LSTM Understanding and Applications

Xiaoya Liao [1]

[1]Beijing EduBrain

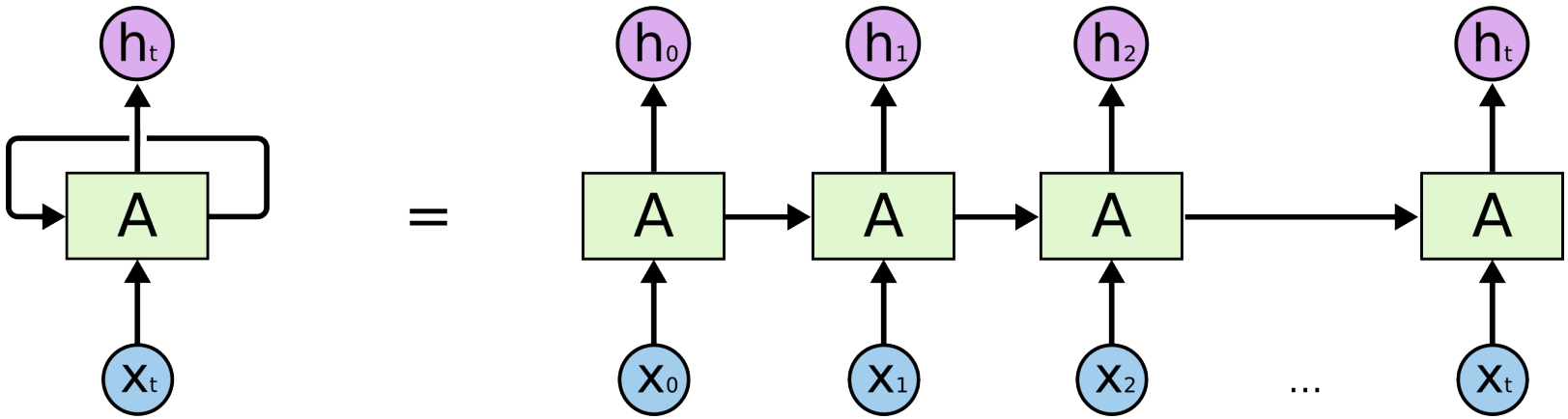
June 1, 2018

Human vs ANNs

Fact: Humans understand each word based on their understanding of previous words.

- Traditional neural networks cant do this
- Recurrent neural networks address this issue. They are networks with loops in them, allowing information to persist.

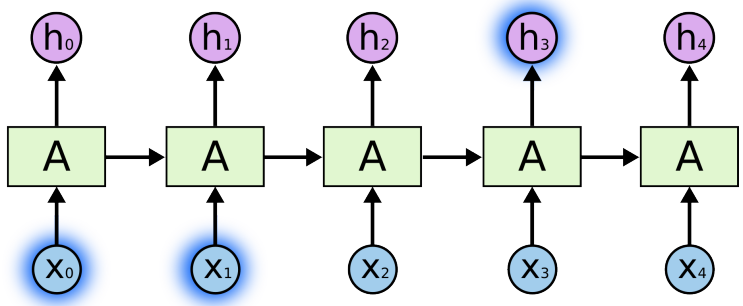
RNNs framework overview



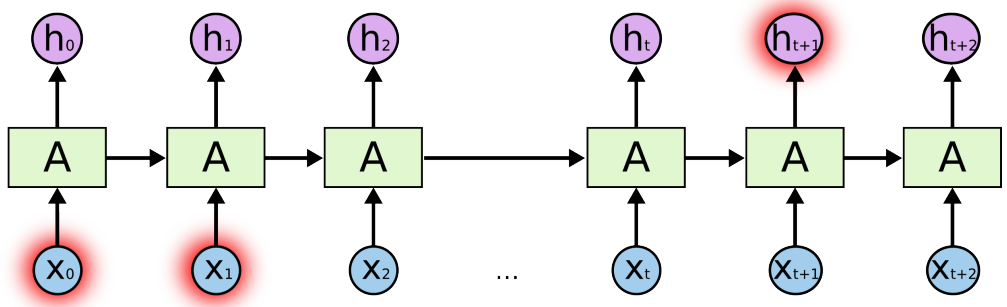
Chain-like nature — related to sequences and lists
Fields: speech recognition, language modeling, translation, image captioning. . . The list goes on.
Essential to these successes is the use of LSTMs, a very special kind of recurrent neural network.

