Long Short Term Memory(LSTM)

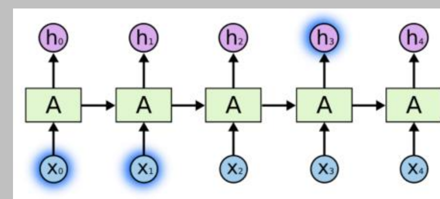
Tiancheng Yuan &
Renzhi Sheng

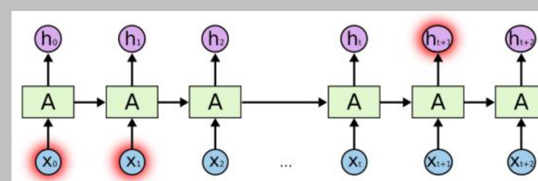
Definition

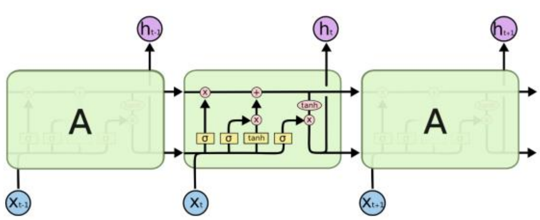
- A special type of Recurrent Neural Networks which has extraordinary performances on a large variety of problems.
- Explicitly designed to have default behaviour of remembering information for a long period.
- A chain-like repeating modules, and more complicated comparing to standard RNN.

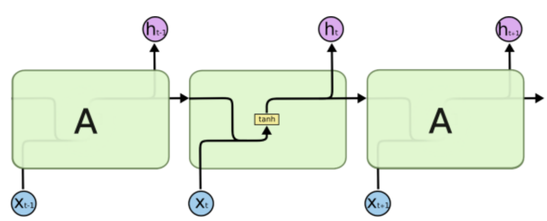
Long-term dependency

- The long-term dependency problem. Theoretically, RNNs are expected to be capable of handling long term dependencies. However, actual practises disagree. In other words, when the gap between previous information and present task grows, the network will become less capable of learning to connect the information.

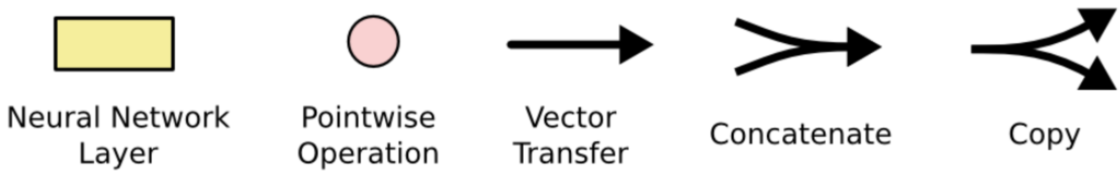








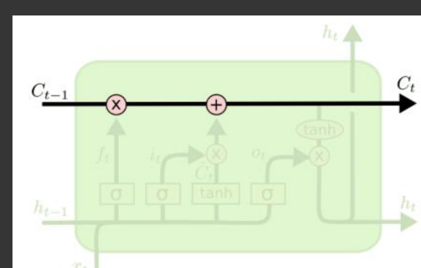
Repeating module(LSTM vs. RNN)

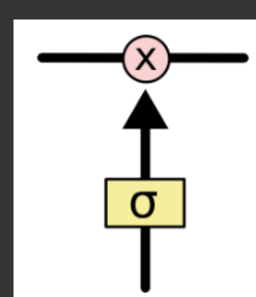


Basic components

The Core ideal of LSTM

- The cell state (**C**), a horizontal line across the top.
- As a conveyor belt, run through the entire chain with only minor interactions. Information can flow along it unchanged.
- A structure called gates is used to remove or add information to **C.**
- A sigmoid function has value [0,1], describing how much of each component should go through.

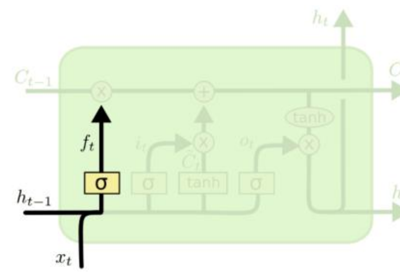




LSTM walk through-1

Decide what information to throw away from cell state – 'forget gate layer'. It looks h_{t-1} and x_t, and output a value [0,1] to determine whether to ignore or keep this component.

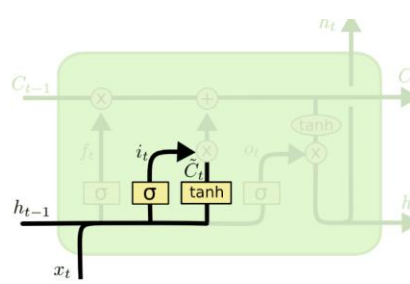
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$



LSTM walk throught-2

- The next step is to decide how to update the cell state,C. There are two components:
  1. a 'input gate layer' to decide the value to update;
  2. a hyperbolic tangent layer creates a vector for new candidate value, $\tilde{C}_t$ ·, to be added to the state.

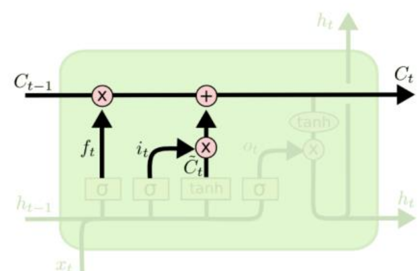
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$



LSTM walk through-3

- In this step, the old cell state C_{t-1} need to be updated to new C_t.
- The old state multiplies f_t to forget the things chosen earlier
- The new candidate value will then be added into the above value to get the new cell state.

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



LSTM walk through-4

- Finally, the structure to decide the output. This output will be a filtered version of the cell state. First, a sigmoid function to decide what parts of cell state to output. Then, we multiply the cell state with sigmoid gate output. Therefore, we can have the desired output.

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