

딥러닝의 통계적이해

12강. 순환신경망 (2)

1. LSTM/GRU
2. 순환신경망의 활용

SK텔레콤 김기온

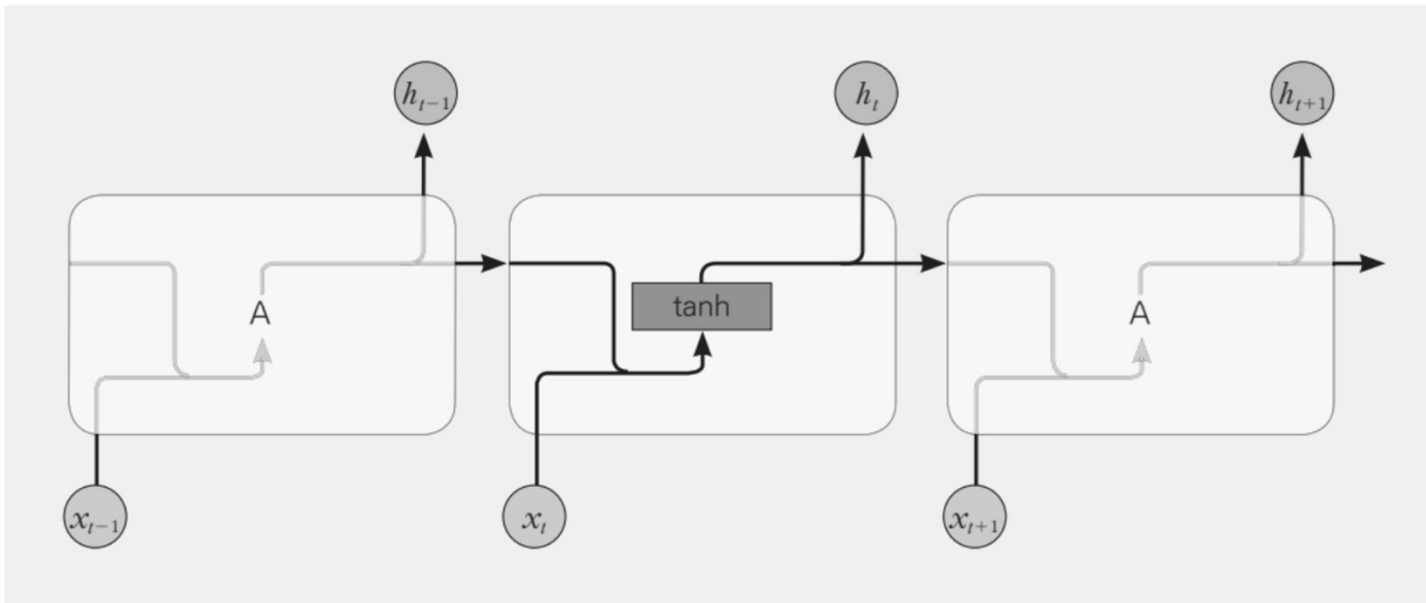
오늘의 학습목표

1. LSTM과 GRU의 구조를 이해한다.
2. 순환신경망의 다양한 활용방법을 이해한다.

1. LSTM/GRU

1. LSTM/GRU

Simple RNN review



Gradient Vanishing / Gradient Exploding

1. LSTM/GRU

LSTM

Gates	$i_t = \sigma(b_t + W_i h_{t-1} + U_i x_t)$	Input Gate
	$f_t = \sigma(b_f + W_f h_{t-1} + U_f x_t)$	Forget Gate
	$o_t = \sigma(b_o + W_o h_{t-1} + U_o x_t)$	Output Gate

