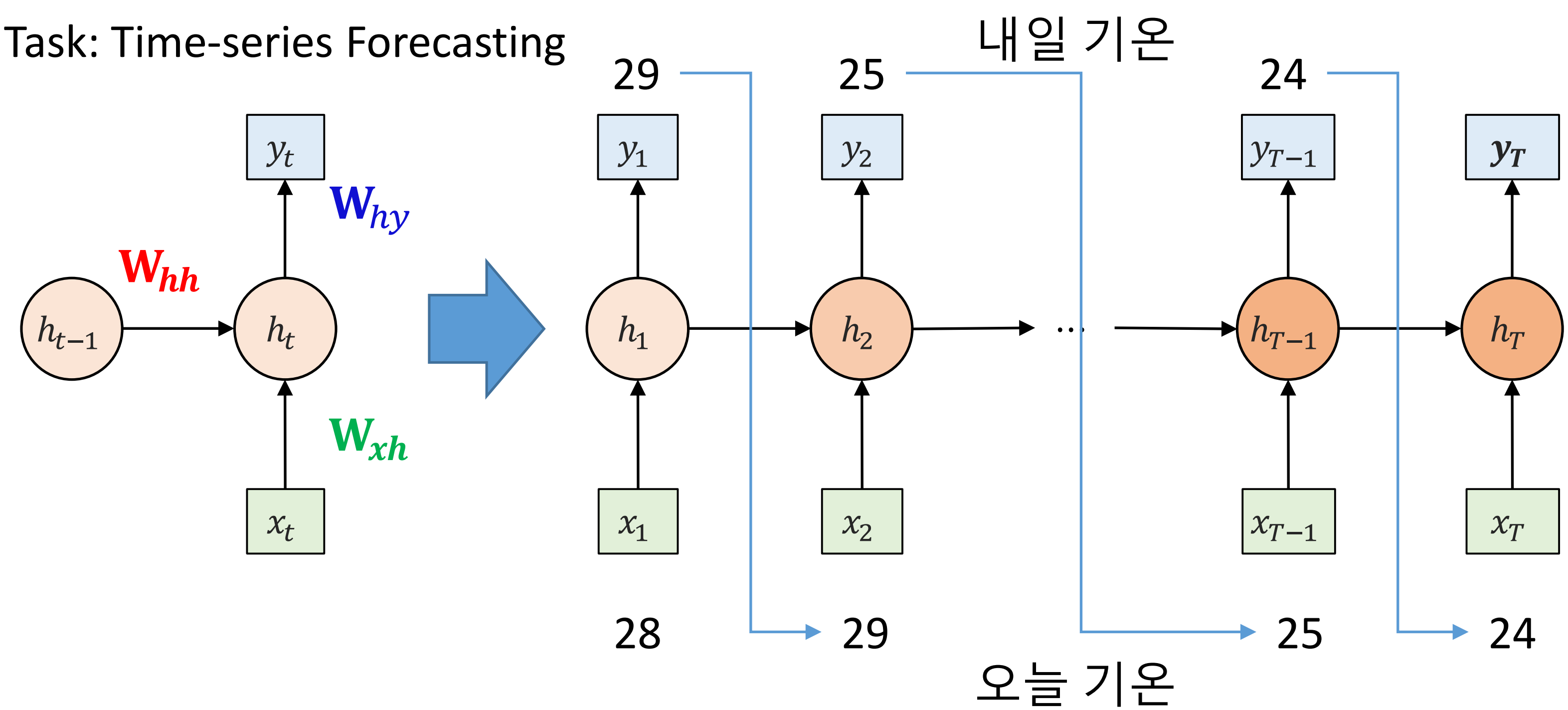
순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

Many to Many RNN

Task: Time-series Forecasting



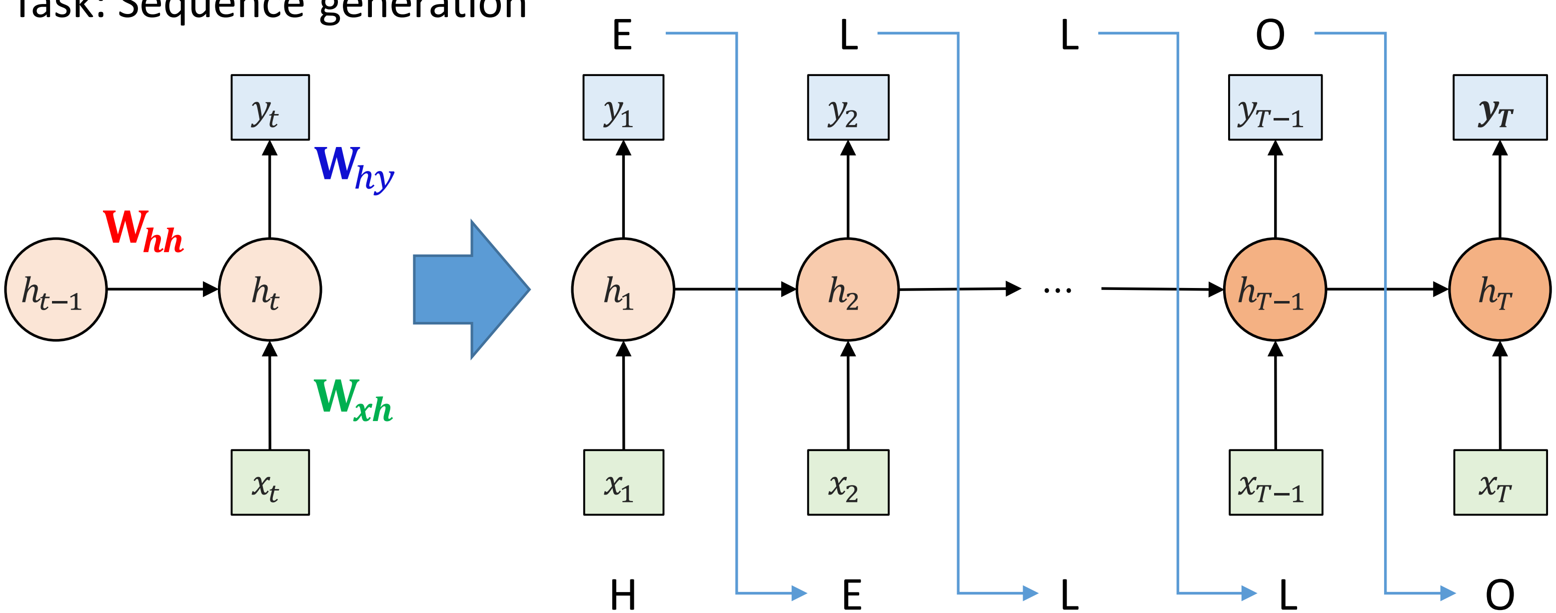
순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

Many to Many RNN

Task: Sequence generation

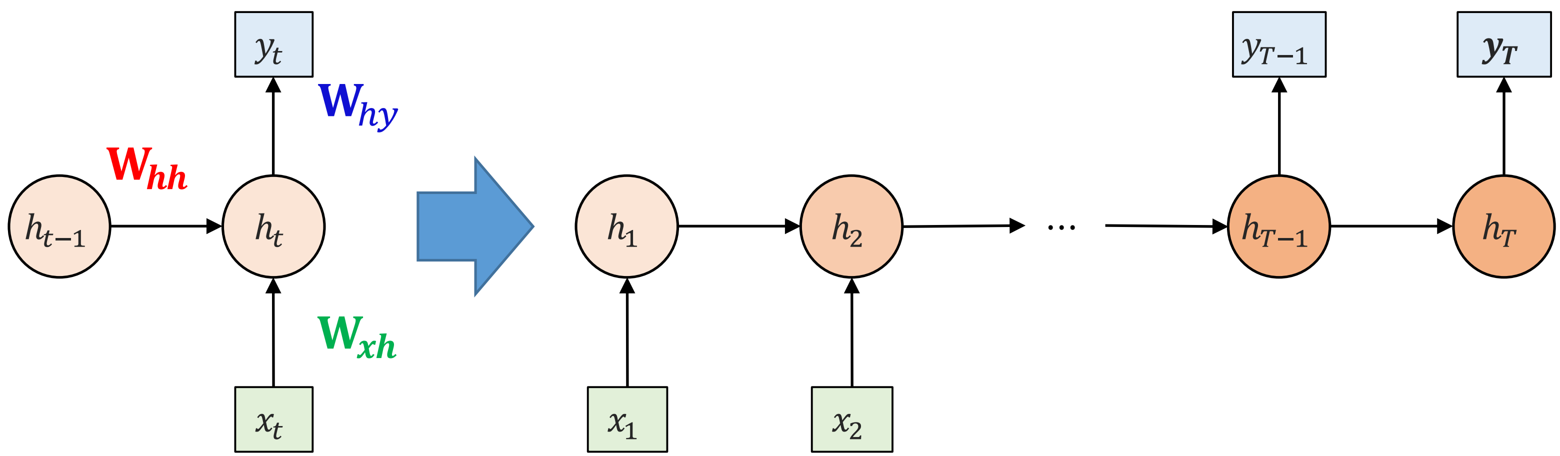


순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

Many to Many RNN



DMQA

순환 신경망 개요

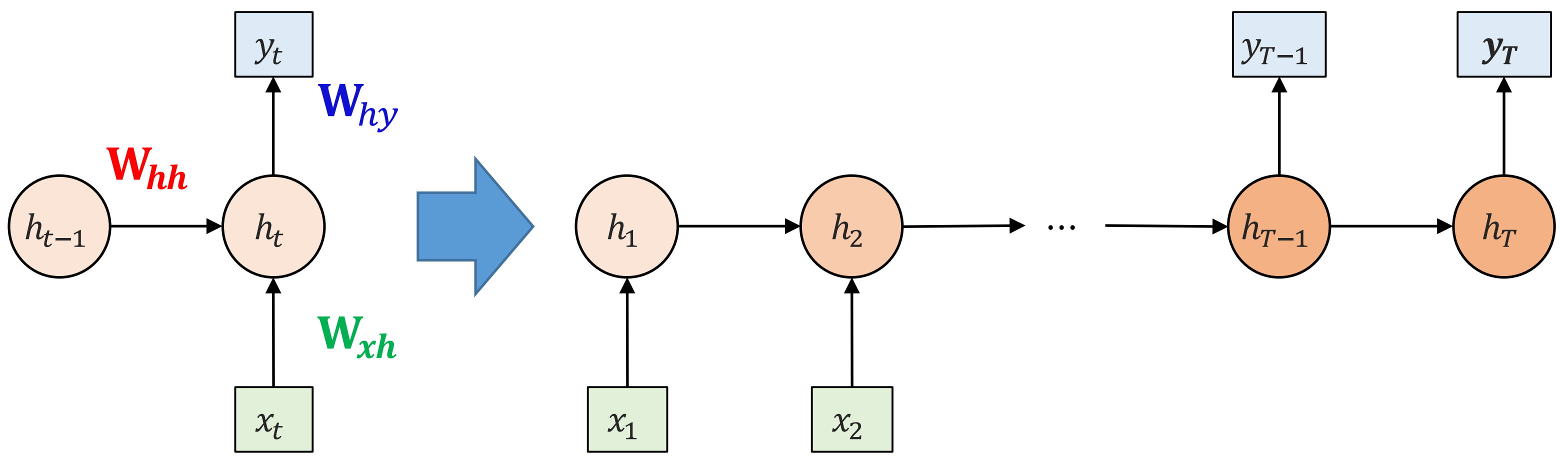
학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

Many to Many RNN

Task: Machine Translation

(한국어) 안녕하세요 조 박사님, 한화입니다.



