

<https://arxiv.org/abs/1412.3555>

GRU (Gated Recurrent Unit)

: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling

KUBIG 15기 남정재

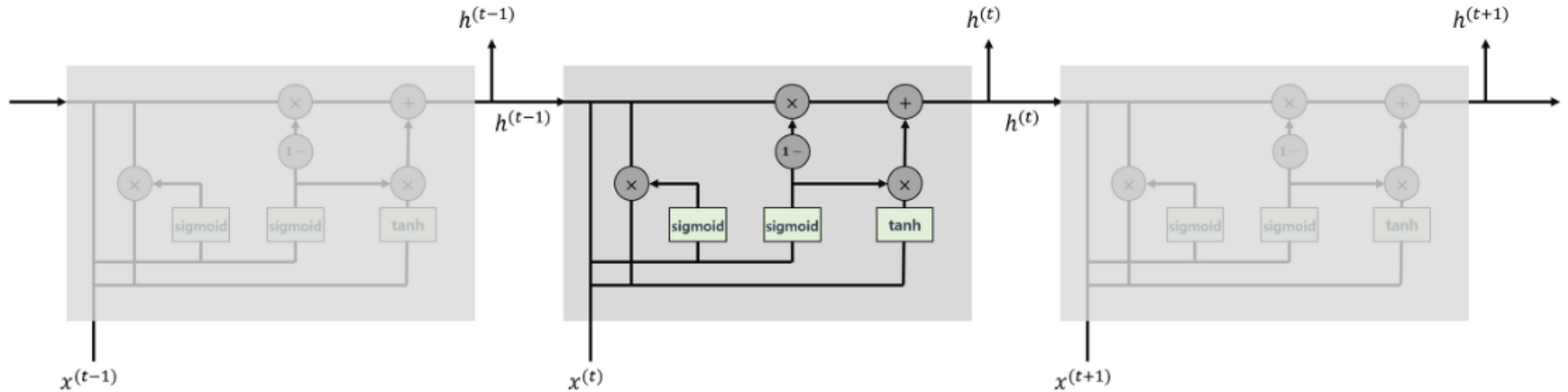
GRU

간단히 말해서...

기존 LSTM의 구조를 조금 더 간단하게 개선한 모델!
cf.) 2014년 한국인 조경현 박사님이 제안했다고 함!

>> LSTM이 RNN의 Long-term dependency 즉, 장기 의존성 문제를 해결하여 긴 시퀀스를 가진 데이터에서도 좋은 성능을 낼 수 있게 되었지만, 복잡한 구조 때문에 RNN에 비하여 파라미터가 많이 필요하게 됨

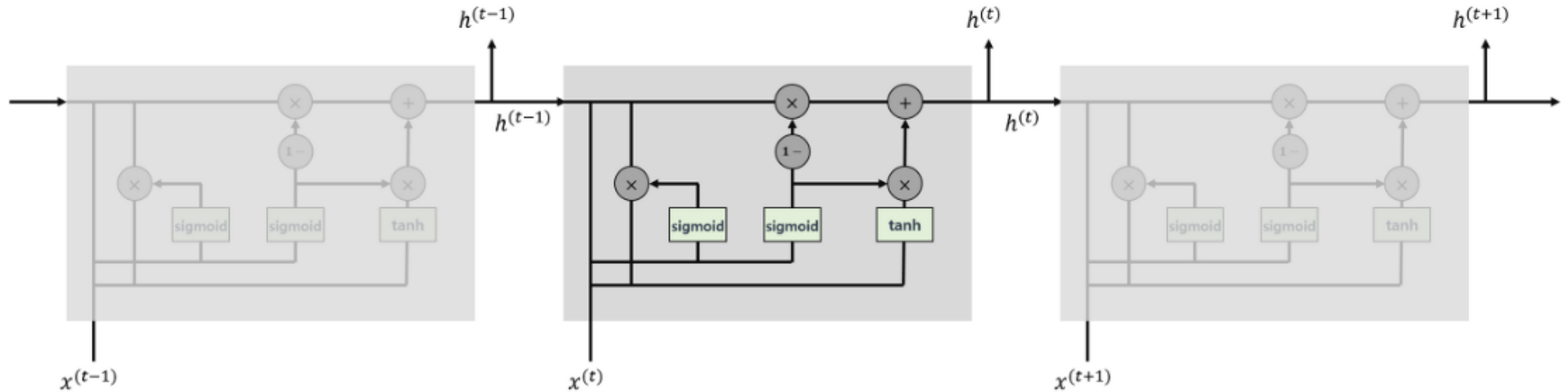
>> 이에 대하여 hidden state를 업데이트하는 계산 즉, GRU가 학습할 가중치가 적게 만들었음