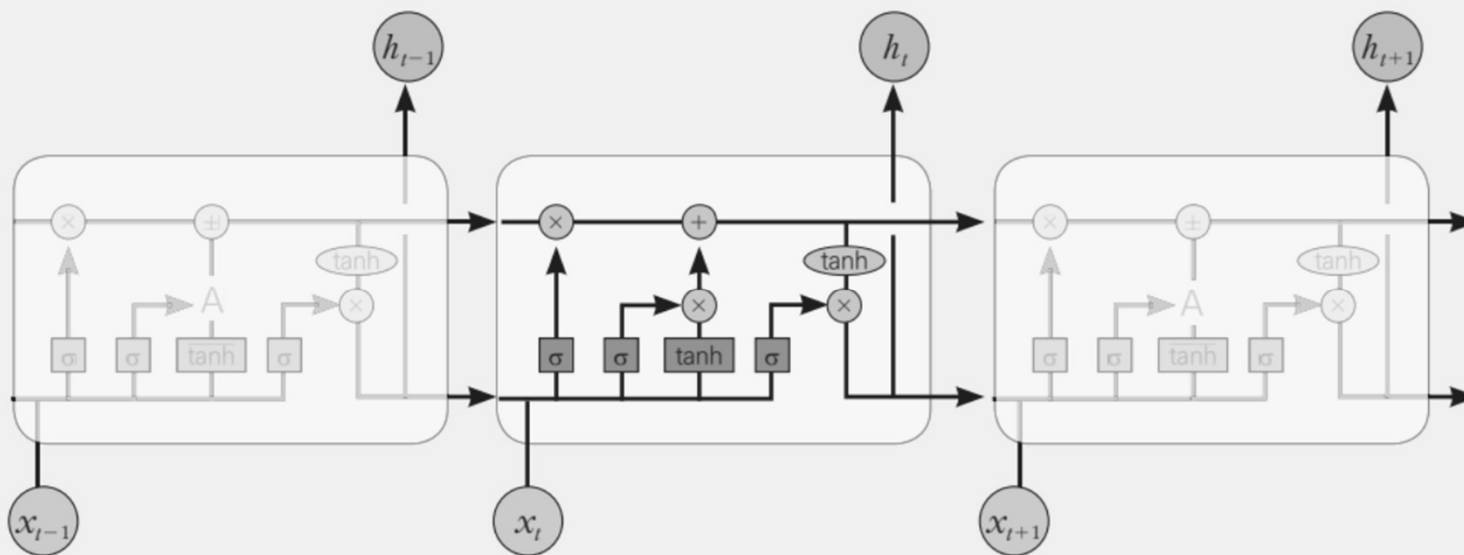
$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$	Cell State
$\tilde{C}_t = \tanh(b_c + W_c h_{t-1} + U_c x_t)$	

$h_t = o_t \odot \tanh(C_t)$	Hidden State
------------------------------	--------------

1. LSTM/GRU

LSTM

$$\begin{aligned} i_t &= \sigma(b_t + W_i h_{t-1} + U_i x_t) & \tilde{C}_t &= \tanh(b_C + W_C h_{t-1} + U_C x_t) \\ f_t &= \sigma(b_f + W_f h_{t-1} + U_f x_t) & C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(b_o + W_o h_{t-1} + U_o x_t) & h_t &= o_t \odot \tanh(C_t) \end{aligned}$$



1. LSTM/GRU

LSTM Gradient

$$\left. \frac{\partial E_t}{\partial W} \right|_{step=1} = \frac{\partial E_t}{\partial o_t} \cdot \frac{\partial o_t}{\partial C_t} \cdot \left(\prod_{k=2}^t \underbrace{\frac{\partial C_k}{\partial C_{k-1}}}_{(1)} \right) \cdot \frac{\partial C_1}{\partial W}$$

$$\frac{\partial C_k}{\partial C_{k-1}} = \frac{\partial}{\partial C_{k-1}} (i_t \odot \tilde{C}_{k-1}) + \text{diag}(f_t)$$

$$\frac{\partial C_k}{\partial C_{k-1}} \approx \text{diag}[\sigma(W_f + U_f x_k)]$$

1. LSTM/GRU

GRU

Gates

$$\begin{aligned} z_t &= \sigma(b_z + W_z h_{t-1} + U_z x_t) && \text{Update Gate} \\ r_t &= \sigma(b_r + W_r h_{t-1} + U_r x_t) && \text{Reset Gate} \end{aligned}$$