The Problem of Long-Term Dependencies

- The clouds are in the ?, predict that ? = sky

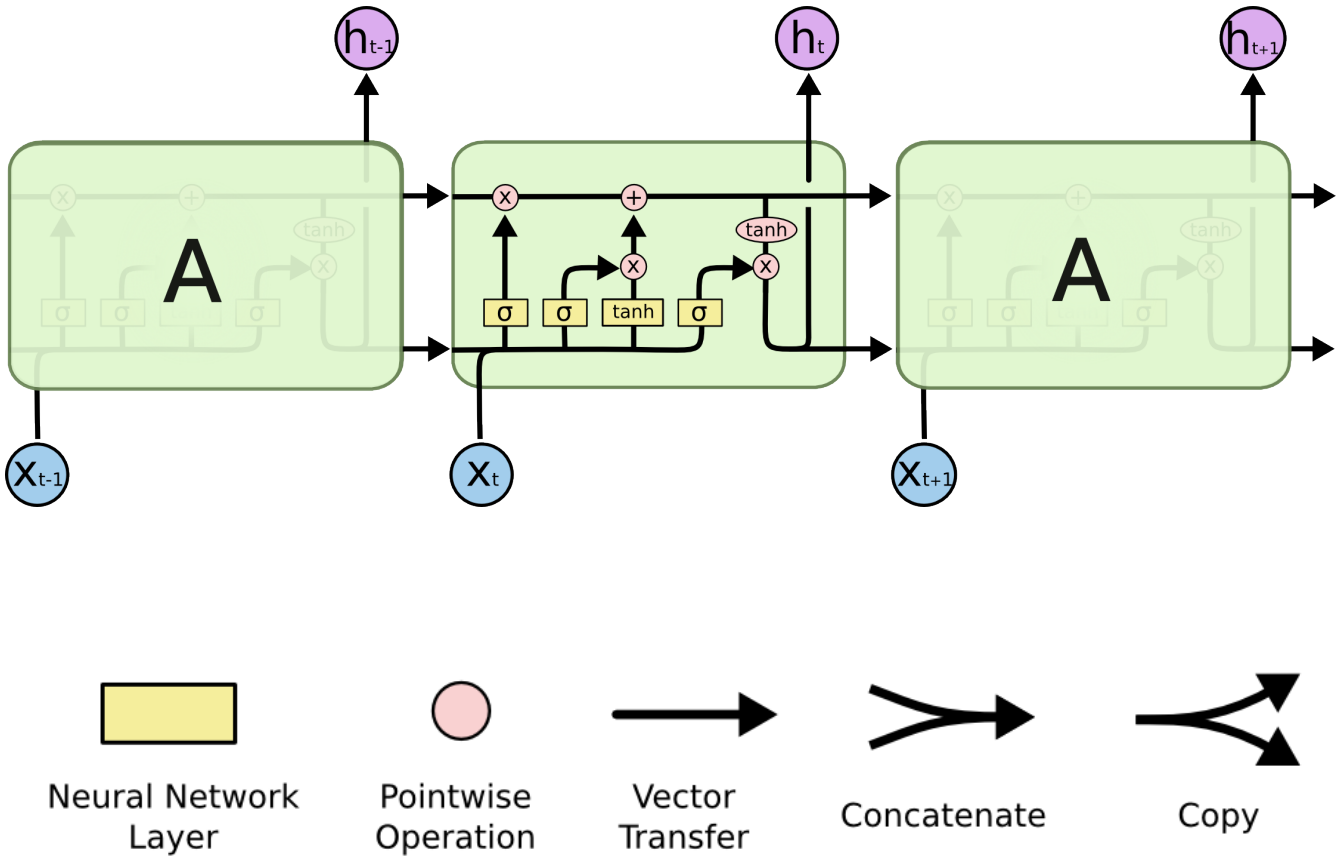


- I grew up in France...I speak fluent ?. predict that ? = French



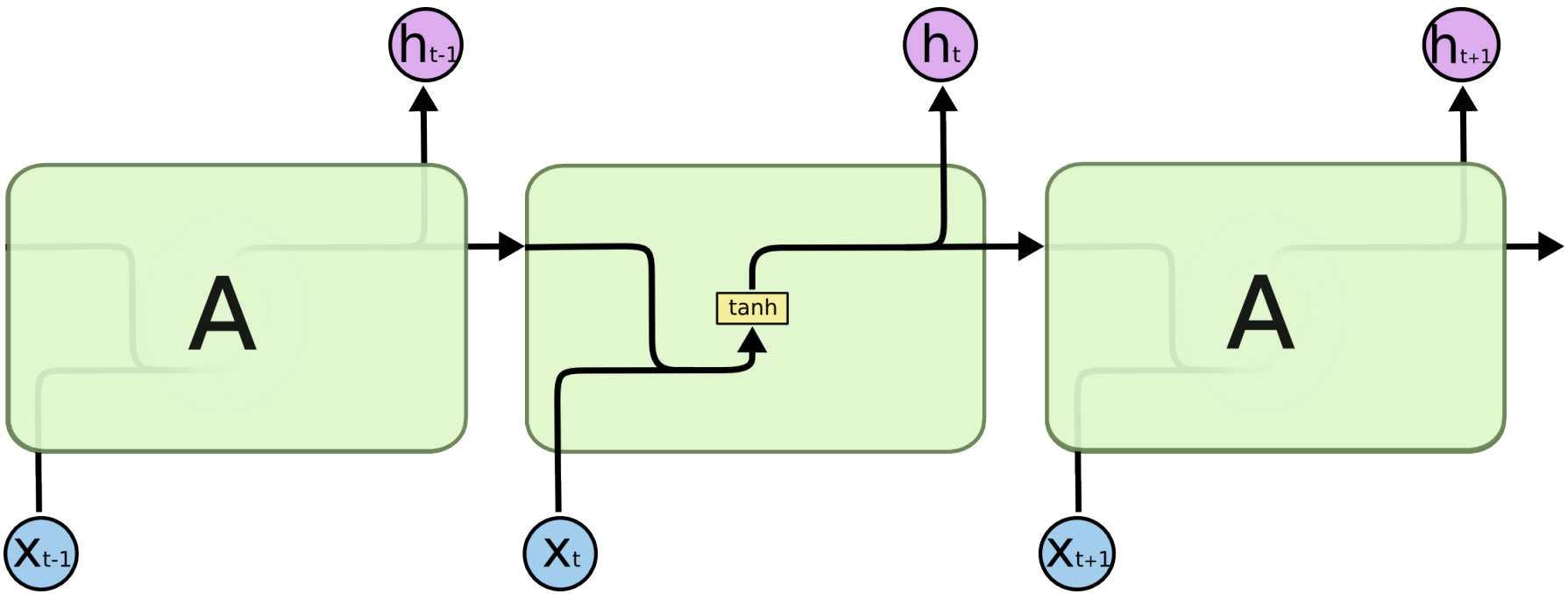