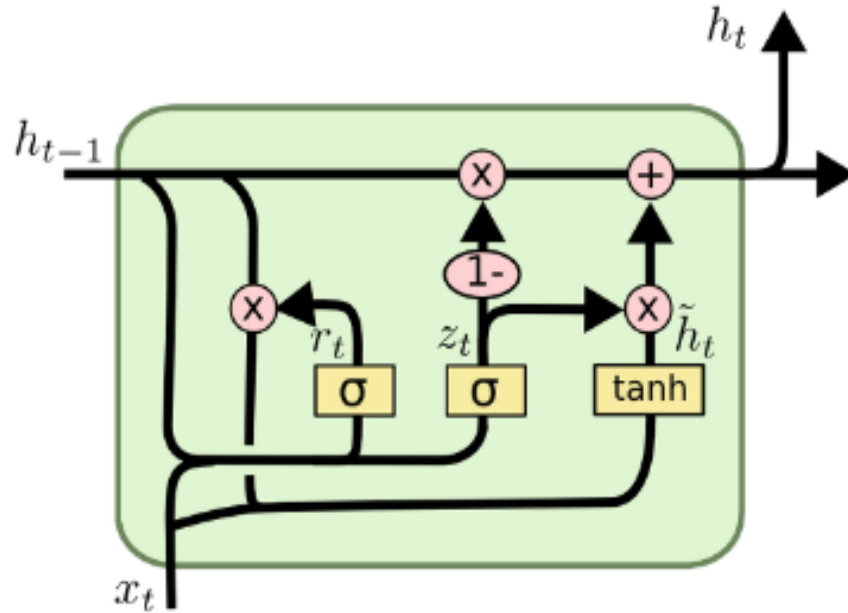
GRU

GRU의 핵심은...

- 1) LSTM의 forget gate와 input gate를 통합하여 하나의 'Update Gate'를 만든 것
- 2) Cell state와 Hidden state를 통합하는 것



GRU



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

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

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

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

z_t : 이전 은닉 상태와 입력 x_t 를 받아 sigmoid 처리
이전 은닉상태에서 얼마나 값을 반영할지 정하는 게이트

h_t~ : 현재 셀 상태

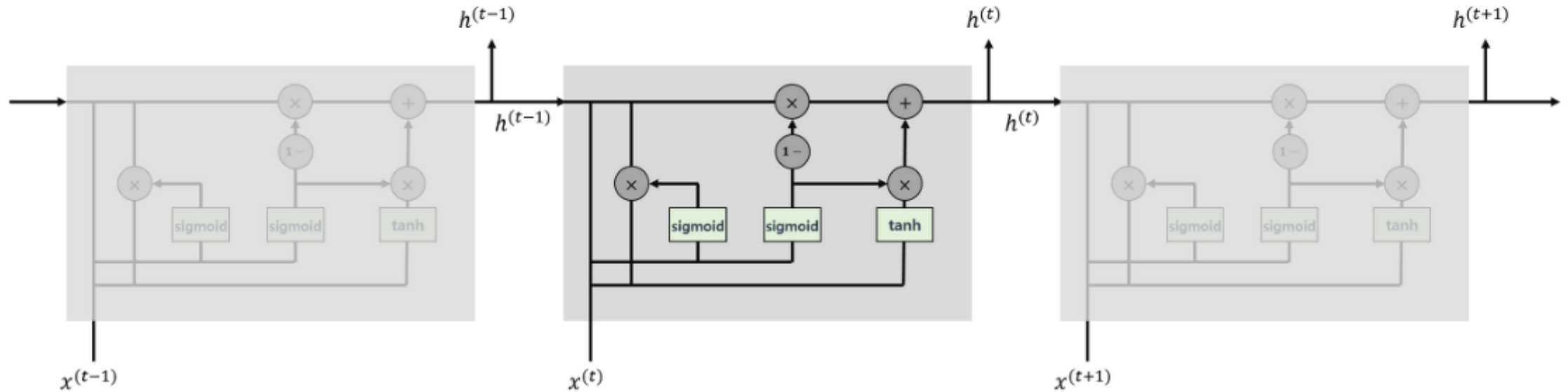
r_t : 이전 상태의 은닉상태와 입력 x_t 를 받아 sigmoid 처리