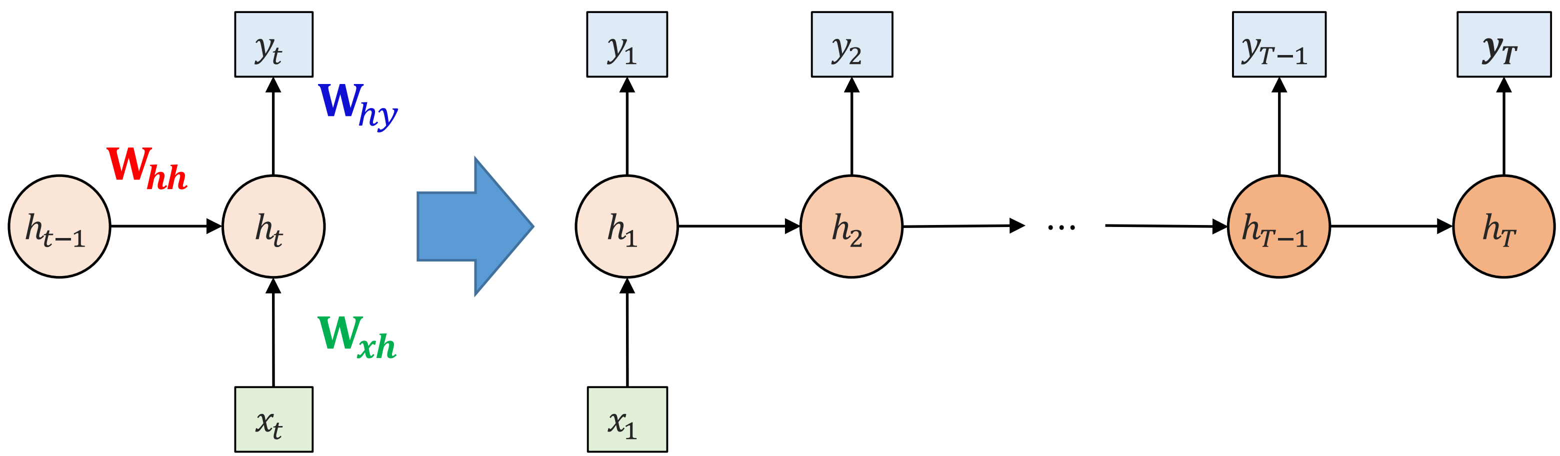
(English) Hello Dr. Cho, This is Hanhwa.

순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

One to Many RNN



순환 신경망 개요

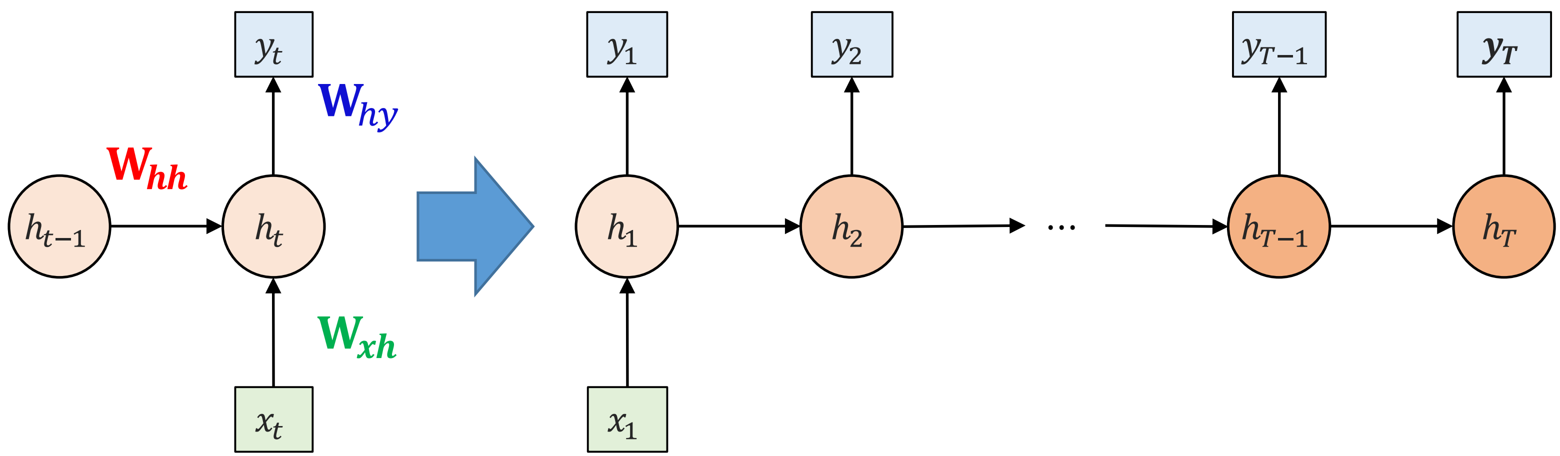
학습의 대상: $(\mathbf{W}_{xh}, \mathbf{W}_{hh}, \mathbf{W}_{hy})$

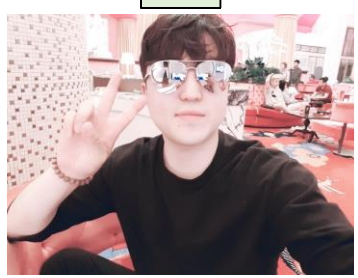
❖ RNN 구조

One to Many RNN

Task: Image Captioning

… 선글라스 낀 남자가 V 를 하고 있다. …





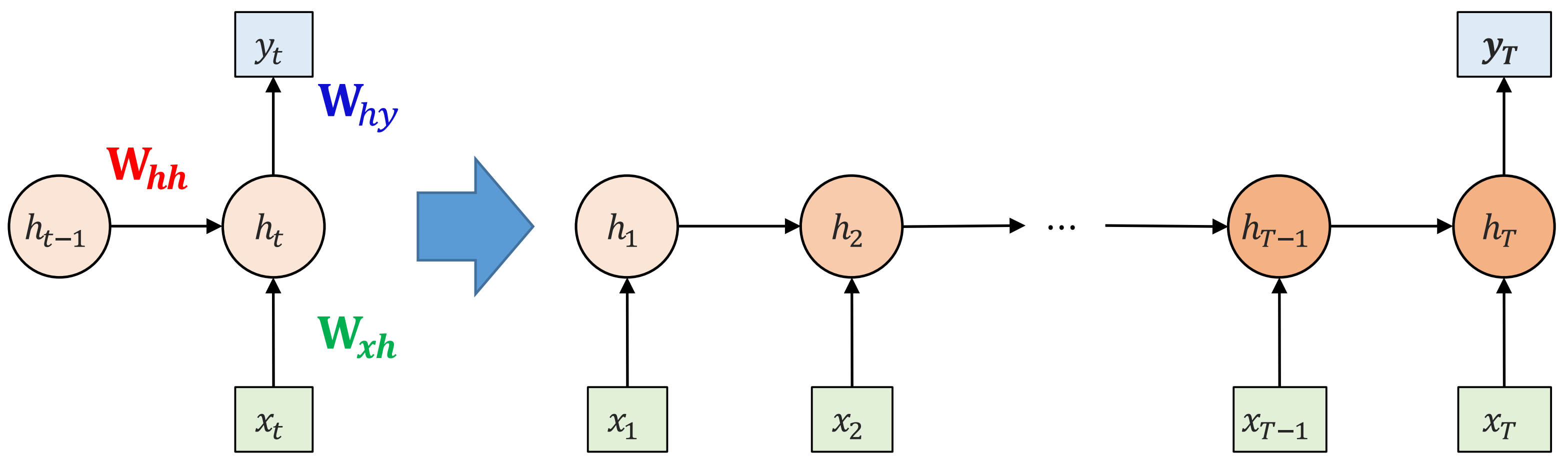
DMQA

순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

Many to One RNN



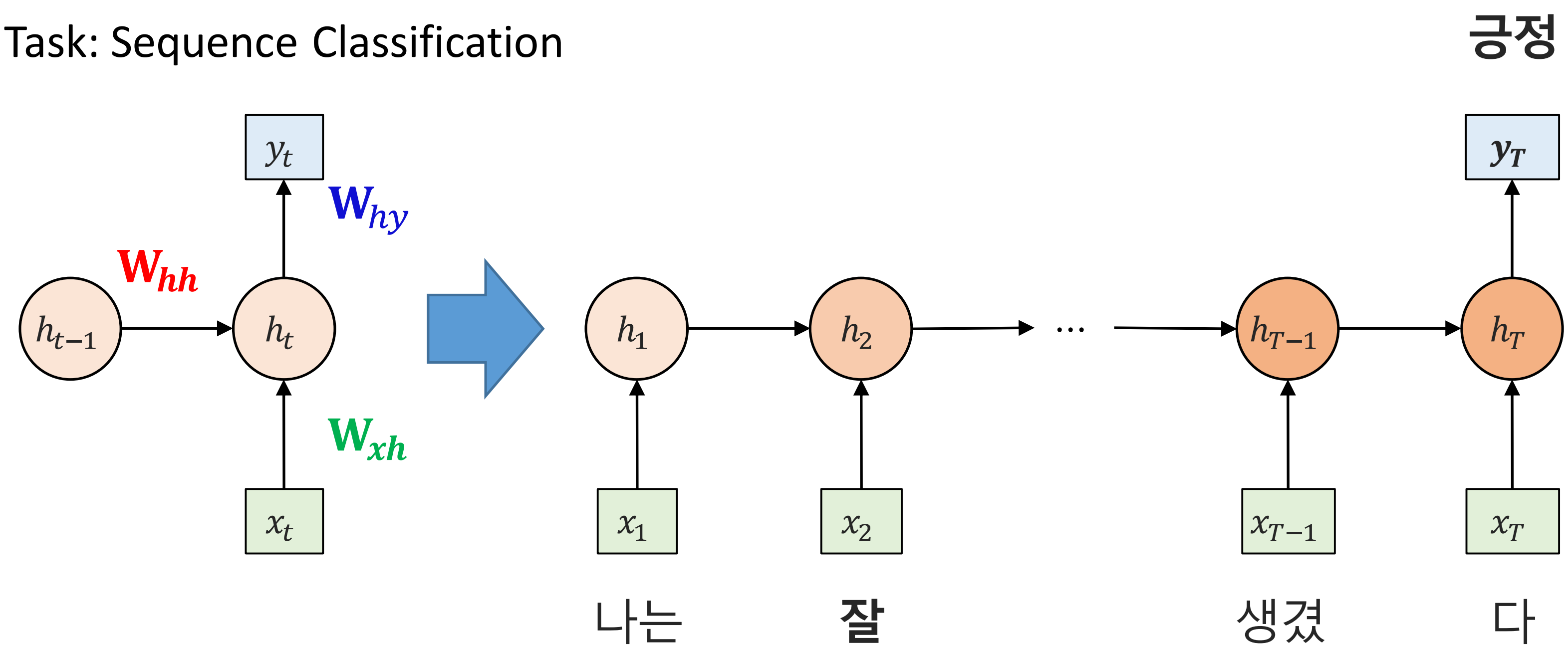
순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