Thankfully, LSTMs dont have this problem!

LSTM framework

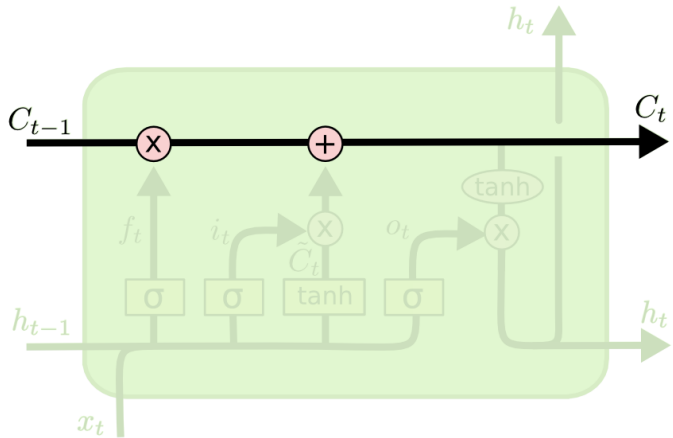


Source:
https://colah.github.io/posts/2015-08-Understanding-LSTMs/

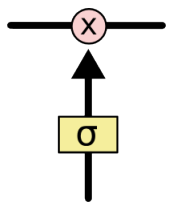
RNNs (LSTM like)