Hidden
State

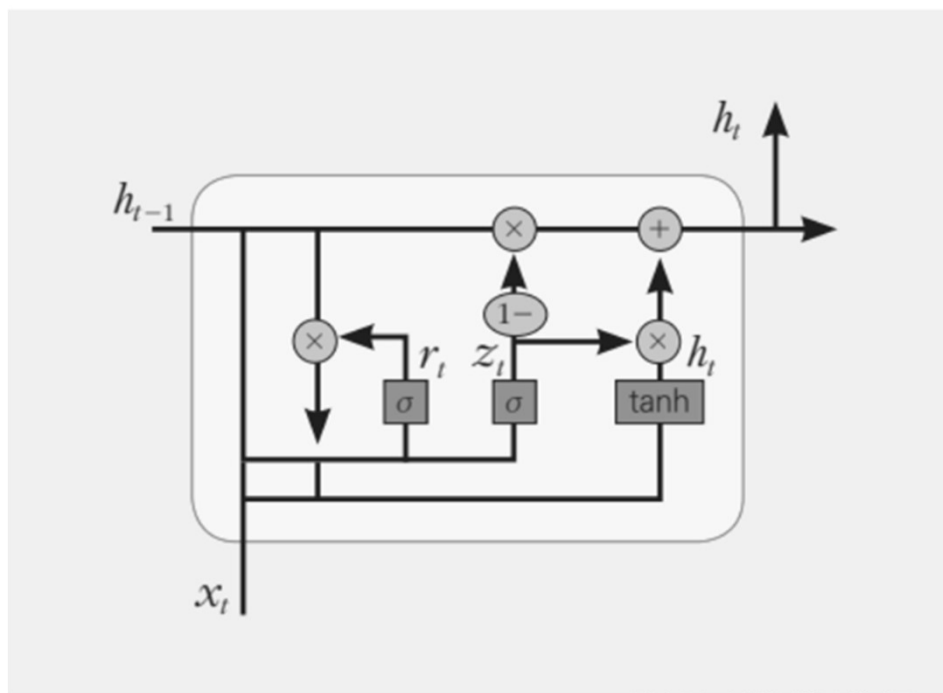
$$\begin{aligned} \tilde{h}_t &= \tanh(b_h + W_h(r_t \odot h_{t-1}) + U_h x_t) \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \end{aligned}$$

$$o_t = \sigma_v(b_o + V h_t)$$

1. LSTM/GRU

GRU

$$\begin{aligned} z_t &= \sigma(b_z + W_z h_{t-1} + U_z x_t) & \tilde{h}_t &= \tanh(b_h + W_h(r_t \odot h_{t-1}) + U_h x_t) \\ r_t &= \sigma(b_r + W_r h_{t-1} + U_r x_t) & h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \\ & & o_t &= \sigma_v(b_o + V h_t) \end{aligned}$$



1. LSTM/GRU

GRU Gradient

$$\left. \frac{\partial E_t}{\partial W} \right|_{step=} = \frac{\partial E_t}{\partial o_t} \cdot \frac{\partial o_t}{\partial h_t} \cdot \left(\prod_{k=2}^t \underbrace{\frac{\partial h_k}{\partial h_{k-1}}}_{(1)} \right) \cdot \frac{\partial h_1}{\partial W}$$