Many to One RNN

Task: Sequence Classification



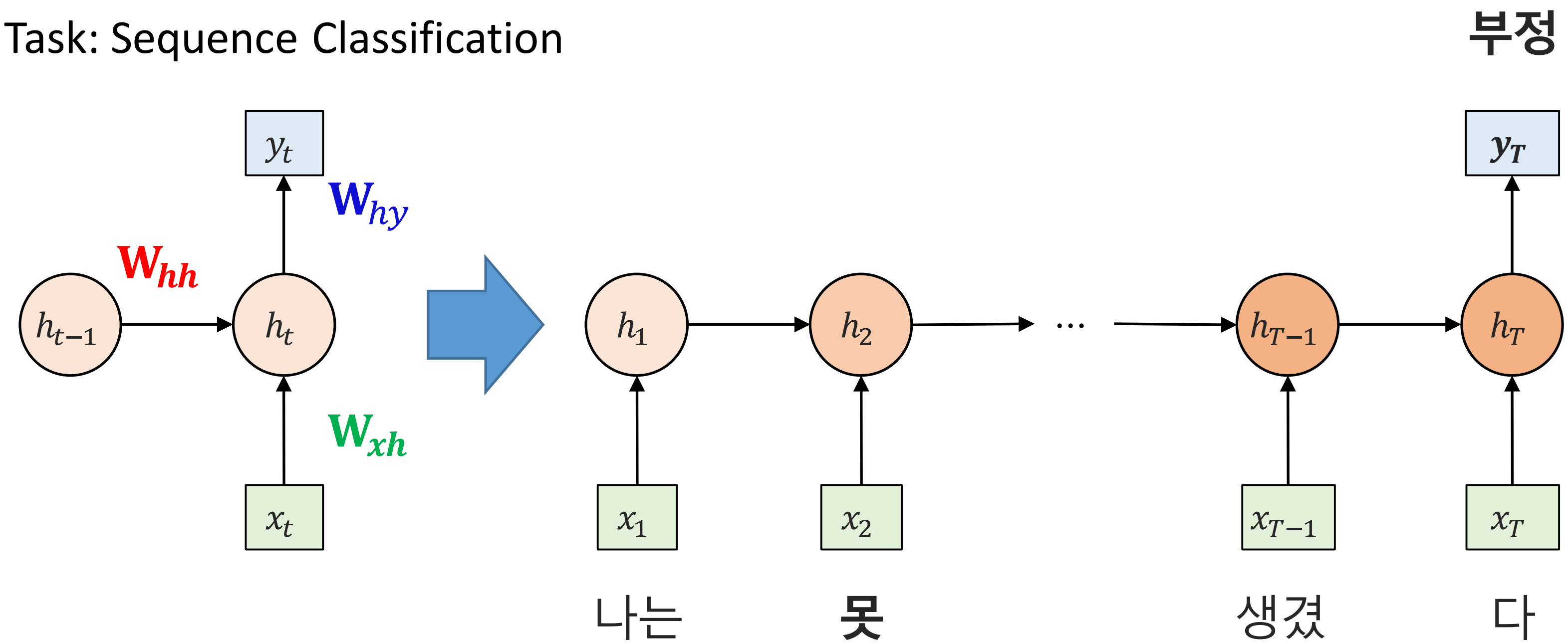
순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조

Many to One RNN

Task: Sequence Classification



DMQA

목차

❖ 순환 신경망 개요

❖ **순환 신경망 학습(Backpropagation Trough Time)**

❖ 순환 신경망 모델링

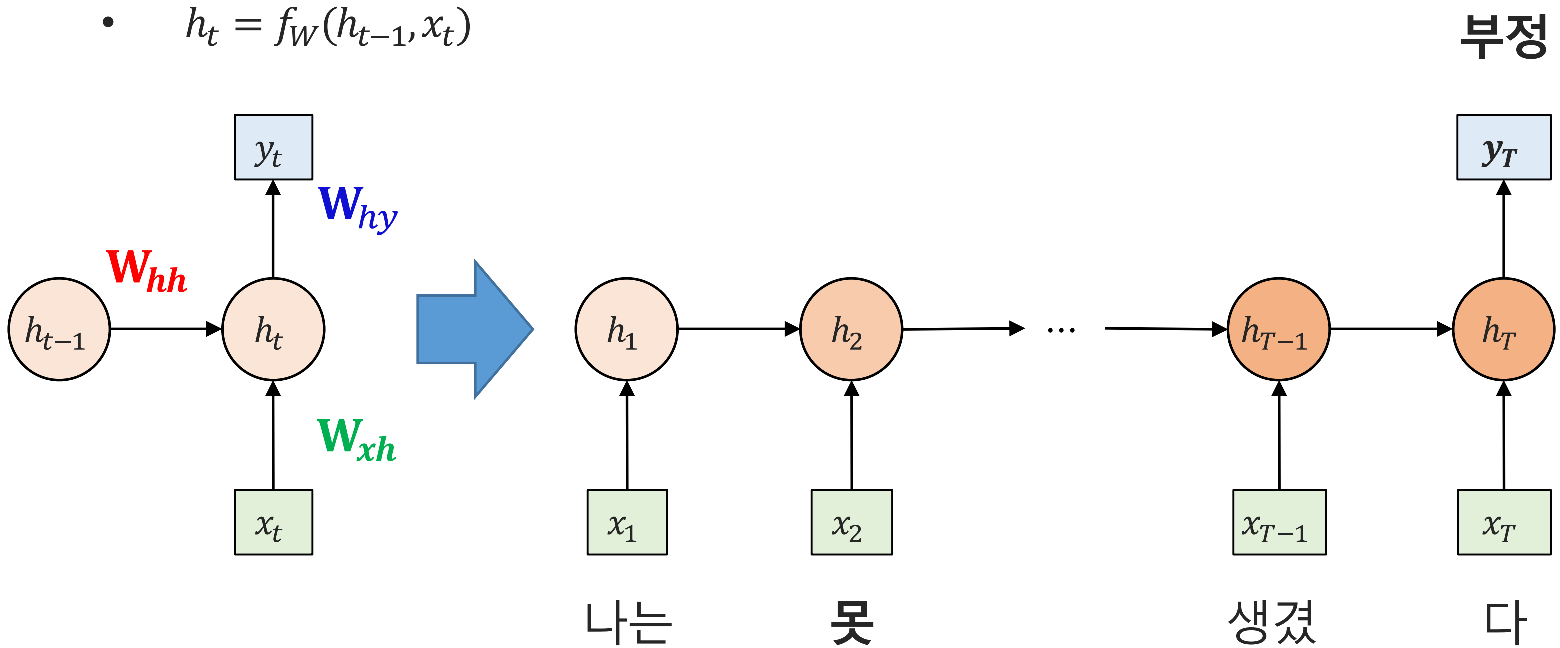
❖ 순환 신경망 한계

순환 신경망 개요

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

❖ RNN 구조: t 시점 이전 정보, t시점 데이터 X, 를 가지고 t 시점 Y를 예측

- $y_t = g(W_{xh}h_t + b_y)$
- $h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_x)$
  - $h_t = f_W(h_{t-1}, x_t)$



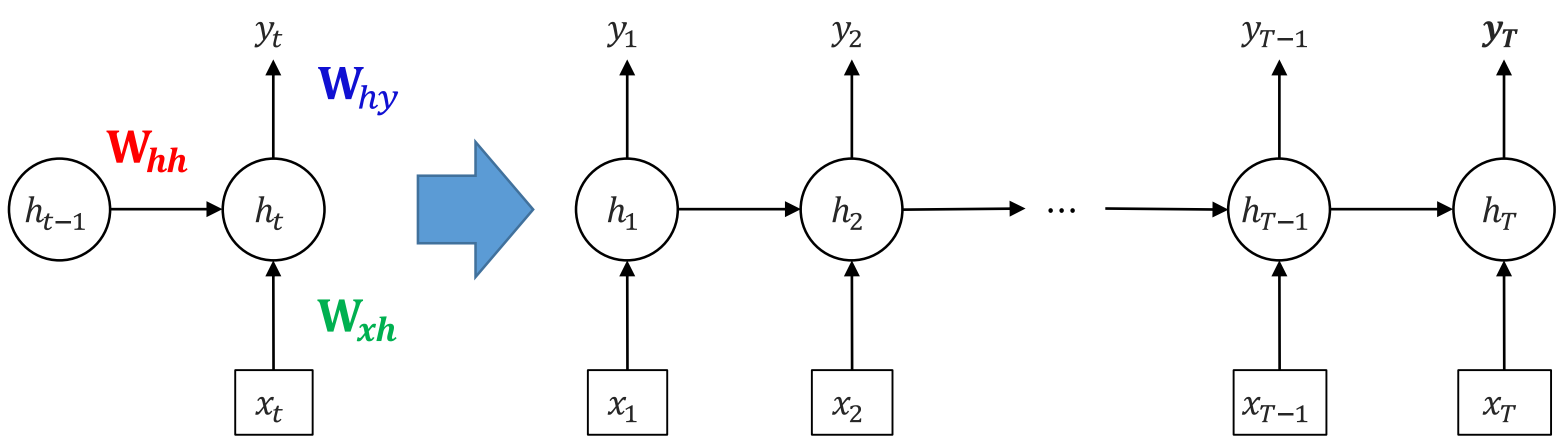
학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$) 학습 방법: Backpropagation **Trough Time**

순환 신경망 개요

❖ RNN 구조: t 시점 이전 정보, t시점 데이터 X, 를 가지고 t 시점 Y를 예측

- $y_t = g(W_{xh}h_t + b_y)$
- $h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_x)$