The key to LSTMs is the cell state

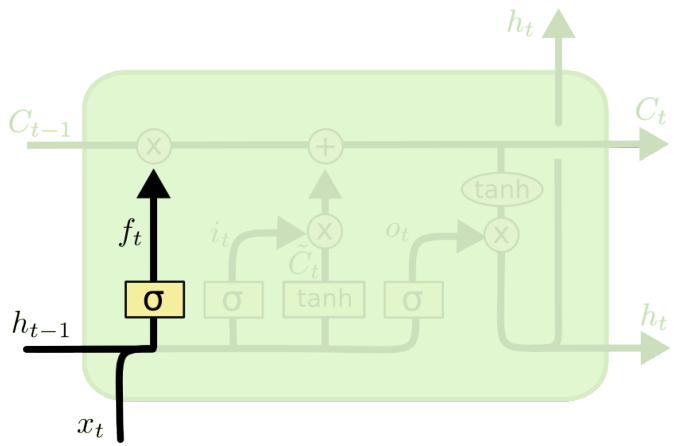


The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates.



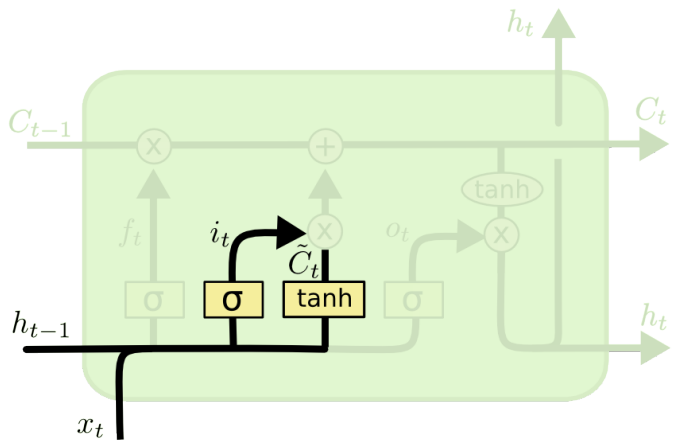
Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and a pointwise multiplication operation.

Forget gate layer



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

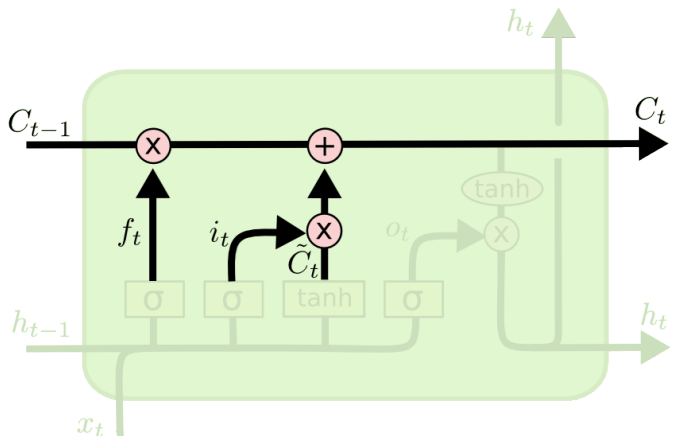
Input gate layer



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

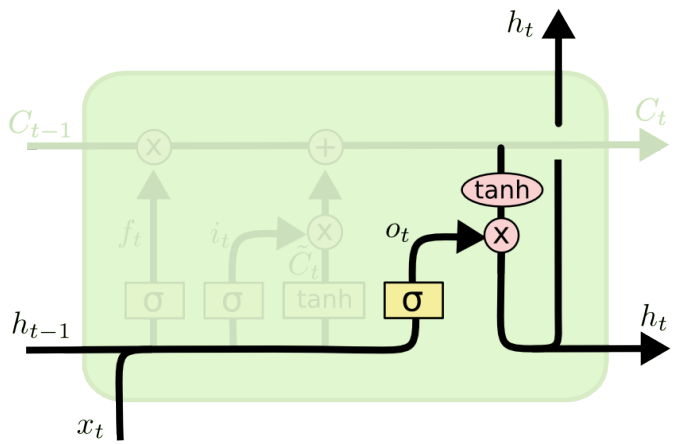
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Update the old cell state