$z_k = 1$: Update,
 $z_k = 0$: Do not Update

$r_k = 1$: Do not reset
 $r_k = 0$: Reset

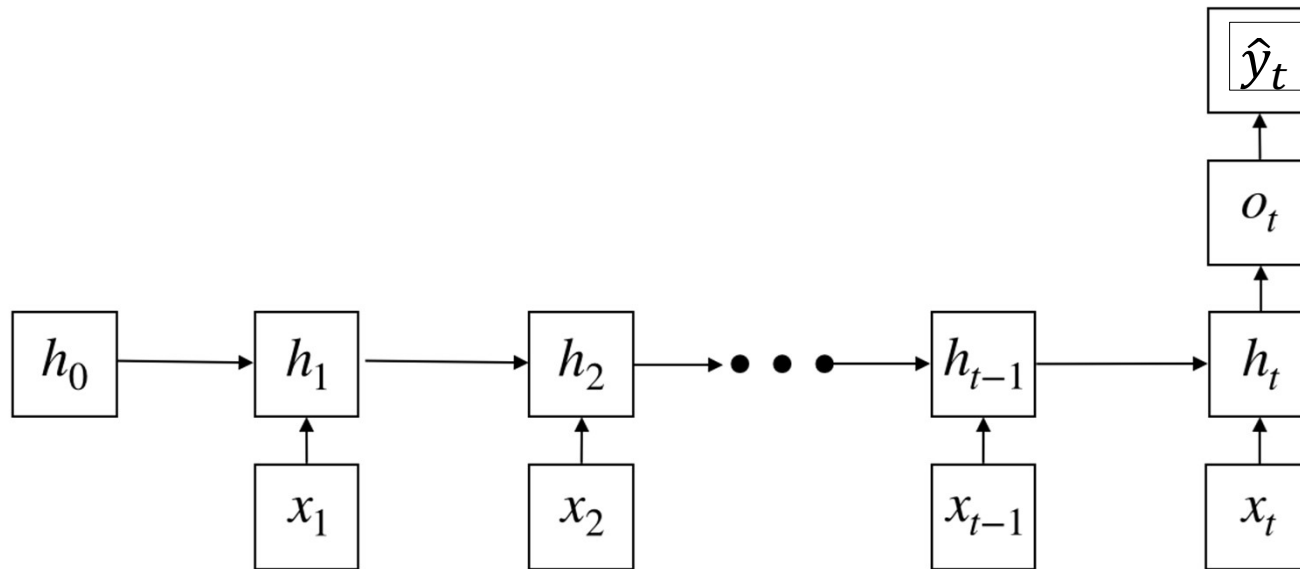
$$\frac{\partial h_k}{\partial h_{k-1}} = \text{diag}(1 - z_k) + z_k \frac{\partial \widetilde{h}_k}{\partial h_{k-1}}$$

$$\frac{\partial \widetilde{h}_k}{\partial h_{k-1}} = \frac{\partial}{\partial h_{k-1}} \tanh(b_h + W_h(r_k \odot h_{k-1}) + U_h x_t) = (1 - \widetilde{h}_k^2)(r_k \odot W_h)$$