Many to Many RNN



학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

학습 방법: Backpropagation **Trough Time**

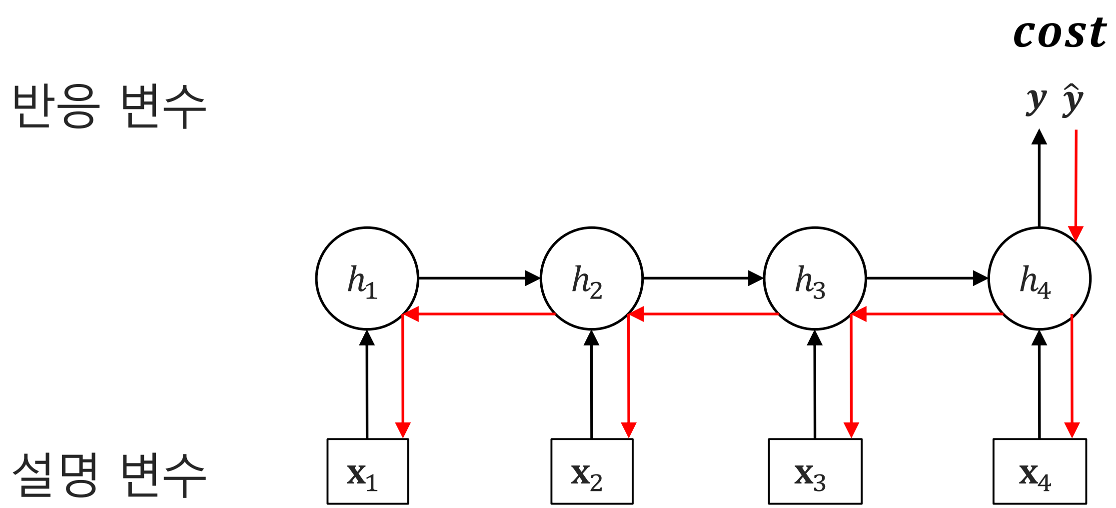
순환 신경망 학습

❖ RNN 학습 방식: 시간 흐름에 따라 파라미터 업데이트

❖ Backpropagation **Trough Time (BPTT)**

학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

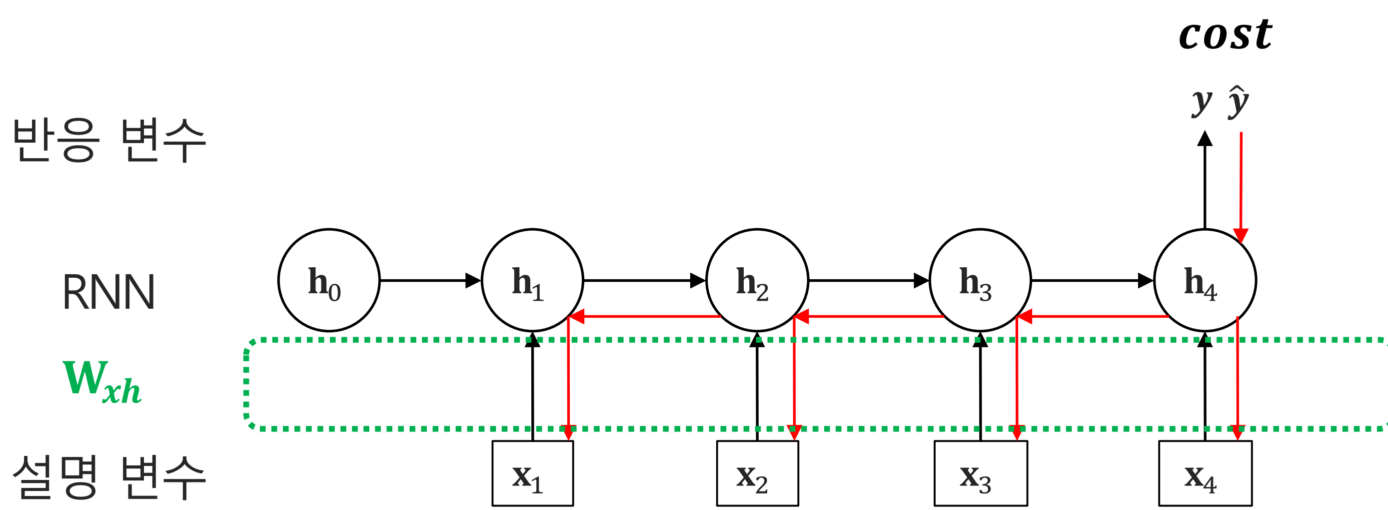
- $y_t = g(W_{xh}h_t + b_y)$
- $h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_x)$ → $f(\cdot)$를 tanh 함수라 하자



DMQA

순환 신경망 학습

- ❖ RNN 학습 방식: Backpropagation **<u>Trough Time (BPTT)</u>**
- ❖ 학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)



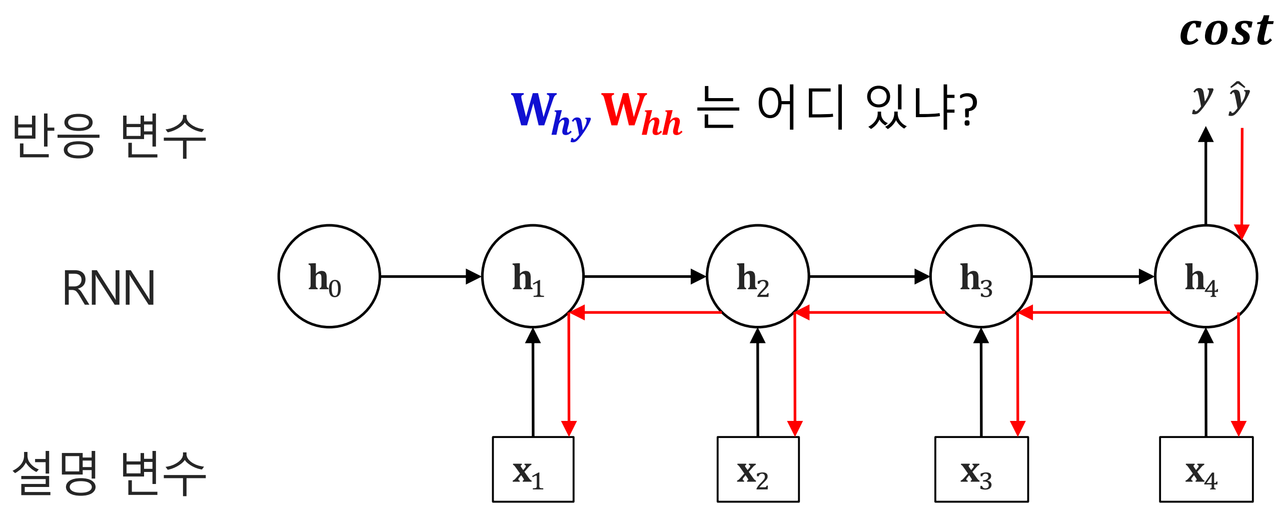
$$\frac{\partial Cost}{\partial \mathbf{W}_{xh}} = \frac{\partial Cost}{\partial y} \times \frac{\partial y}{\partial \mathbf{h}_4} \times \frac{\partial \mathbf{h}_4}{\partial \mathrm{W}_{xh}} + \frac{\partial Cost}{\partial y} \times \frac{\partial y}{\partial \mathbf{h}_4} \times \frac{\partial \mathbf{h}_4}{\partial \mathbf{h}_3} \times \frac{\partial \mathbf{h}_3}{\partial \mathrm{W}_{xh}} +$$

$$\frac{\partial Cost}{\partial y} \times \frac{\partial y}{\partial \mathbf{h}_4} \times \frac{\partial \mathbf{h}_4}{\partial \mathbf{h}_3} \times \frac{\partial \mathbf{h}_3}{\partial \mathbf{h}_2} \frac{\partial \mathbf{h}_2}{\partial \mathrm{W}_{xh}} + \frac{\partial Cost}{\partial y} \times \frac{\partial y}{\partial \mathbf{h}_4} \times \frac{\partial \mathbf{h}_4}{\partial \mathbf{h}_3} \times \frac{\partial \mathbf{h}_3}{\partial \mathbf{h}_2} \times \frac{\partial \mathbf{h}_2}{\partial \mathbf{h}_1} \times \frac{\partial \mathbf{h}_2}{\partial \mathbf{h}_1}$$

순환 신경망 학습

❖ RNN 학습 방식: Backpropagation **<u>Trough Time (BPTT)</u>**

❖ 학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)



$\mathbf{W}_{hy}$ $\mathbf{W}_{hh}$ 는 어디 있냐?

$$\frac{\partial Cost}{\partial \mathbf{W}_{xh}} = \sum_{i=1}^{n} \frac{\partial Cost}{\partial y} \cdot \boxed{\frac{\partial y}{\partial \mathbf{h}_n}} \cdot \left(\boxed{\prod_{j=i}^{n-1} \frac{\partial \mathbf{h}_{j+1}}{\partial \mathbf{h}_j}} \right) \cdot \frac{\partial \mathbf{h}_i}{\partial \mathbf{W}_{xh}}$$

순환 신경망 학습

❖ RNN 학습 방식: Backpropagation **Trough Time (BPTT)**

❖ 학습의 대상: ($\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$, $\mathbf{W}_{hy}$)

$$\frac{\partial Cost}{\partial \mathbf{W}_{xh}} = \sum_{i=1}^{n} \frac{\partial Cost}{\partial y} \cdot \boxed{\frac{\partial y}{\partial \mathbf{h}_n}} \cdot \left(\boxed{\prod_{j=i}^{n-1} \frac{\partial \mathbf{h}_{j+1}}{\partial \mathbf{h}_j}} \right) \cdot \frac{\partial \mathbf{h}_i}{\partial \mathbf{W}_{xh}}$$

$$\mathbf{h}_t = f(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{xh}\mathbf{x}_t + \mathbf{b}_x) = \tanh(\mathbf{z}_t)$$

$$\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}} = \frac{\partial \mathbf{h}_t}{\partial \mathbf{z}_t} \times \frac{\partial \mathbf{z}_t}{\partial \mathbf{h}_{t-1}} = \left(1 - tanh^2(\mathbf{z}_t)\right) \cdot \mathbf{W}_{hh}$$

$$\frac{\partial \mathbf{h}_t}{\partial \mathbf{W}_{xh}} = \frac{\partial \mathbf{h}_t}{\partial \mathbf{z}_t} \times \frac{\partial \mathbf{z}_t}{\partial \mathbf{W}_{xh}} = \left(1 - tanh^2(\mathbf{z}_t)\right) \cdot \mathbf{x}_t$$

DMQA

순환 신경망 학습

❖ RNN 구조

One-to-Many | Many-to-One | Many-to-Many | Many-to-Many

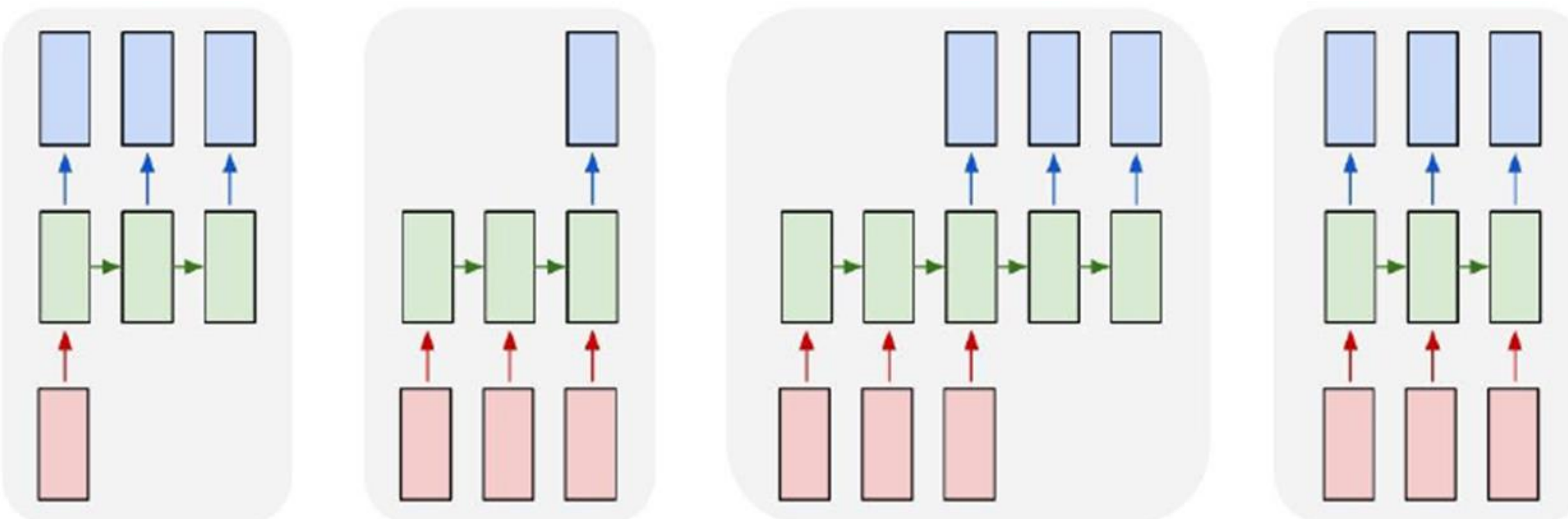


Image Captioning
→ Image to sequence of words

감정 분류

기계 번역

비디오 분류
(프레임단위)