h_t : $(1 - z_t) * (h_{t-1})$ 는 이전 은닉 상태를 얼마나 잊을지를 정해주고,
 $(z_t * h_{t-1}~)$ 는 현재 상태를 얼마나 반영하는지 정해 줌. 다음 스텝의 은닉 상태가 되는 것.

GRU

4STEP

1. Reset Gate
2. Update Gate
3. Candidate
4. Hidden Layer Calculation



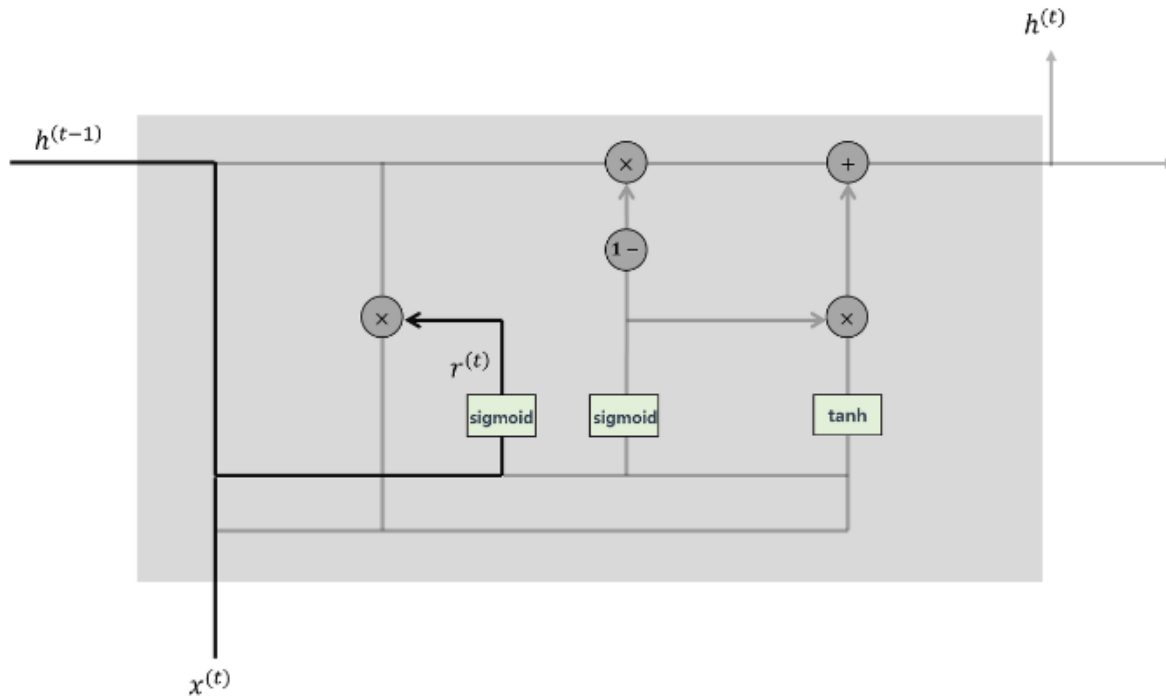
GRU

1. Reset Gate

>> 과거의 정보를 적당히 리셋시키는게 목적

>> sigmoid 함수를 출력으로 이용해 (0,1) 값을 이전 은닉층에 곱해줌