2. 순환신경망의 활용

2. 순환신경망의 활용

Many-to-one



2. 순환신경망의 활용

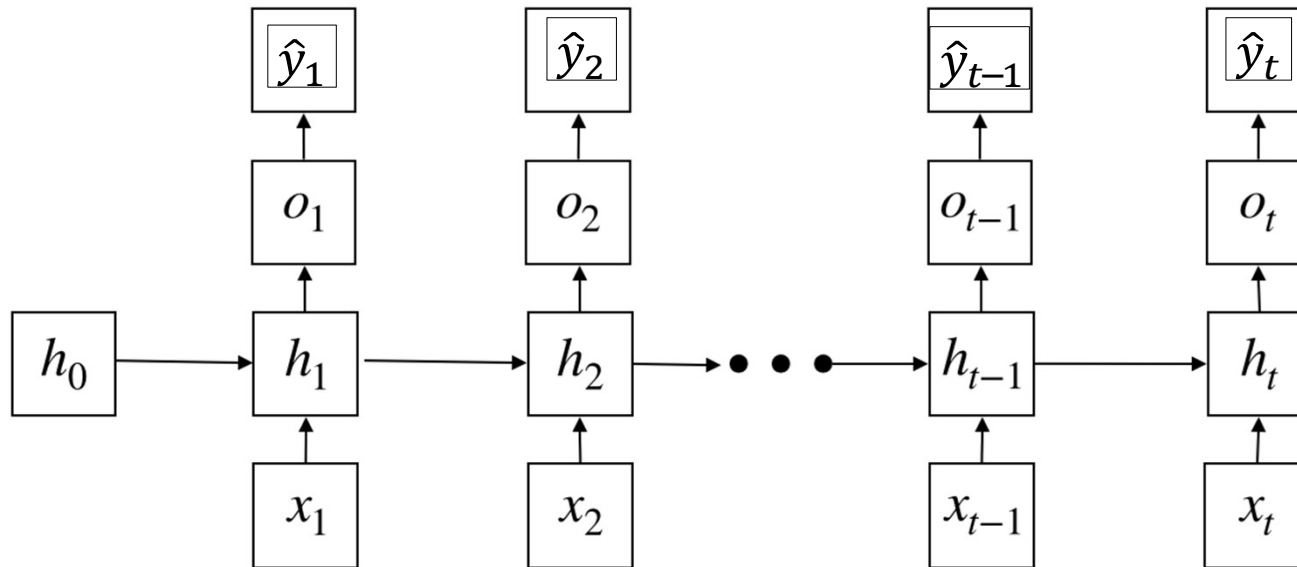
감성분석

출처: <https://wikidocs.net/44249>

	id	document	label
0	9976970	아 더빙.. 진짜 짜증나네요 목소리	0
1	3819312	흠...포스터보고 초딩영화줄....오버연기조차 가볍지 않구나	1
2	10265843	너무재밌었다그래서보는것을추천한다	0
3	9045019	교도소 이야기구먼 ..솔직히 재미는 없다..평점 조정	0
4	6483659	사이몬페그의 익살스런 연기가 돋보였던 영화!스파이더맨에서 늙어보이기만 했던 커스틴 ...	1

2. 순환신경망의 활용

Many-to-Many



2. 순환신경망의 활용

Entity Recognition

<증권거래소:OG>에 주식을 상장하였다

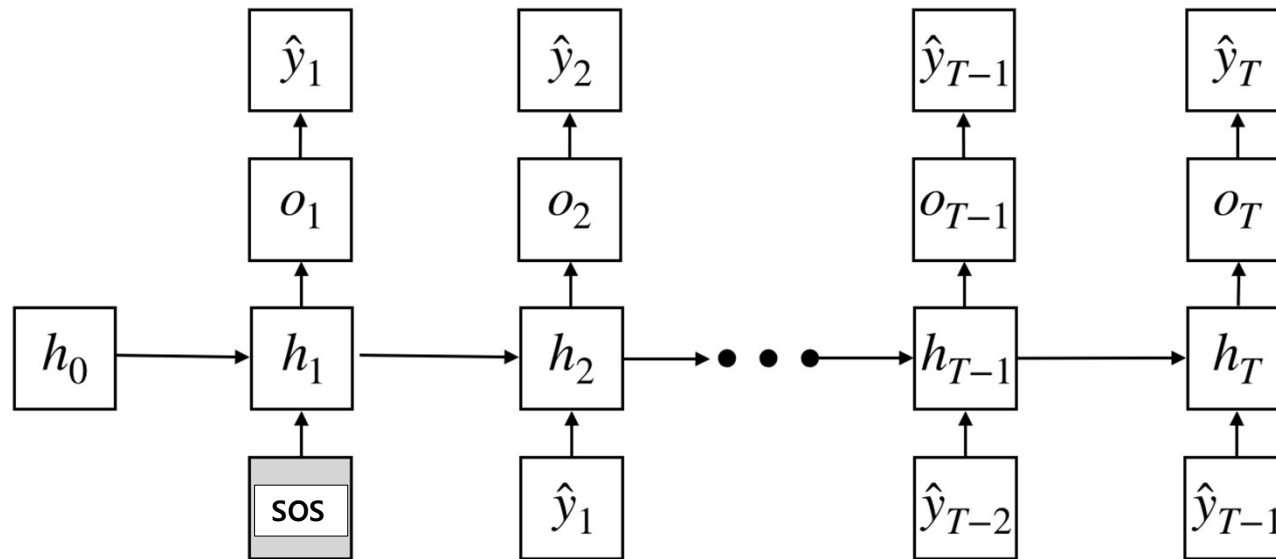


증	##권	##거	##래	##소	##에	주	##식	##을	상	##장	##하	##였	##다
B-OG	I-OG	I-OG	I-OG	I-OG	0	0	0	0	0	0	0	0	0

출처; <http://blog.naver.com/PostView.nhn?blogId=jeanmy1102&logNo=221735304210&categoryNo=204&parentCategoryNo=0>

