

RNN 입문

<https://youtu.be/caQNkz8A09I>

RNN 개요

RNN 구조

LSTM 구조

RNN 개요

RNN(Recurrent Neural Network)

순환 신경망

- 순서가 있는 정보를 입력 데이터로 사용.

기존 뉴런 네트워크와의 차이점

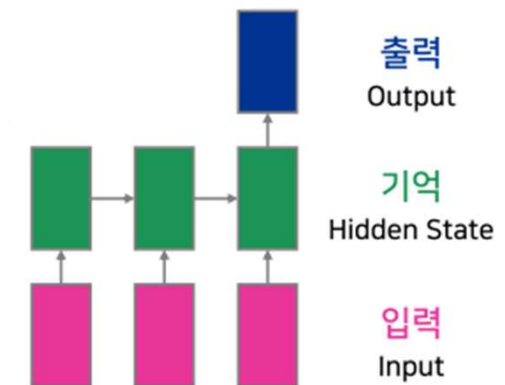
- 새로운 입력으로 이전의 기억을 수정해 새로운 기억을 생성
- 출력은 이전의 모든 기억과 새로운 입력으로 생성
- 과거의 데이터가 미래에 영향을 줄 수 있는 구조

미래 데이터에는 영향을 받지 않음

- bidirectional RNN은 양방향 구조로 확장

앞에 처리된 결과가 뒤에 영향을 미침

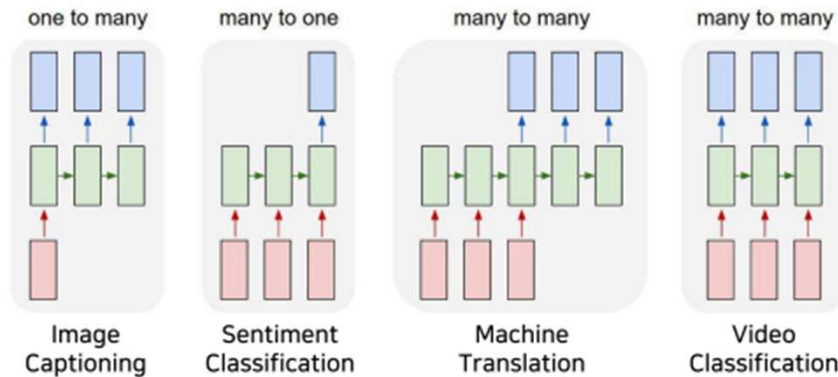
- 현재까지 누적된 입력이 현재 출력 결정



RNN 개요

RNN Applications

RNN에는 여러가지 모델 존재.