>> 직전 시점의 은닉층의 값과 현시점의 정보에 가중치를 곱하여 얻을 수 있음



$$r^{(t)} = \sigma \left(W_r h^{(t-1)} + U_r x^{(t)} \right)$$

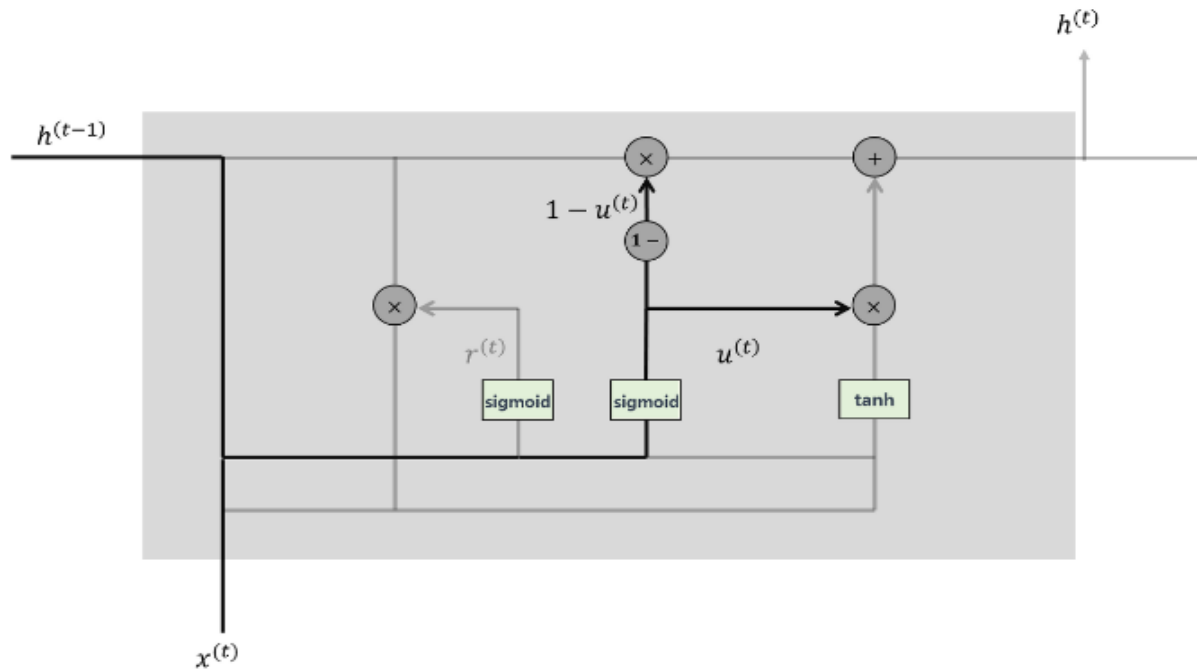
GRU

2. Update Gate

>> LSTM의 forget gate와 input gate를 합쳐 놓은 느낌

>> 과거와 현재의 정보의 최신화 비율을 결정하는 역할

>> sigmoid로 출력된 결과 ($u(t)$)는 현 시점의 정보의 양을 결정하고,
1에서 뺀 값 ($1-u(t)$)는 직전 시점의 은닉층의 정보에 곱해주어,
>> 각각이 LSTM의 input gate와 forget gate와 유사함