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

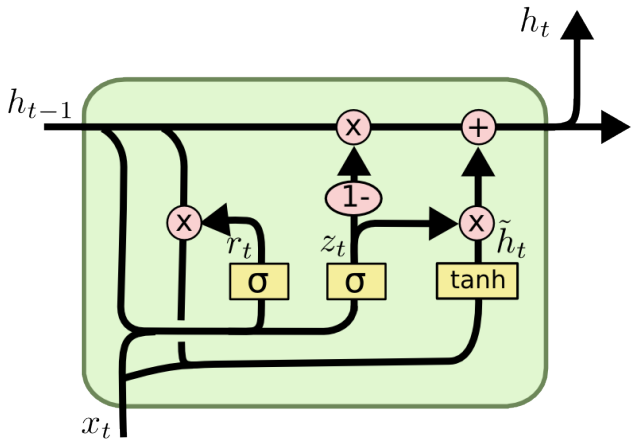
Decide what were going to output



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

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

Variants on Long Short Term Memory



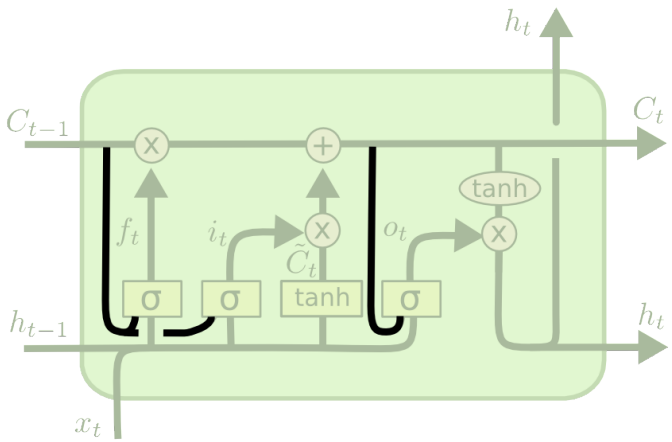
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

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

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

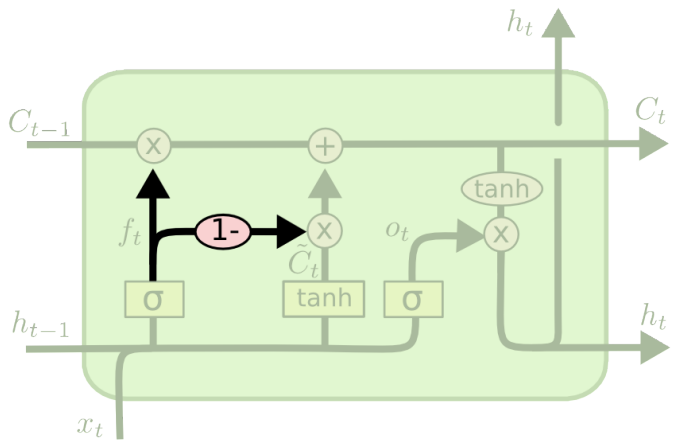
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Variants on Long Short Term Memory



$$f_t = \sigma\left(W_f \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] \; + \; b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] \; + \; b_i\right)$$
$$o_t = \sigma\left(W_o \cdot [\boldsymbol{C_t}, h_{t-1}, x_t] \; + \; b_o\right)$$

Variants on Long Short Term Memory



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

Open source

All Deeplearning libs.
tensorflow, pytorch are the good choices.
mxnet should be implemented by yourself.

Review

1. Human's thoughts have persistence.
2. RNNs applied this feature
3. Application fields: speech recognition, language modeling, translation, image captioning. . . The list goes on.
4. Essential to successes mentioned in 2 is the use of LSTMs, a very special kind of recurrent neural network which has no problem of Long-Term dependencies
5. Open source: all Deep Learning libraries, such as tensorflow, pytorch, mxnet.

The end

Thanks
for your attention!