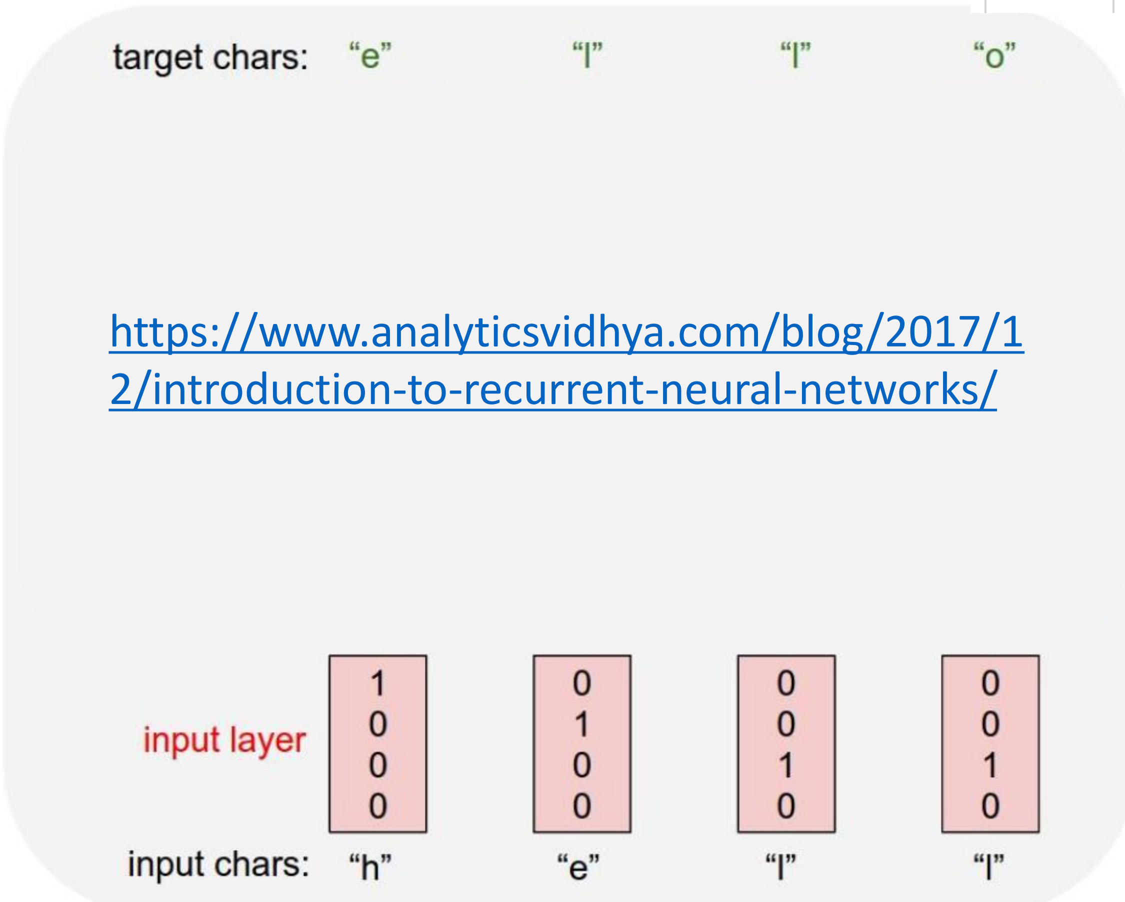
- https://cs224d.stanford.edu/

DMQA

순환 신경망 모델링

❖ Many to Many RNN for Language modeling

wxh			
0.287027	0.84606	0.572392	0.486813
0.902874	0.871522	0.691079	0.18998
0.537524	0.09224	0.558159	0.491528



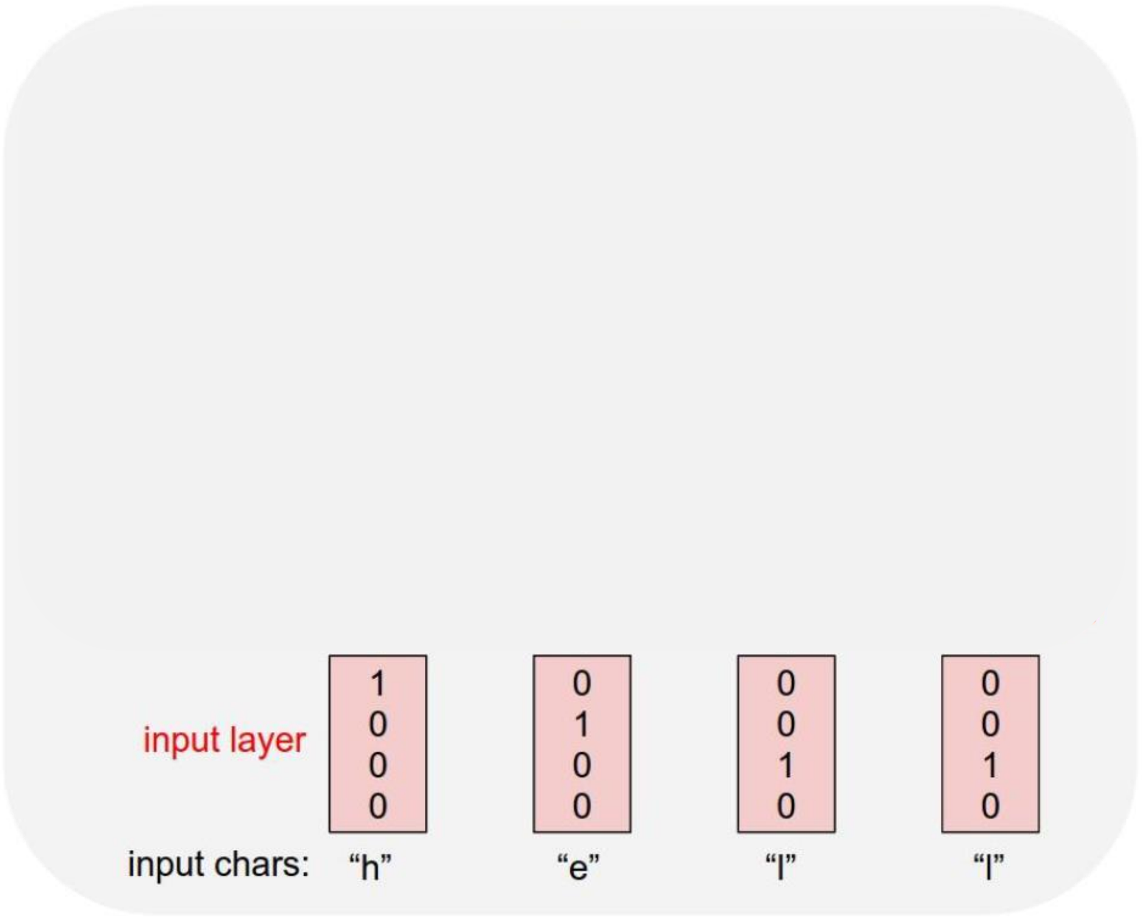
https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