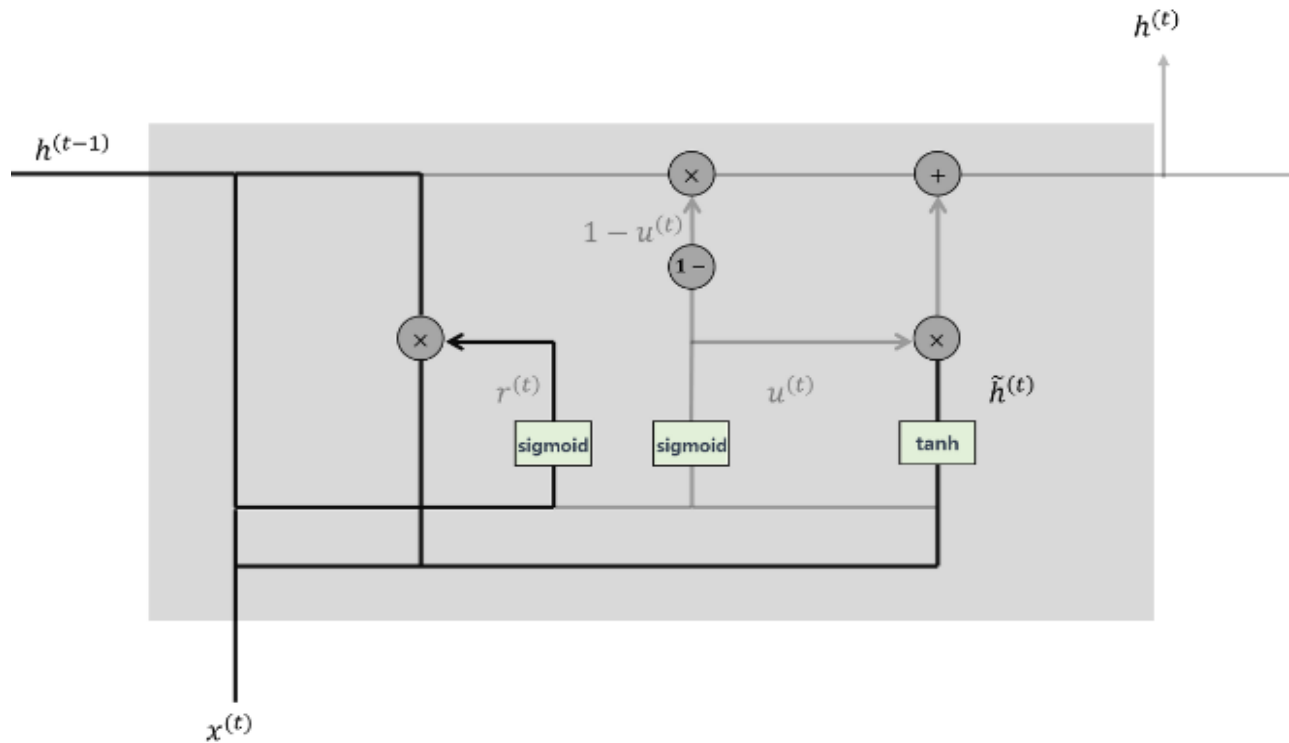
$$u^{(t)} = \sigma \left(W_u h^{(t-1)} + U_u x^{(t)} \right)$$