On-to-many : 이미지에 대해 설명을 붙일 때 사용
ex) 한 장의 그림에 대해 여러 개의 단어 형태로 표현

Many-to-one : 여러 개의 입력에 대해 하나의 결과
ex) 말을 통해 심리 상태를 한 단어로 결과를 예측할 때

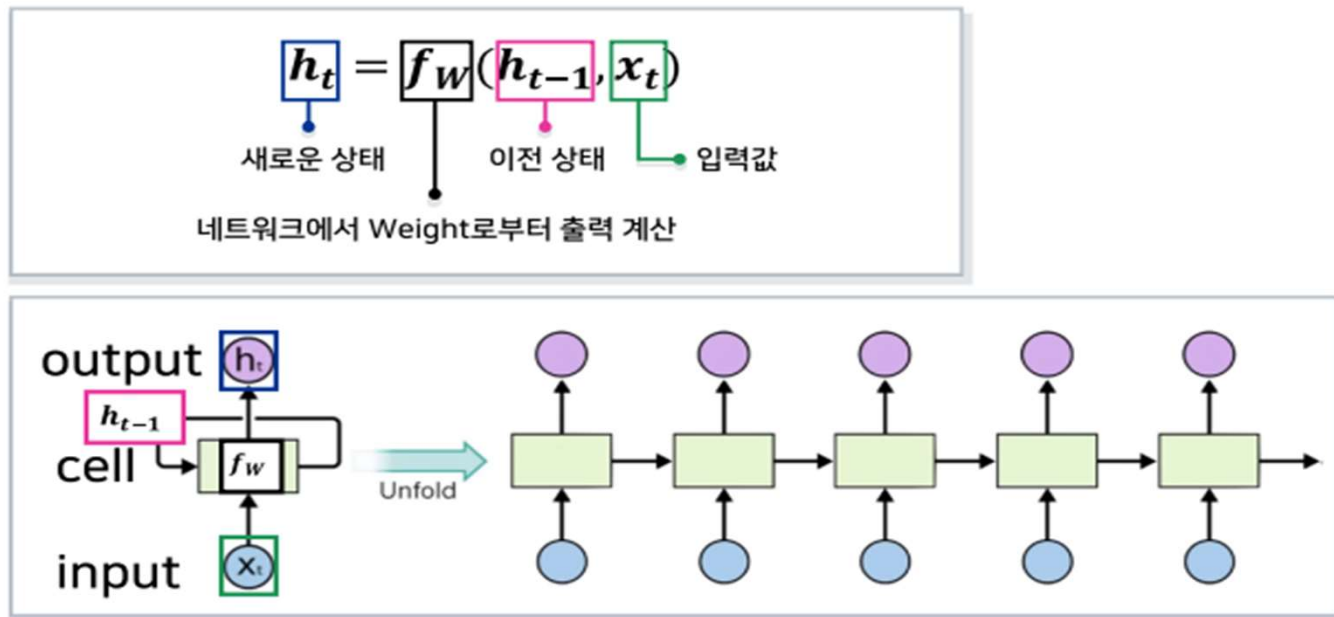
Many-to-many(1) : 기계 번역에서 주로 사용
ex) 여러 개의 단어로 구성된 문장을 받아 여러 개의 단어로 구성된 문장을 반환

Many-to-many(2) : 위의 모델의 또 다른 형태
ex) 동영상의 여러 개의 이미지 프레임에 대해 설명이 나 번역 형태로 결과를 반환

RNN 구조

RNN(Recurrent Neural Network)

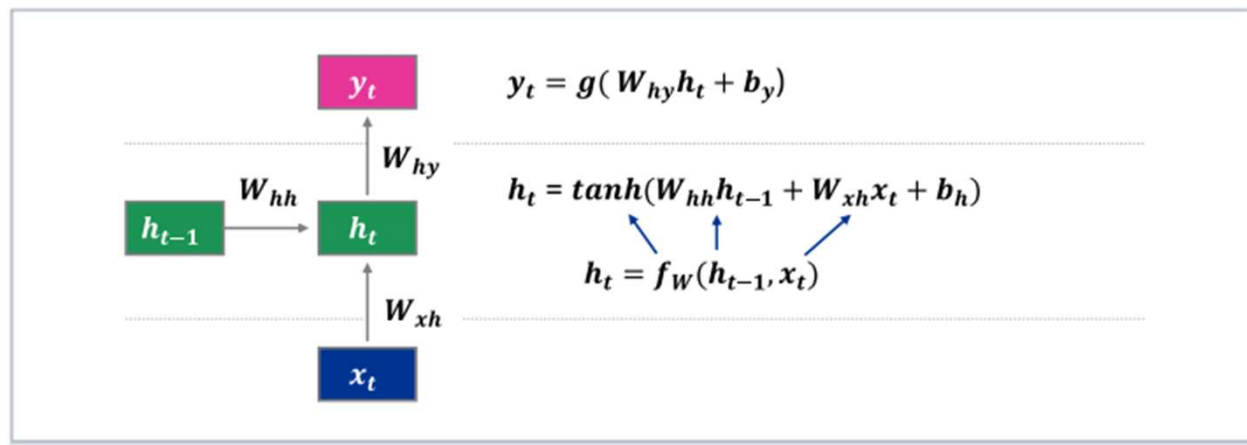
- 이전의 상태값이 결과에 영향
- RNN에서는 activation 함수를 tanh에 사용



