Basic RNN / LSTM / GRU Unit Structures

YeLab Reading Group

2018/8/10

This slide is mainly based on Colah's Blog (http://colah.github.io/posts/2015-08-Understanding-LSTMs/)

Outline

- Overview
- Basic RNN Unit
- Long Short Term Memory (LSTM) Unit
- Gated Recurrent Unit (GRU)
- References

Overview

- Recurrent Neural Network
  - Input feature $X_t \in \mathbb{R}^{d_1}$
  - Learned feature $h_t \in \mathbb{R}^{d_2}$

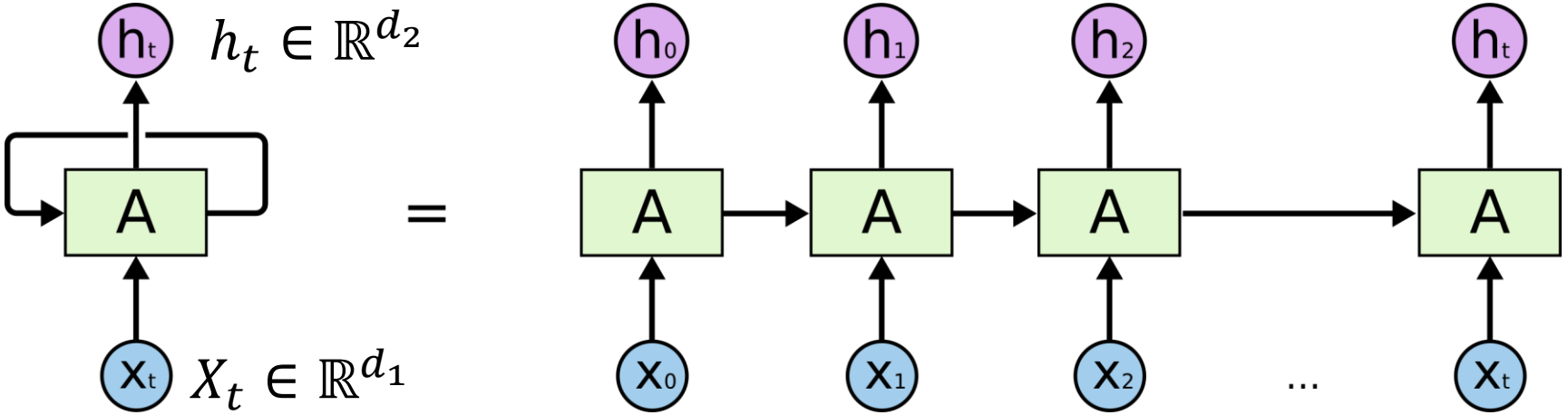


Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Overview

- Goal: Design a recurrent unit ‘A’ such that
  - (1) learn and control short-term / long-term dependencies
  - (2) computationally efficient / simple structure

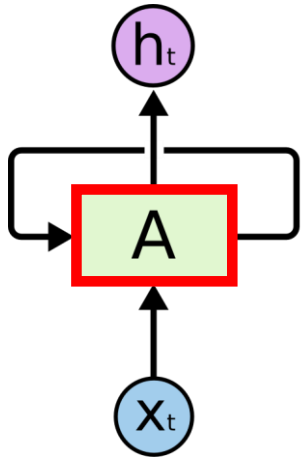


Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Overview

- Basic RNN: simple, less control on information flow.
- LSTM: complex, more control on information flow.
- GRU: simpler than LSTM, intermediate control on information flow.