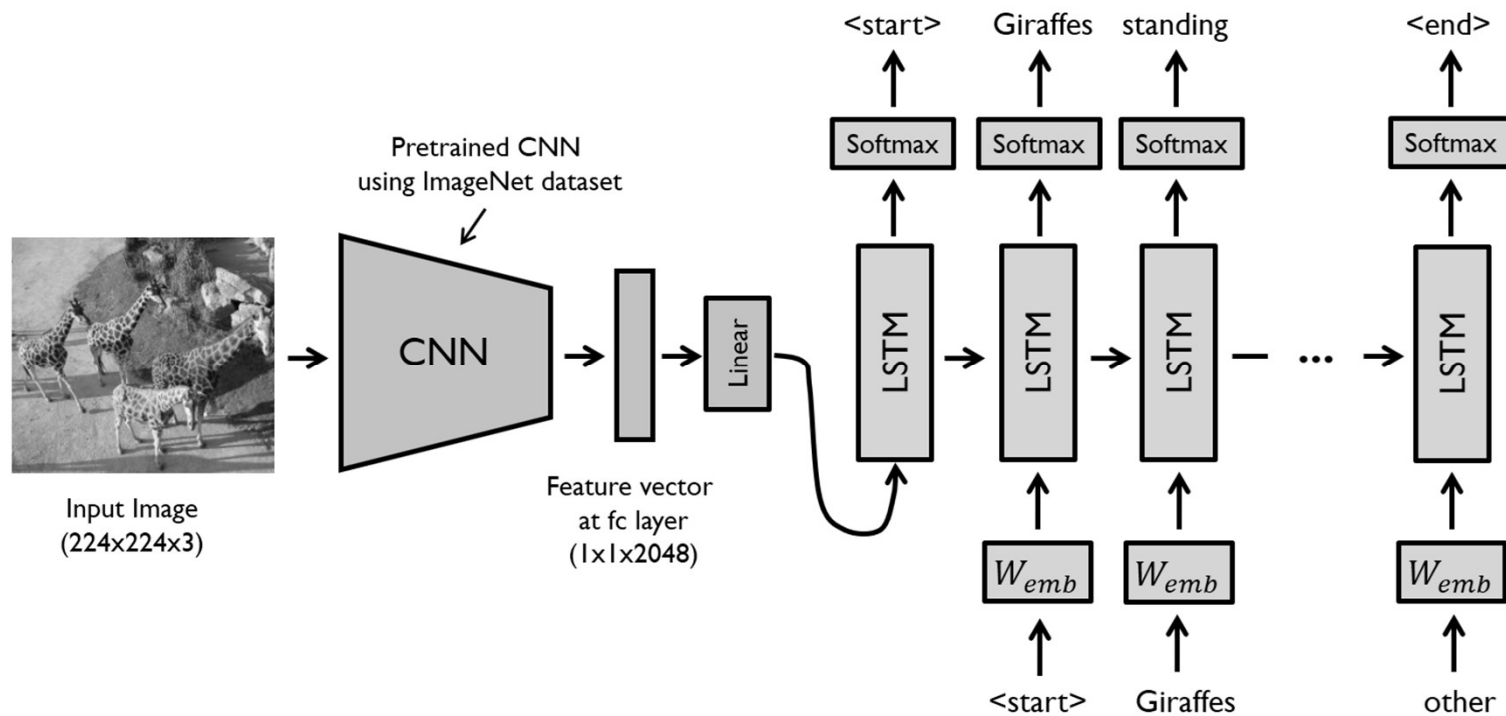
2. 순환신경망의 활용

One-to-Many



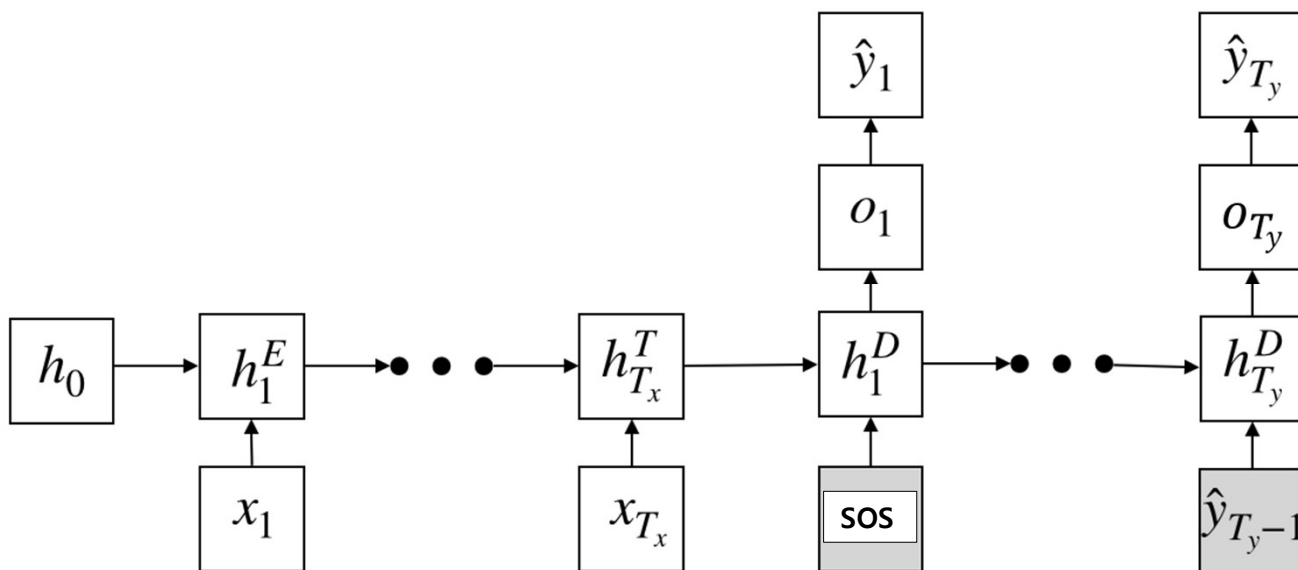
2. 순환신경망의 활용

이미지 캡션



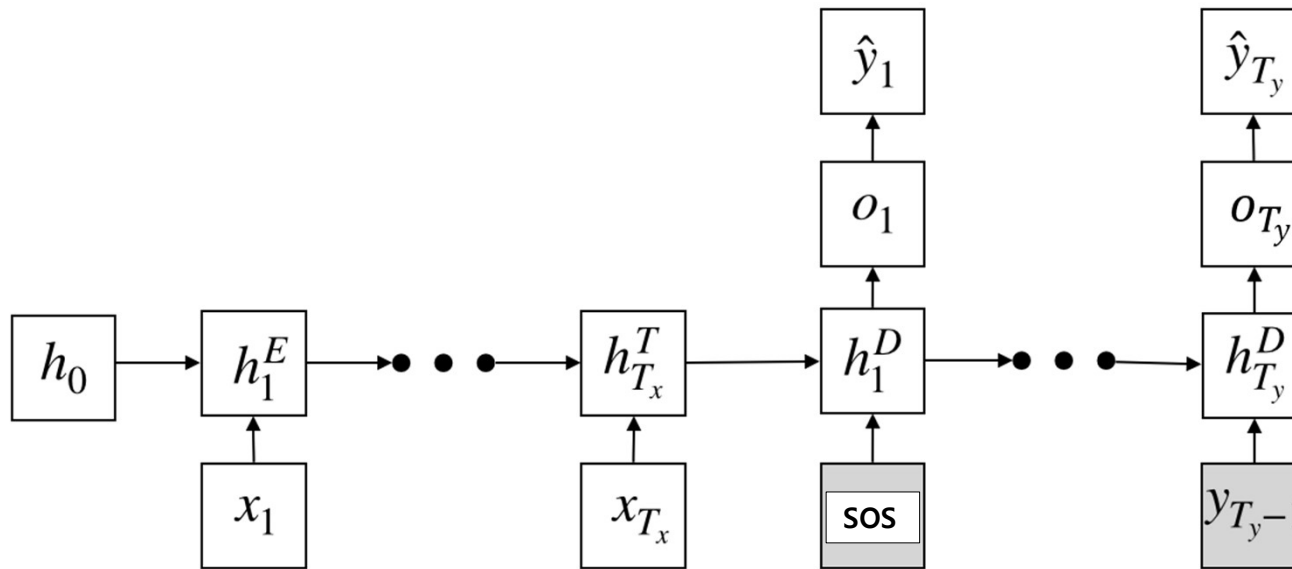
2. 순환신경망의 활용

Seq2Seq (Sequence-to-sequence)



2. 순환신경망의 활용

Seq2Seq



Teacher forcing

학습정리

- ✓ 경사소실/경사폭발의 문제를 해결하기 위해 LSTM과 GRU가 제안되었다.
- ✓ RNN은 손실함수를 다양하게 구성함으로써, 감성분석, 개체명 인식, 기계번역 등 다양한 자연어 이해 (NLU) 문제에 활용할 수 있다.

딥러닝의 통계적 이해
다음시간안내

13강. 딥러닝 모형을 이용한 자연어 처리 (1)