RNN 구조

RNN의 Data Flow

- 3개의 서로 다른 Weight 존재
- Y의 값이 몇 개의 벡터로 나올 것인가는 W_{hy} 의 shape이 결정
- $H(t)$ 의 출력값의 개수는 각각의 weight의 shape에 의해 결정
- 모든 셀이 동일한 weight를 가짐



RNN 구조

Tanh function

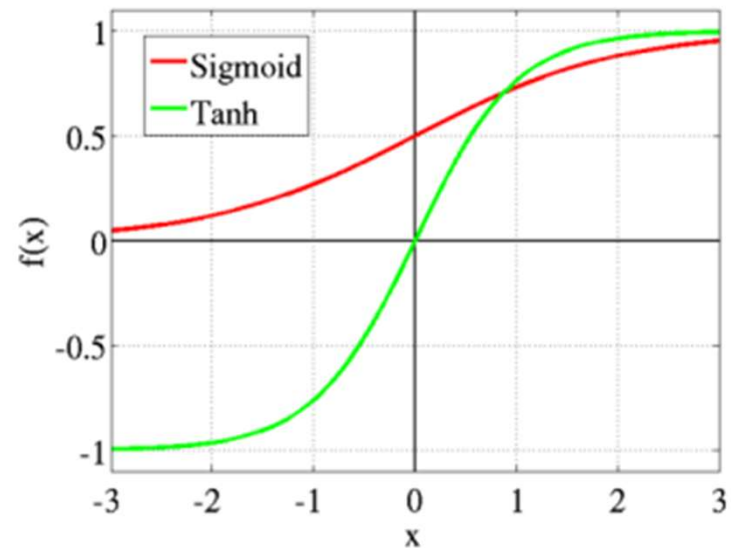
Sigmoid는 0~1사이의 값 존재

- 경사면이 작아 느리게 수렴하여 학습

Tanh는 -1~1사이의 값 존재

- 경사면이 커 빠르게 수렴하여 학습

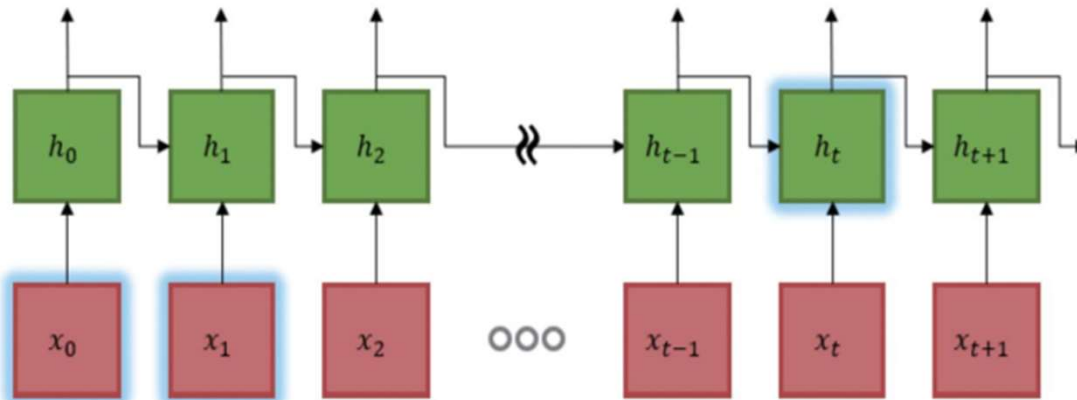
Sigmoid 보다 Tanh 함수가 속도와 연산량이 적어
activation function으로 사용



LSTM 구조

RNN의 장기 의존성 문제

- 입력 데이터가 길어질수록 학습 능력이 떨어지는 문제.