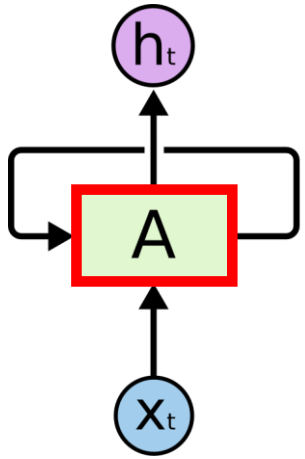
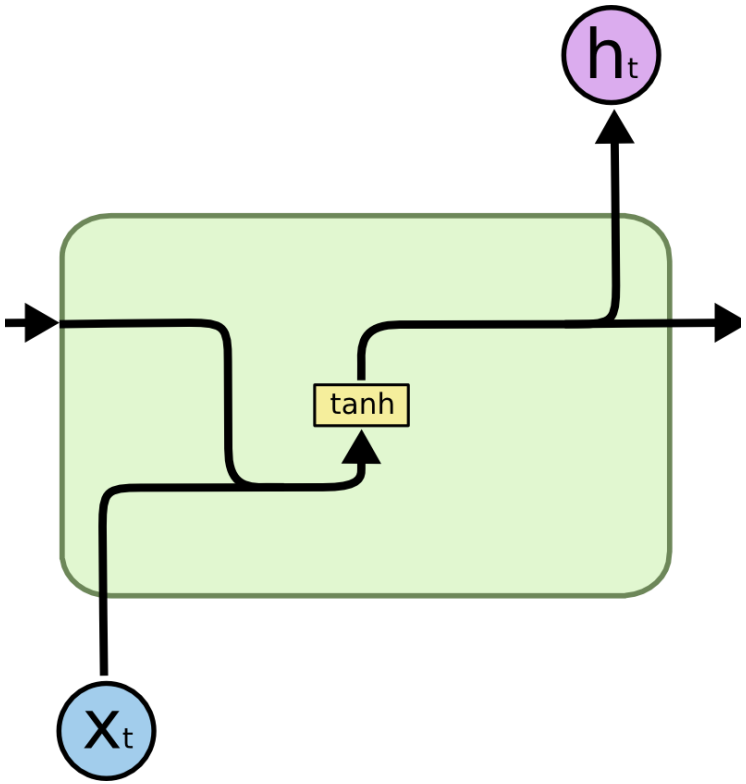


Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

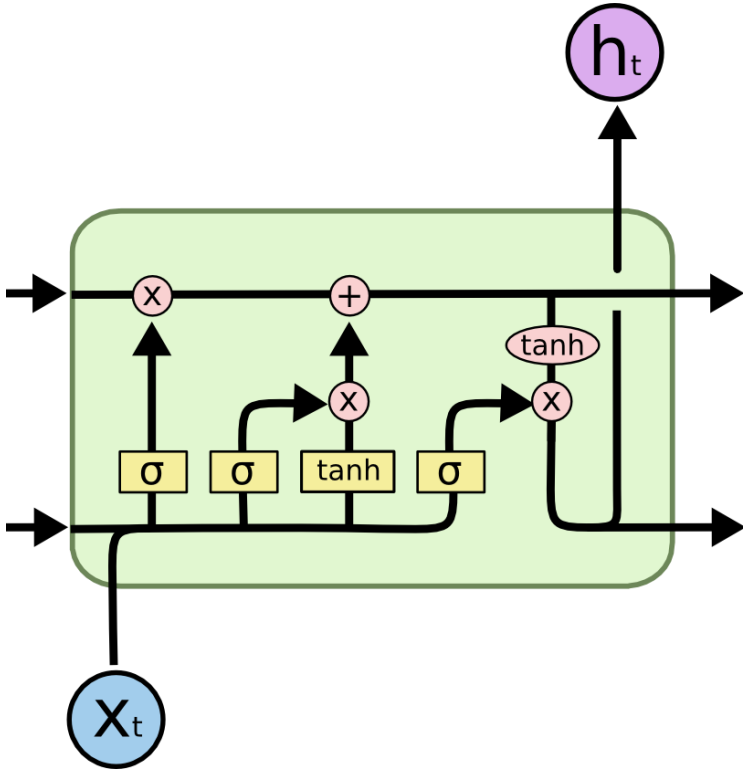
Basic RNN Unit

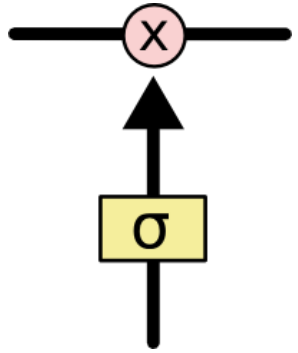


$$h_t = \tanh\left(W \cdot [h_{t-1}, x_t] \; + \; b\right)$$

Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTM Unit



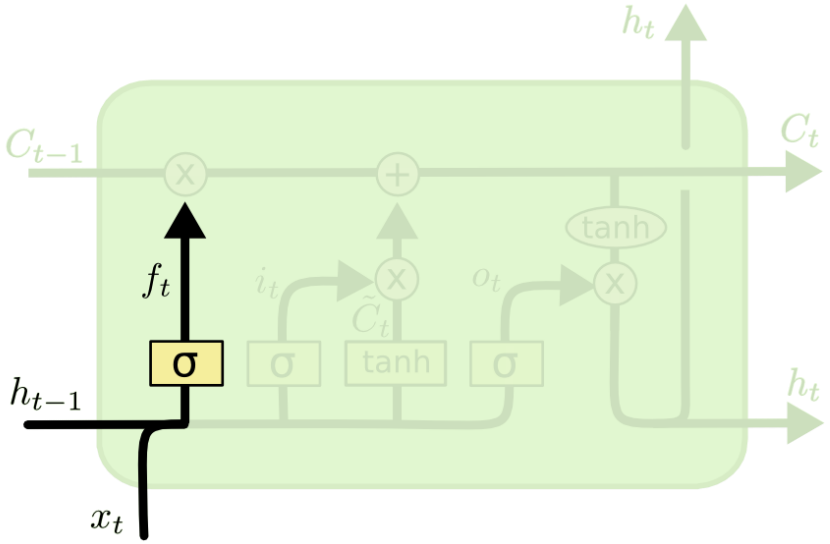


Gates

Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTM Unit

- Forget Gate

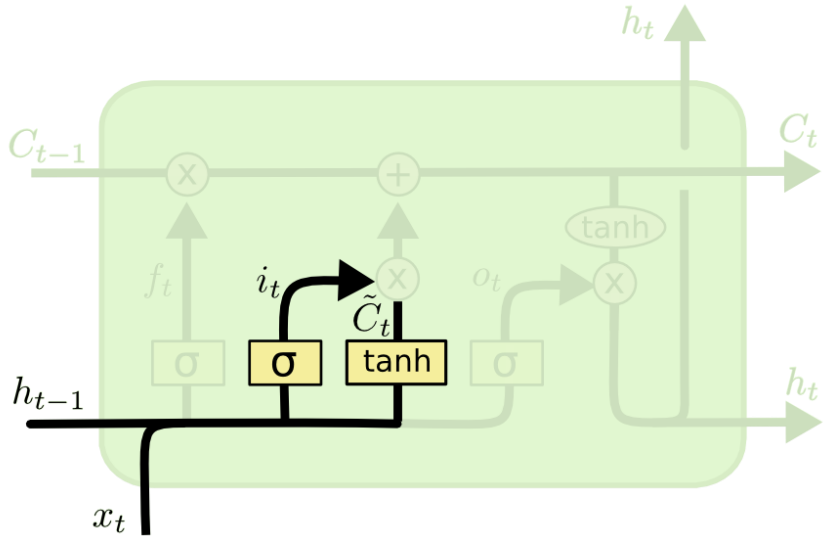


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTM Unit

- Input Gate



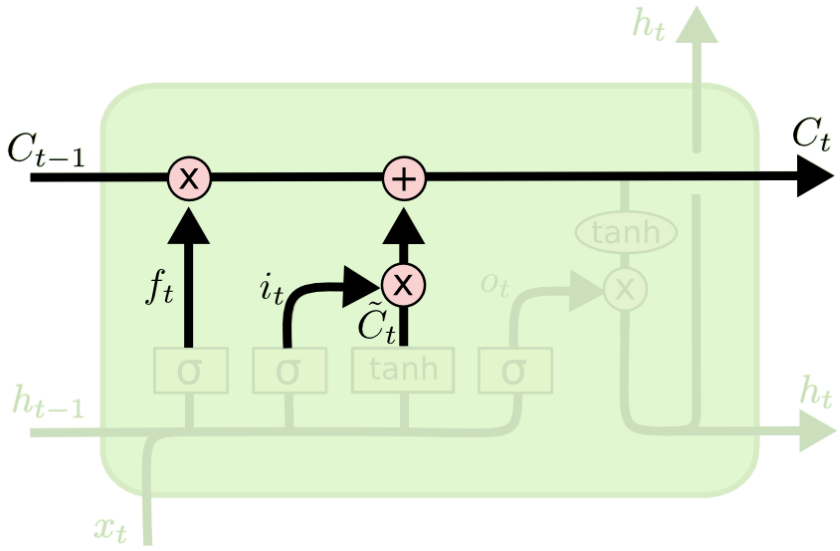
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTM Unit

