

Review of Recurrent Neural Network and Long Short-Term Memory Network

Guannan Hu

April 11, 2018

In Figure 3, each line carries an entire vector, from the output of one node to the inputs of others. The pink circles represent pointwise operations, like vector addition, while the yellow boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denote its content being copied and the copies going to different locations.

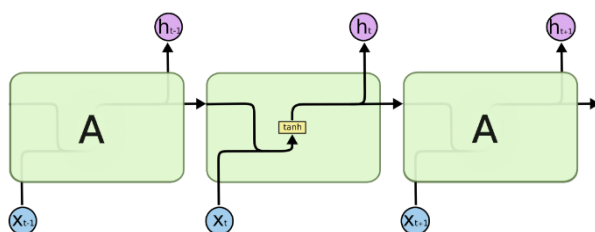


Figure 1: The repeating module in a standard RNN contains a single layer

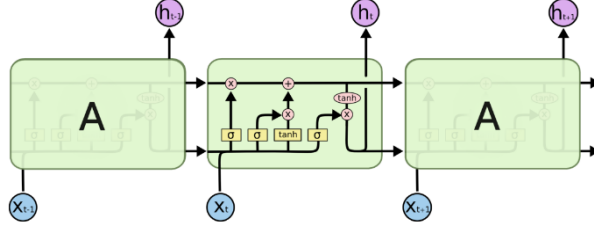


Figure 2: The repeating module in an LSTM contains four interaction layers

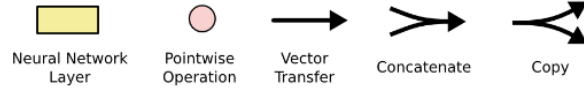


Figure 3: the operations in LSTM

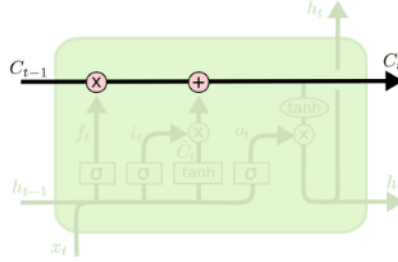


Figure 4: The details of LSTM Cell

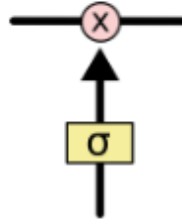
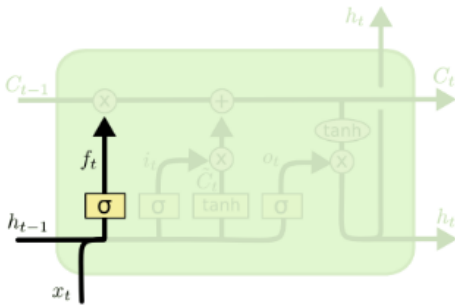
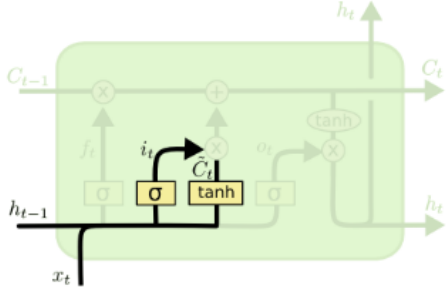


Figure 5: The sigmoid layer in LSTM Cell



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

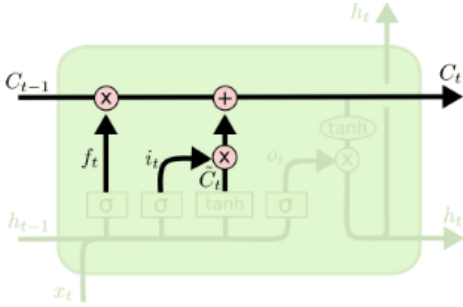
Figure 6: The details of LSTM Cell 2



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

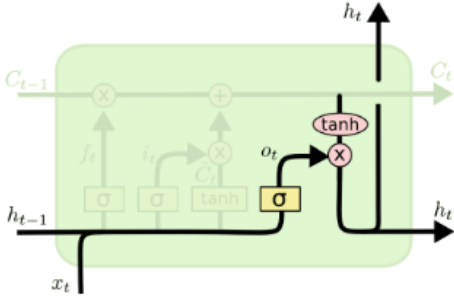
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Figure 7: The details of LSTM Cell 3



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

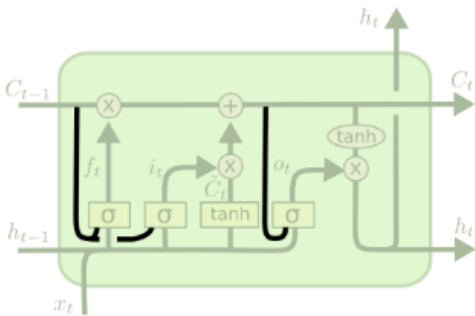
Figure 8: The details of LSTM Cell 3



$$o_t = \delta(W_o \cdot [h_{t-1}, x_t] + b_o)$$

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

Figure 9: The details of LSTM Cell 4

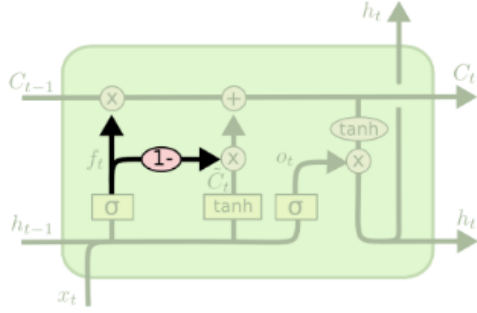


$$f_t = \delta(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f)$$

$$i_t = \delta(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i)$$

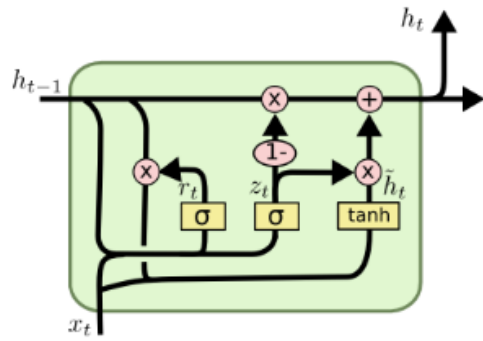
$$o_t = \delta(W_o \cdot [C_t, h_{t-1}, x_t] + b_o)$$

Figure 10: Variant of LSTM 1 : adding peephole connections



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

Figure 11: Variant of LSTM 2 : couple forget and input gates



$$\begin{aligned} z_t &= \delta(W_z \cdot [h_{t-1}, x_t]) \\ r_t &= \delta(W_r \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t &= \tanh(W \cdot [r_t * h_{t-1}, x_t]) \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{aligned}$$

Figure 12: Variant of LSTM 3: Gated Recurrent Unit, GRU