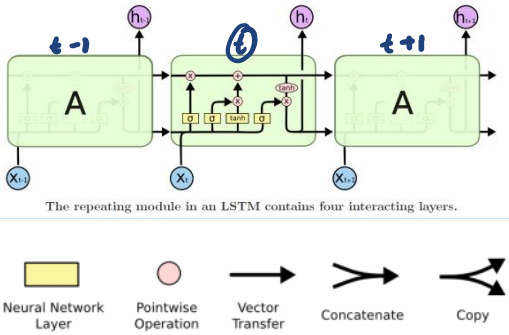
20_Variations_in_LSTM

12 September 2024 21:35

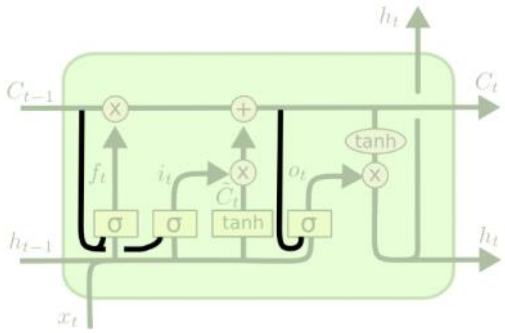


The repeating module in an LSTM contains four interacting layers.

Variations in LSTM:-

① Peephole connections :- 2000s Gers & Schmidhuber

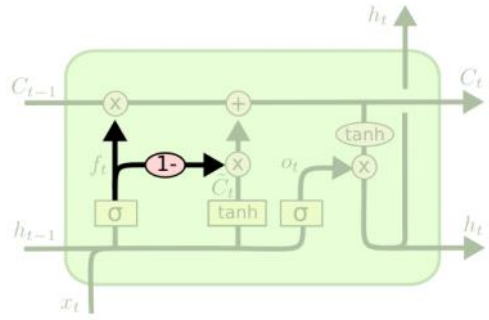
⇒ let the gate layers look at the cell states.



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

② Coupled forget and input gate.

instead of seperately deciding what to forget and what we should add new info to, we make those decisions together.



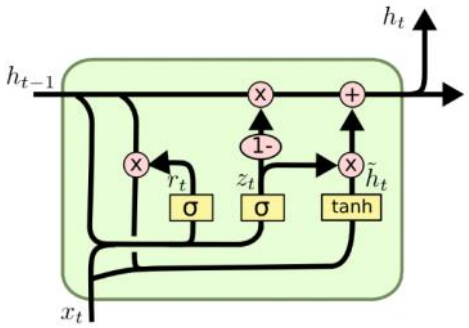
$$C_t = f_t \cdot C_{t-1} + (1 - f_t)\tilde{C}_t$$

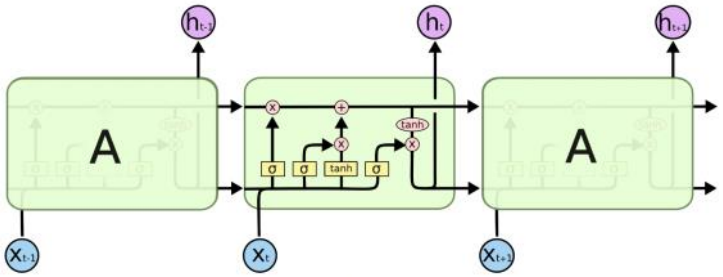
GRU (Gated Recurrent Unit)

2014 Cho et. al.

① Combines forget and input gates into a single "update gate".

② Merges the cell state and hidden state, and make some other changes.





The repeating module in an LSTM contains four interacting layers.

"simpler model than LSTM".

$$z_t = \sigma(W_z [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r [h_{t-1}, x_t])$$
$$\tilde{h}_t = \tanh(W [r_t \cdot h_{t-1}, x_t])$$
$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t$$