- Long term memory output

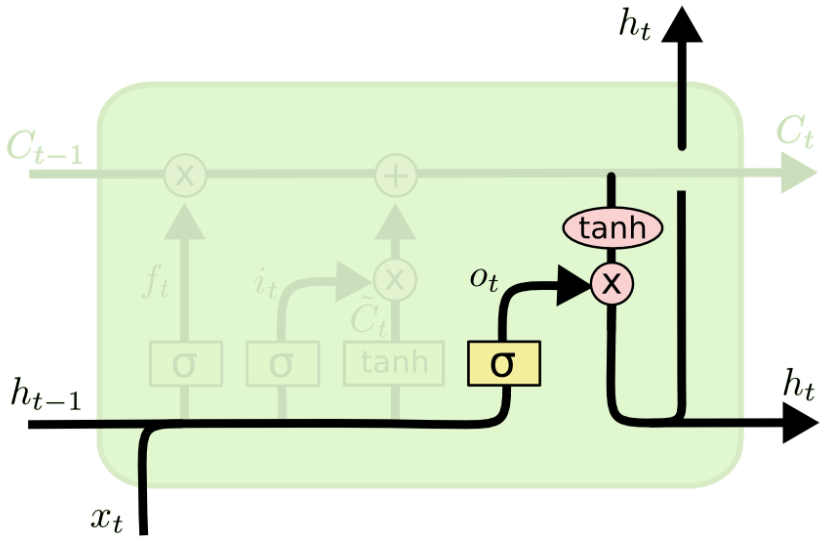


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTM Unit

- Output Gate



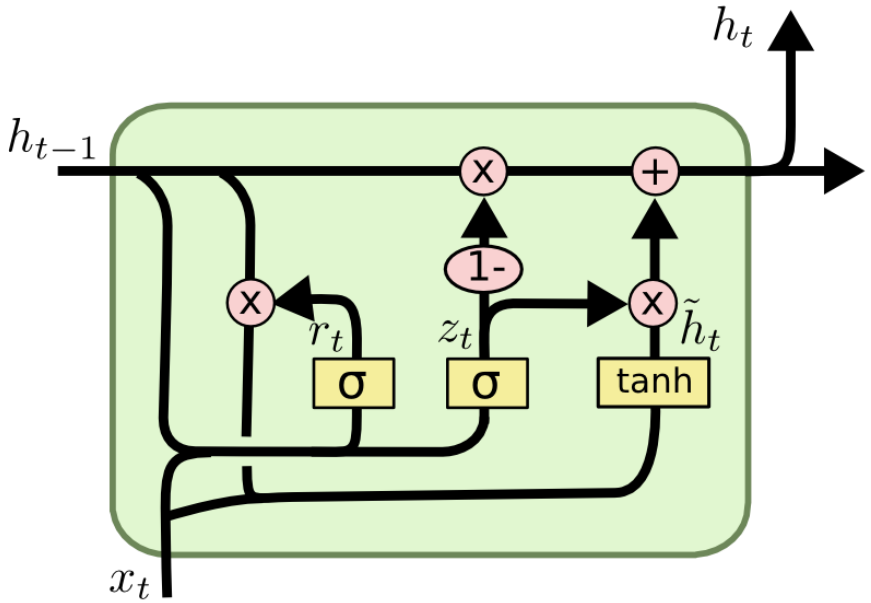
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

GRU

- Combine Forget and Input Gate into one



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

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

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

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

Image from: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

References

- [1] Colah’s Blog: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/
- [2] https://theneuralperspective.com/2016/11/17/recurrent-neural-network-rnn-part-4-custom-cells/