W_hy
W_hh
W_xh

- https://cs224d.stanford.edu/

순환 신경망 모델링

❖ Many to Many RNN for Text modeling

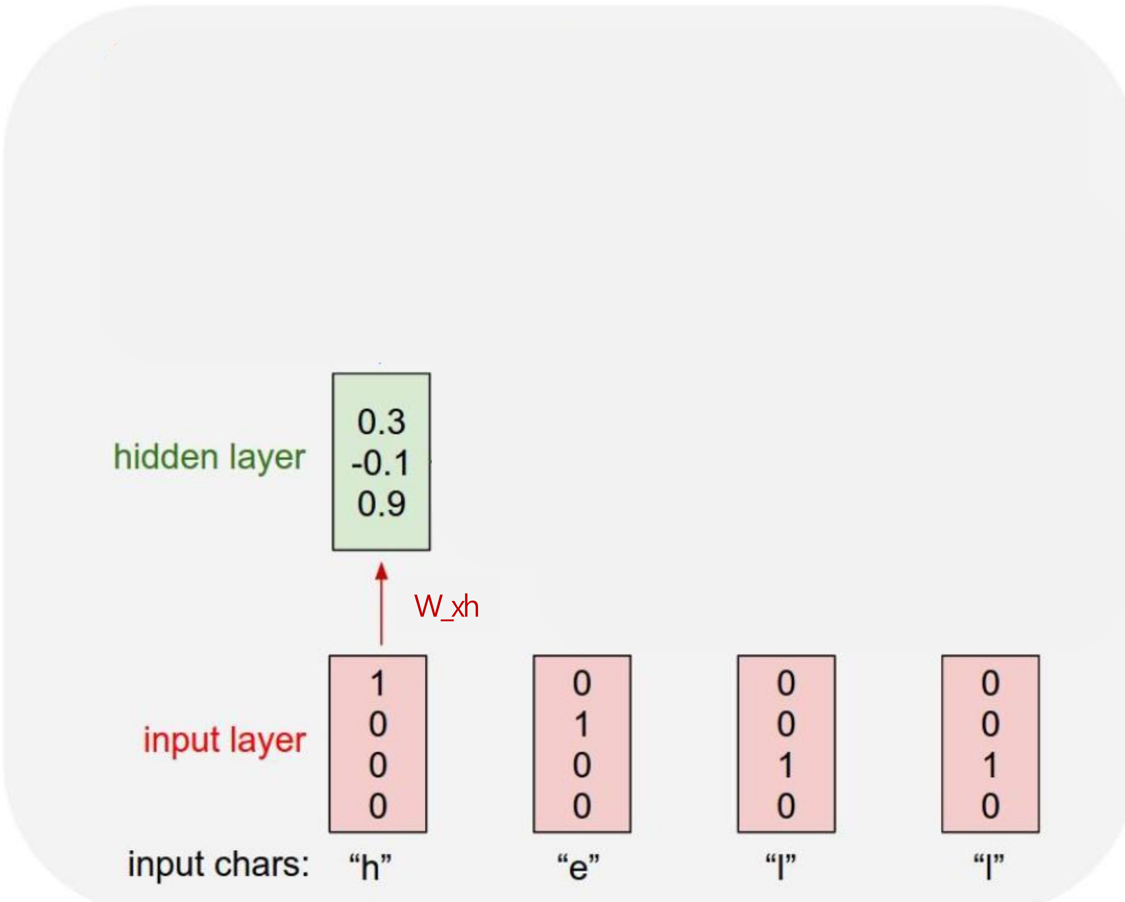


- https://cs224d.stanford.edu/

DMQA

순환 신경망 모델링

❖ Many to Many RNN for Text modeling

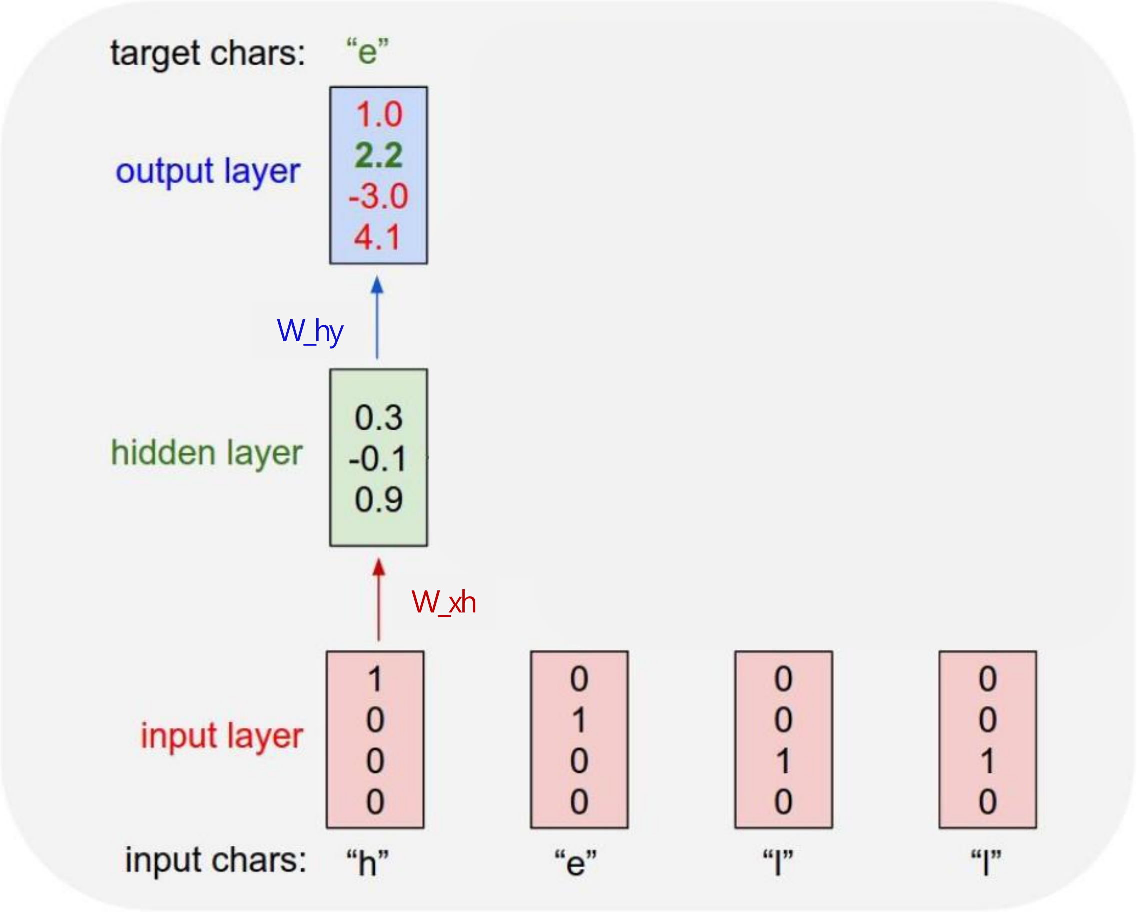


- https://cs224d.stanford.edu/

DMQA

순환 신경망 모델링

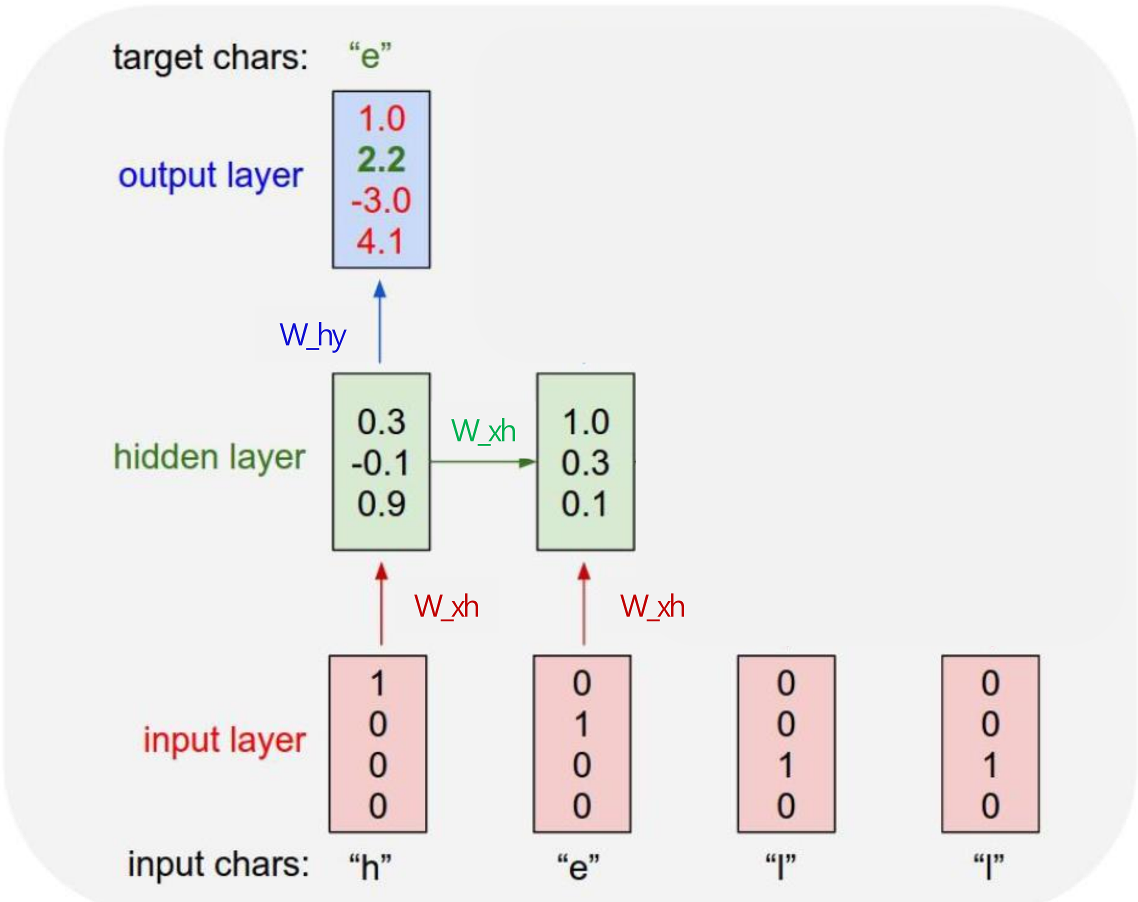
❖ Many to Many RNN for Text modeling



[- https://cs224d.stanford.edu/](https://cs224d.stanford.edu/)