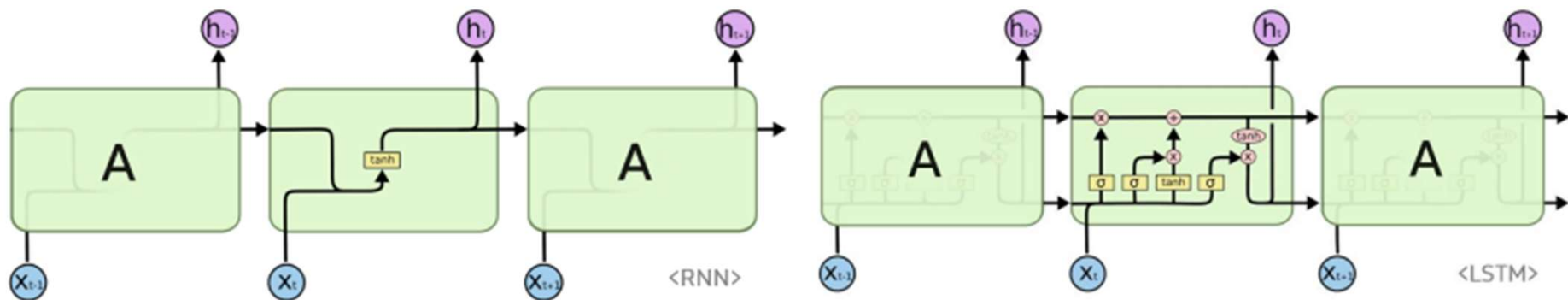
RNN은 관련 정보와 지점사이 거리가 멀 경우 gradient가 감소

→ LSTM : RNN의 Vanishing Gradient Problem을 해결하기 위해 등장.

LSTM 구조

LSTM : RNN의 Vanishing Gradient Problem 해결하기 위한 모델

- RNN의 hidden state에 Cell-state를 추가한 구조
- gate라는 요소를 활용하여 정보를 더하거나 제거
- gate는 선택적으로 정보들이 흘러들어갈 수 있도록 만든 장치



LSTM 구조

LSTM 구조

- Forget gate : f_t 에 의해 제어되며 c_{t-1} 의 어느 부분을 삭제할지 제어
 - 과거 정보를 잊기 위한 게이트
- input gate : i_t 에 의해 제어되며 g_t 의 어느 부분이 더해지는지 제어
 - 현재 정보를 기억하기 위한 게이트
- output gate : o_t 에 의해 제어되며 c_t 의 어느 부분을 h_t, y_t 로 출력하는지 제어

$$i_t = \text{sigmoid}(x_t U^i + h_{t-1} W^i)$$

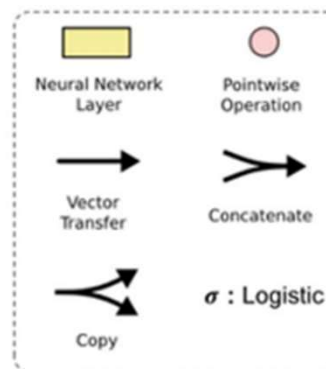
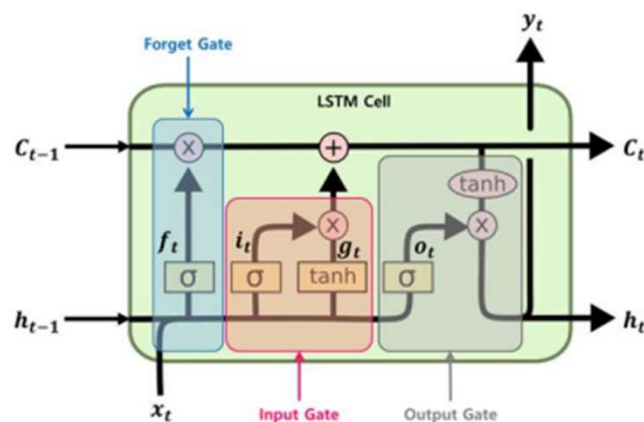
$$f_t = \text{sigmoid}(x_t U^f + h_{t-1} W^f)$$

$$o_t = \text{sigmoid}(x_t U^o + h_{t-1} W^o)$$

$$g_t = \tanh(x_t U^{\tilde{c}} + h_{t-1} W^{\tilde{c}})$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t$$

$$h_t = \tanh(c_t) \times o_t$$



Q & A