GRU

3. Candidate

>> 현 시점의 정보 후보군을 계산하는 단계

>> 과거 은닉층의 정보를 그대로 이용하지 않고 리셋 게이트의 결과를 곱하여 이용해주는 것



$$\tilde{h}^{(t)} = \tau \left(W h^{(t-1)} * r^{(t)} + U x^{(t)} \right)$$

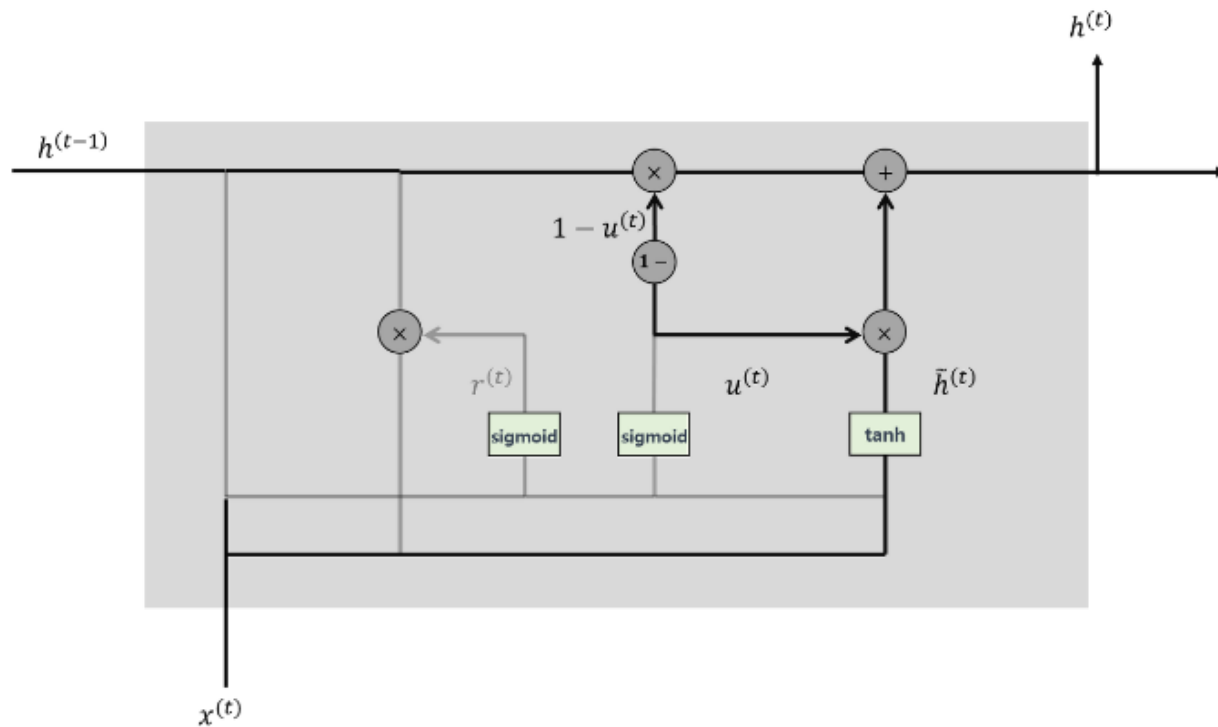
GRU

4. Hidden Layer Calculation

>> update gate 결과와 candidate 결과를 결합하여 현시점의 은닉층을 계산하는 단계

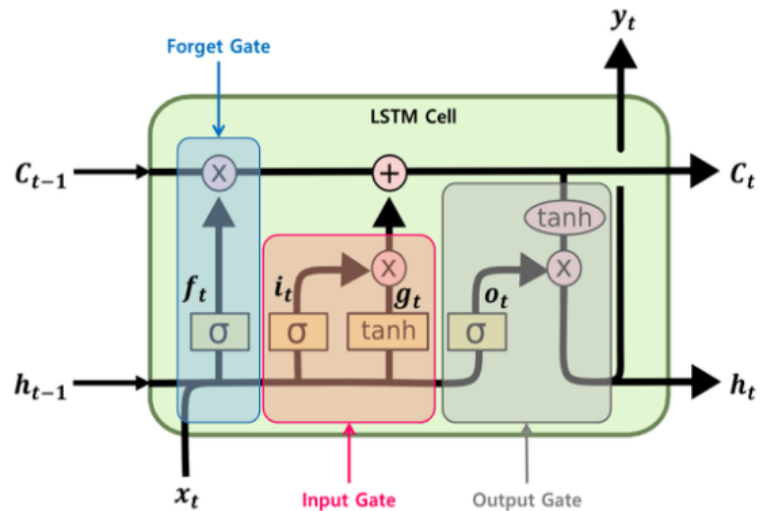
>> sigmoid 함수의 결과는 현시점 결과의 정보의 양을 결정해줌

>> 1-sigmoid 함수의 결과는 과거 시점의 정보의 양을 결정해줌

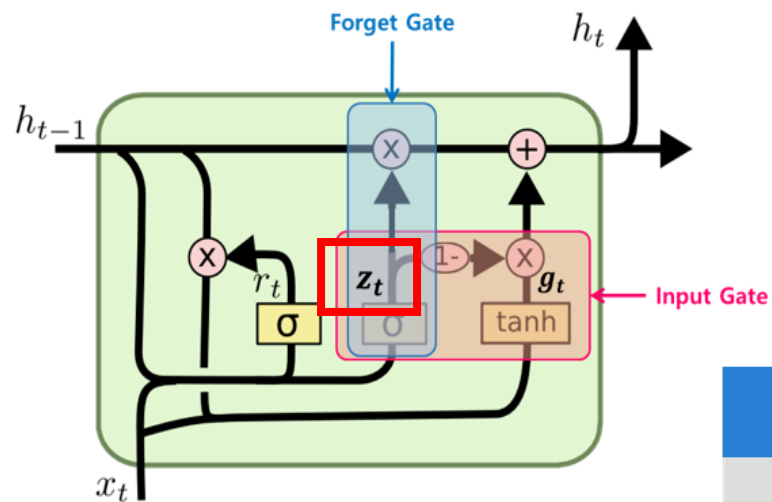


$$h^{(t)} = (1 - u^{(t)}) * h^{(t-1)} + u^{(t)} * \tilde{h}^{(t)}$$

LSTM



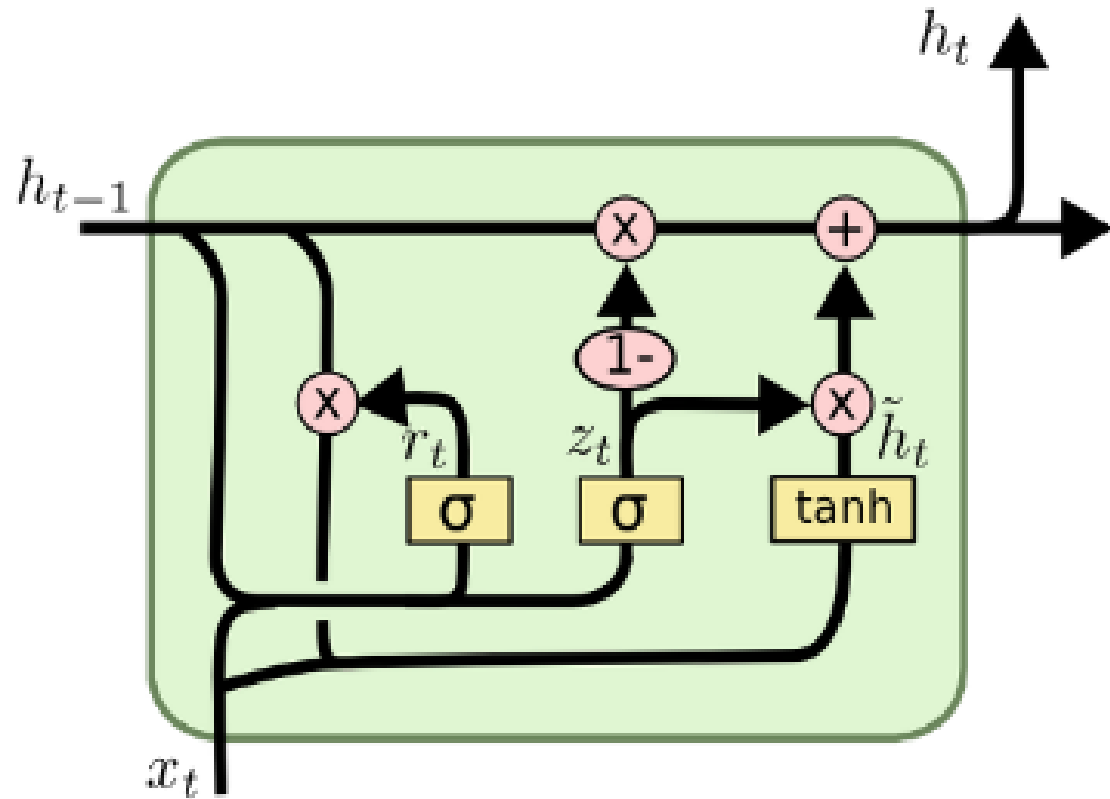
GRU



LSTM vs GRU

LSTM	GRU
gate 수 3개(forget, input, output)	gate 수 2개(reset, update)
Control the exposure of memory content (cell state)	Expose the entire cell state to other units in the network
Has separate input and forget gates	Performs both of these operations together via update gate
More parameters	Fewer parameters

GRU : Contribution



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

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

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

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

GRU

- The end -