순환 신경망 모델링

❖ Many to Many RNN for Text modeling

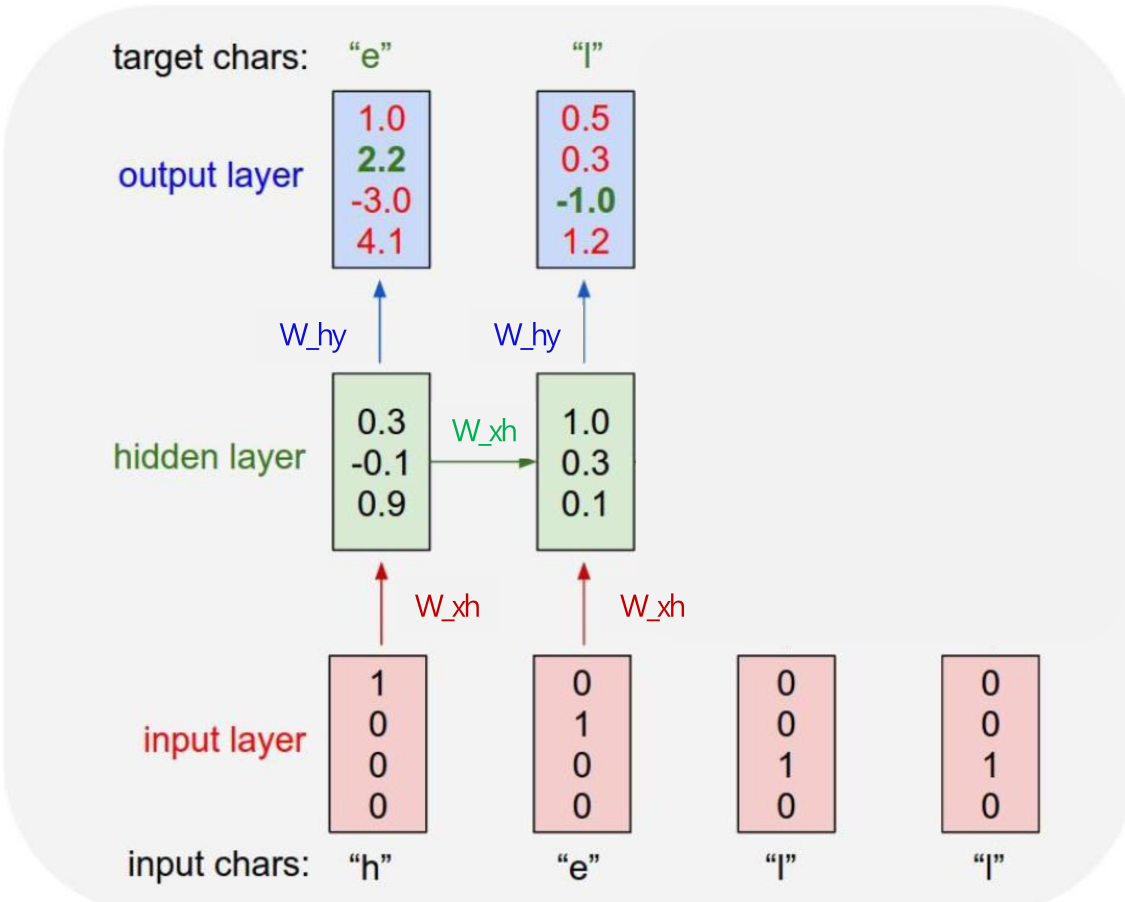


- https://cs224d.stanford.edu/

DMQA

순환 신경망 모델링

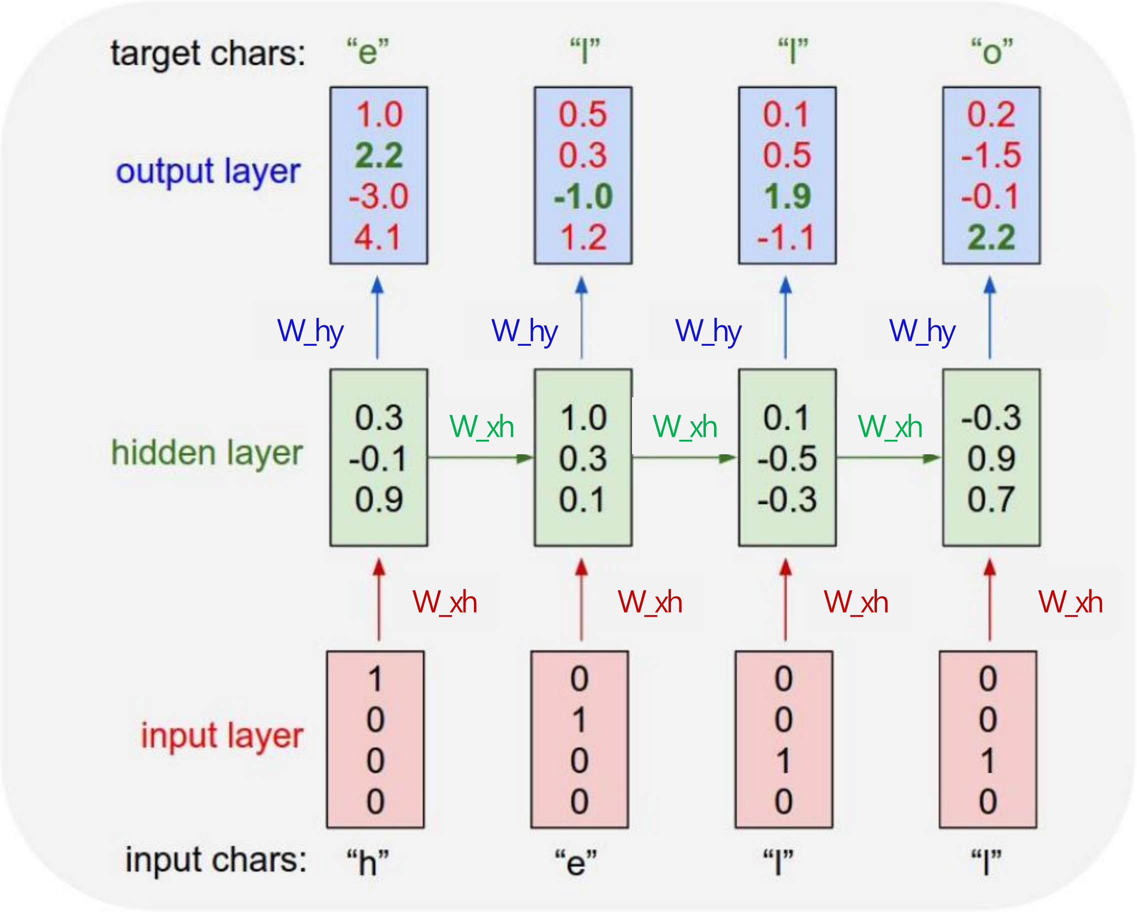
❖ Many to Many RNN for Text modeling



- https://cs224d.stanford.edu/

순환 신경망 모델링

❖ Many to Many RNN for Text modeling



- https://cs224d.stanford.edu/

목차

❖ 순환 신경망 개요

❖ 순환 신경망 학습(Backpropagation Trough Time)

❖ **순환 신경망 모델링**

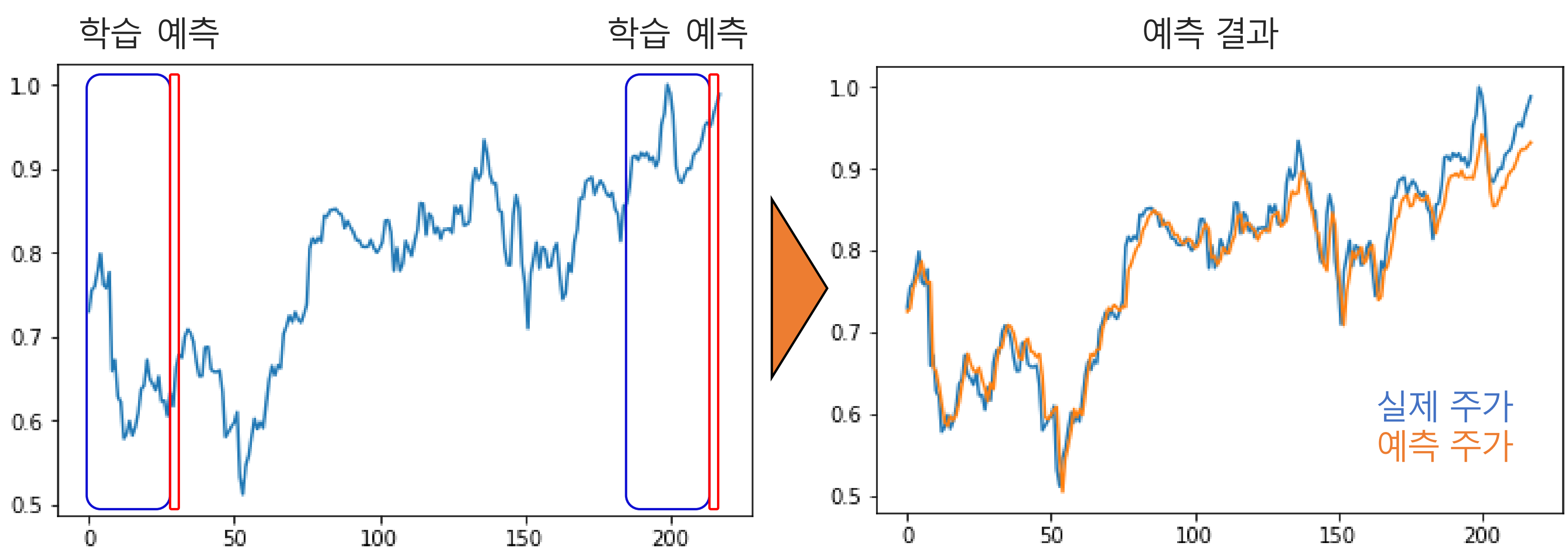
❖ 순환 신경망 한계

순환 신경망 모델링

❖ Many to Many RNN for Time series data modeling: Univariate

- ➢ 설명 변수 $(y_{t-10}, \cdots, y_{t-1})$: 주가 (이전 시점)
- ➢ 반응 변수 (y_t) : 주가 건수 (현재 or 예측 시점)
- ➢ 하이퍼파라미터(hyperparmameter): 주기 (time length) = 10

주가 예측

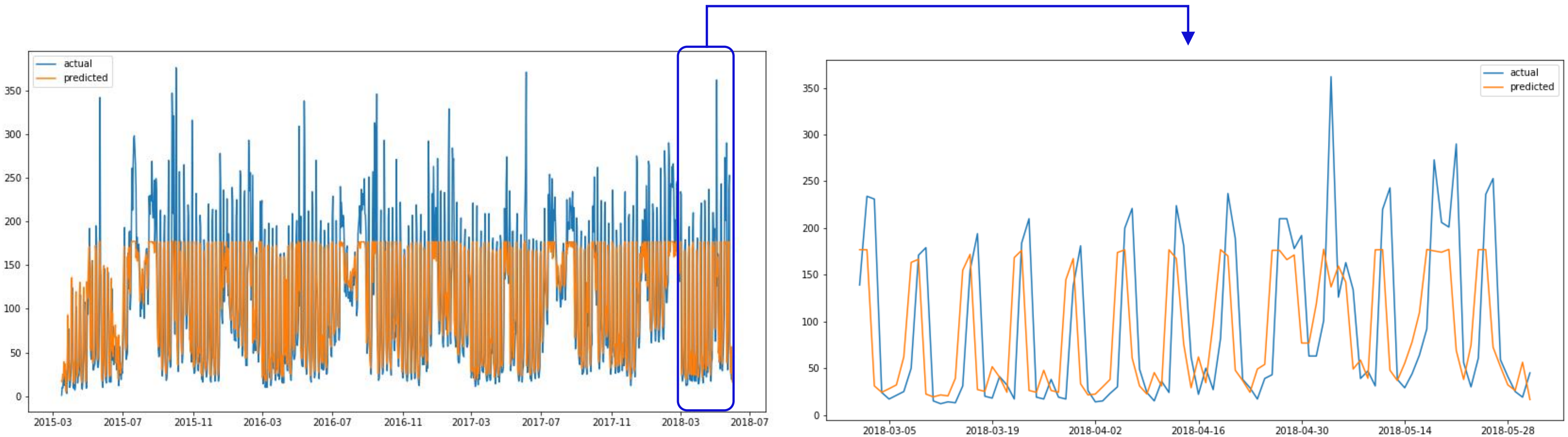


DMQA

순환 신경망 모델링

- Many to Many RNN for Time series data modeling: Univariate
  - 설명 변수 $(y_{t-10}, \cdots, y_{t-1})$: 리조트 취소 건수 (이전 시점)
  - 반응 변수 (y_t) : 리조트 취소 건수 (현재 or 예측 시점)
  - 하이퍼파라미터(hyperparmameter): 주기 (time length) = 10

한화 리조트 객실 취소 건수 예측(단변량)

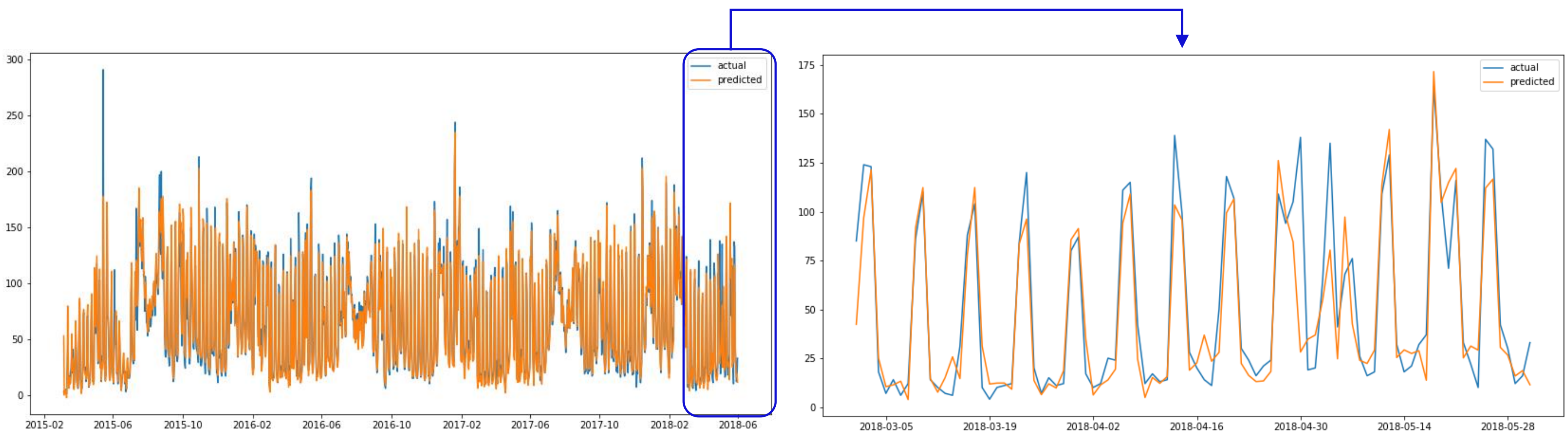


순환 신경망 모델링

❖ Many to Many RNN for Time series data modeling: Multivariate

- 설명 변수 $(x_{1,t}, \cdots, x_{i,t})$: 예약자 정보, 객실 정보(평수, 방 개수 등) 등
- 반응 변수 (y_t) : 리조트 취소 건수

한화 리조트 객실 취소 건수 예측(다변량)



