

## 제6장: 세가지기본신경망III: RNN

**Soyoung Park**

Pusan National University  
Department of Statistics

**`soyoung@pusan.ac.kr`**

# Recurrent Neural Network

- RNN은 순서정보가 분류나 회귀 등의 분석에 중요한 요인이 될 때 사용하는 신경망모형
- ex) 신문기사의 주제가 경제, 문화, 사회 등의 분류, 트윗이 특정 주제에 대해 찬성인지 반대인지를 구별 하는 등
- ex) 한언어에서 다른 언어로의 번역, 내일의 주식가격 예측 등
- RNN의 입력자료는 (표본수, 시간스텝, 특성변수 수)로 3D텐서

# 1분단위 주식가격 예측문제

# Simple RNN architecture

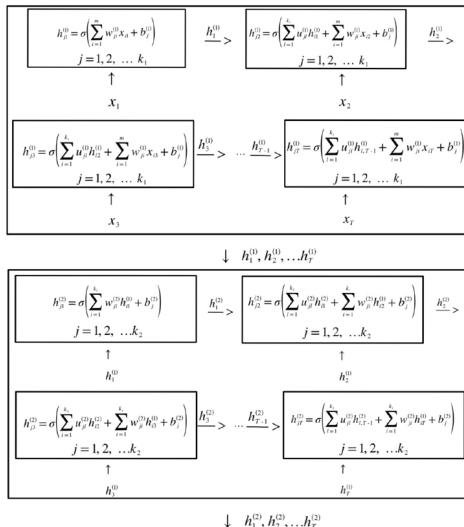


그림 2-15 간단한 RNN의 개념도

# Recurrent Neural Network

# RNN vs. MLP

# RNN vs. CNN

# Simple RNN의 구조



# Simple RNN의 구조

# Simple RNN의 구조

# RNN의 도식화

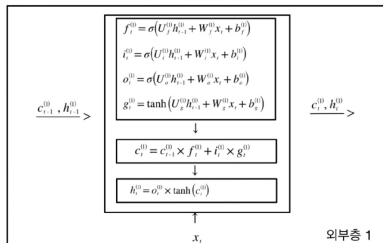
# Simple RNN의 모수

# Simple RNN의 모수

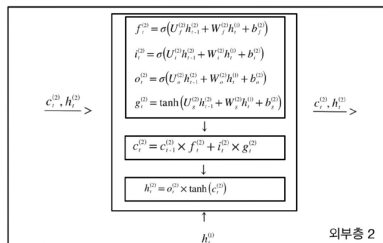
# Simple RNN의 모수

# Simple RNN의 약점

# LSTM 모형



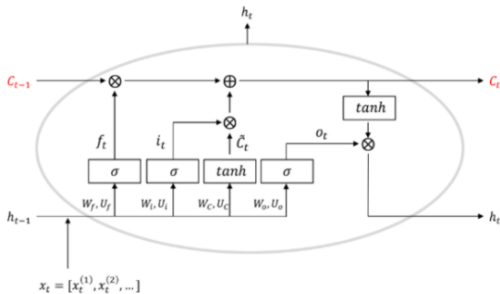
↓  $(h_1^{(0)} h_2^{(0)} \dots h_T^{(0)}), h_T^{(0)}, c_T^{(0)}$



↓  $(h_1^{(2)} h_2^{(2)} \dots h_T^{(2)}), h_T^{(2)}, c_T^{(2)}$



# LSTM 모형



$$(1) f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f)$$

$$(2) i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i)$$

$$(3) \tilde{C}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c)$$

$$(4) C_t = f_t \otimes C_{t-1} \oplus i_t \otimes \tilde{C}_t$$

$$(5) o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o)$$

$$(6) h_t = o_t \otimes \tanh(C_t)$$

(f: forget i: ignore o: output  $\sigma$ : sigmoid)

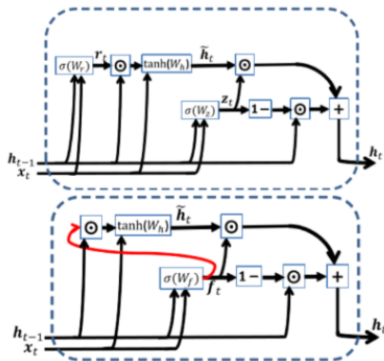
# LSTM 모형

# LSTM 모형

# LSTM과 GRU 모형

- RNN의 gradient vanishing/ exploding 문제를 해결하는 모형
- $h_t$ 와  $h_{t-1}$ 을 간접적으로 연결하는 모형
- GRU는 LSTM을 간소화한 모형
- LSTM과 GRU는 간단한 RNN 구조를 가지고 있으나 내부층을 구성하는  $h_t$ 가 좀 더 복잡하게 구성되어 있음.

## GRU 모형



Gated Recurrent Unit (GRU)

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r)$$

$$\tilde{h}_t = \tanh(W_h[r_t \odot h_{t-1}, x_t] + b_h)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

Minimal Gated Unit (MGU)

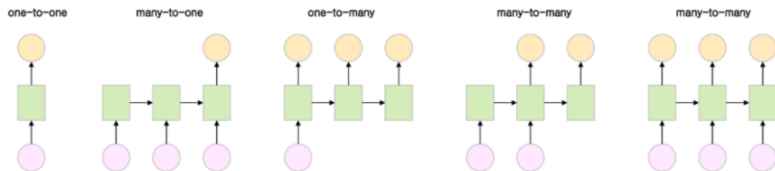
$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

$$\tilde{h}_t = \tanh(W_h[f_t \odot h_{t-1}, x_t] + b_h)$$

$$h_t = (1 - f_t) \odot h_{t-1} + f_t \odot \tilde{h}_t$$

# RNN의 학습유형

- RNN은 분석 목적에 따라 아래와 같이 여러 가지 유형으로 학습
- RNN의 유형은 입력값에 대응하는 출력값의 개수에 따라 아래 4가지 유형으로 분류
- 예시) Classification 목적을 위해서는 many-to-one 유형, 시계열 분석은 many-to-one이나 many-to-many 유형, 이미지를 말로 설명하는 image captioning 기능은 one-to-many 유형, 자동번역기 (machine translation)는 (아래 3번째 유형의) many-to-many 유형으로 학습



# Summary

- 세개의 신경망에 있는 모수는 표본에 의존하지 않는다.
  - 추가적으로 RNN의 모수는 시간스텝에 의존하지 않음
  - 표본은 서로간에 독립
  - RNN의 시간스텝은 최소한 정상성 조건을 만족
- 노드수는 은닉층에서 만들어진 특성변수의 수
- 은닉층에서 만들어진 특성변수들은 학습에 의해 만들어지며 해석이 어려움