목차

❖ 순환 신경망 개요

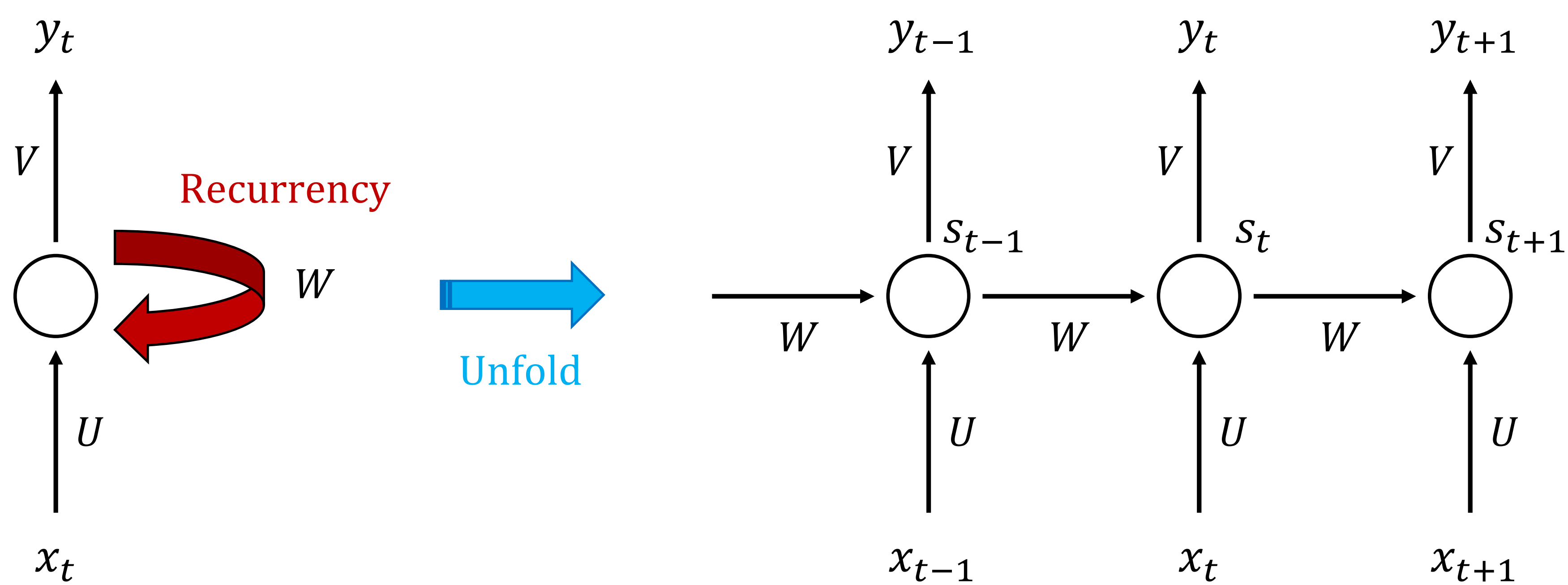
❖ 순환 신경망 학습(Backpropagation Trough Time)

❖ 순환 신경망 모델링

❖ **순환 신경망 한계**

순환 신경망 한계

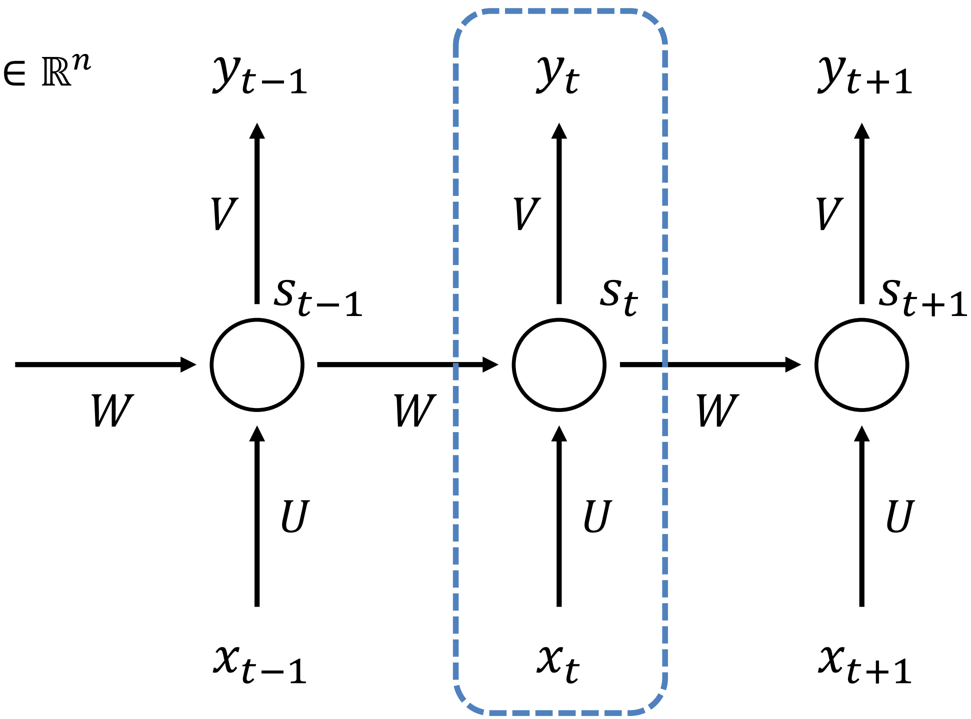
❖ RNN 용어



순환 신경망 한계

❖ RNN 용어: 약어 표시 예시

- x_t : $input\ at\ time\ step\ t \in \mathbb{R}^m$
- s_t : $hidden\ state\ at\ time\ step\ t \in \mathbb{R}^n$
- y_t : $output\ at\ time\ step\ t \in \mathbb{R}^l$
- $U \in \mathbb{R}^{n \times m}$
- $V \in \mathbb{R}^{l \times n}$
- $W \in \mathbb{R}^{n \times n}$
- $s_t = \tanh(Ux_t + Ws_{t-1})$
- $y_t = softmax(Vs_t)$
- $L(y, \hat{y}) = \sum_t L_t(y_t, \hat{y}_t) = -\sum_t y_t \ln \hat{y}_t$



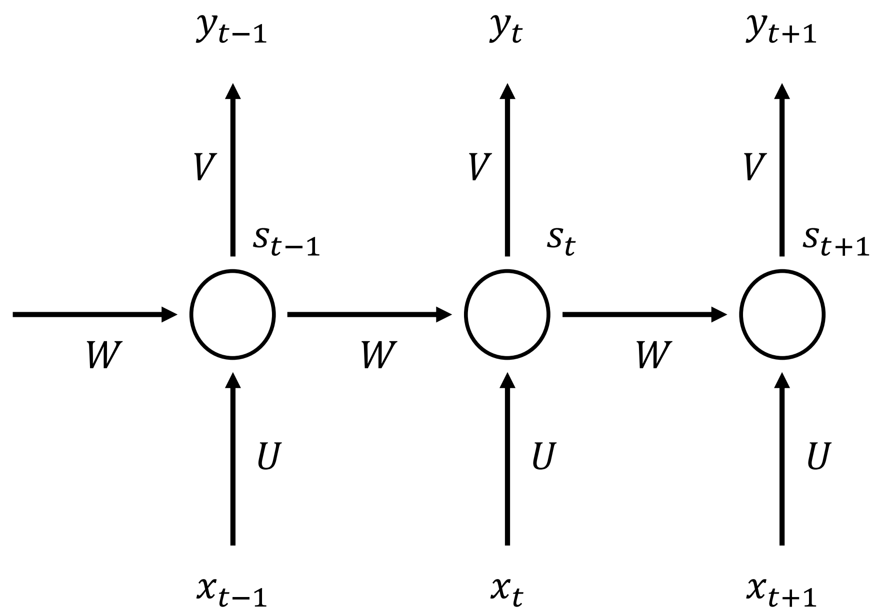
순환 신경망 한계

❖ RNN 용어: 약어 표시 예시

1) $\frac{\partial L_t}{\partial V} = \frac{\partial L_t}{\partial \hat{y}_t} \times \frac{\partial \hat{y}_t}{\partial V s_t} \times \frac{\partial V s_t}{\partial V}$

2) $\frac{\partial L_t}{\partial W} = \sum_{k=0}^{t} \left(\frac{\partial L}{\partial \hat{y}_t} \times \frac{\partial \hat{y}_t}{\partial s_t} \times \frac{\partial s_t}{\partial s_k} \times \frac{\partial s_k}{\partial W} \right)$

3) $\frac{\partial L_t}{\partial U} = \sum_{k=0}^{t} \left(\frac{\partial L}{\partial \hat{y}_t} \times \frac{\partial \hat{y}_t}{\partial s_t} \times \frac{\partial s_t}{\partial s_k} \times \frac{\partial s_k}{\partial U} \right)$



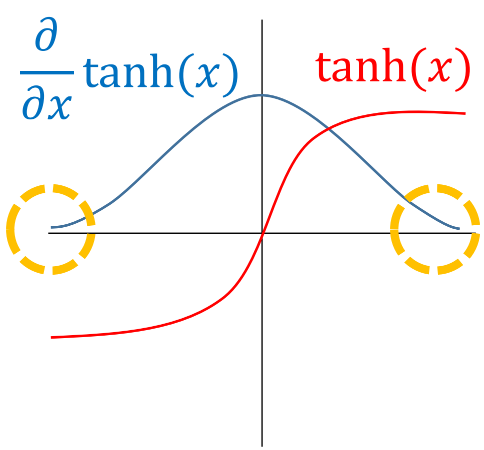
순환 신경망 한계

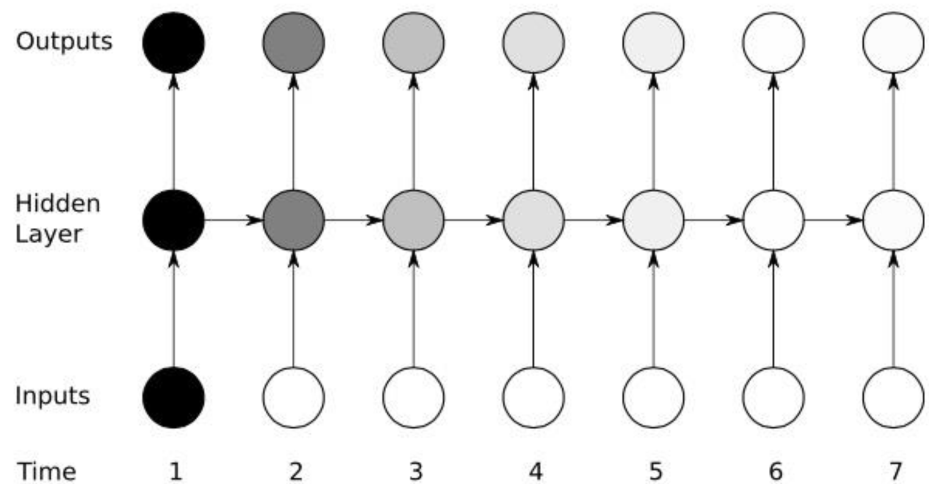
❖ 길이기 긴 sequence 의 경우 Gradient Vanishing / Exploding 문제 발생

- Vanishing gradient: 다른 모델 사용 →(LSTM, GRU 등)
- Exploding gradient: clip gradient

$$\frac{\partial \mathcal{E}_3}{\partial \hat{y}_3}\frac{\partial \hat{y}_3}{\partial s_3}\left(\frac{\partial s_3}{\partial W}+\frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial W}+\frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial W}+\frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial s_0}\frac{\partial s_0}{\partial W}\right)$$

Long-Term Dependency problem





DMQA

EOD