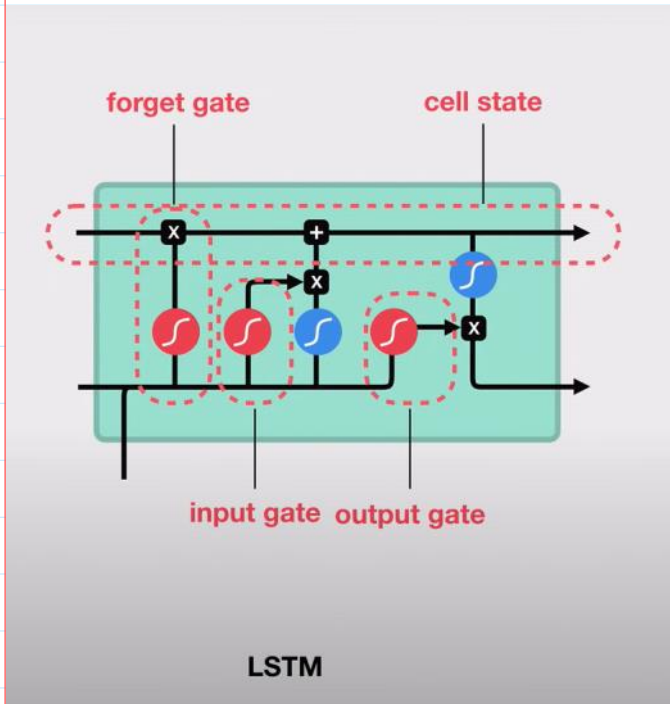


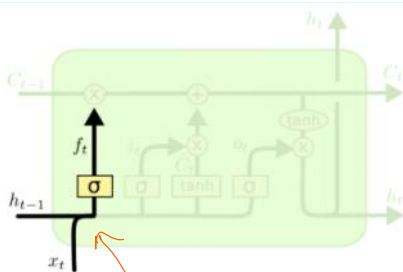
⇒ LSTM (Long short term memory)

These are used to overcome problem of vanishing gradients in RNN.



⇒ cell state :-

→ It is like a conveyor belt which runs down through the entire network. It carries the info from previous cells and that info can be added or removed in current cell as follows :-

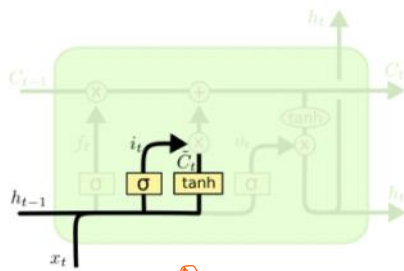


$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

forget gate (decides which info to through away)

This gate has a sigmoid activation function by which we decide should we keep the information (number) from previous cell (1) or through it away (0 or σ).

decide should we keep the information (number) from previous cell (i) or through it away (0 of σ).

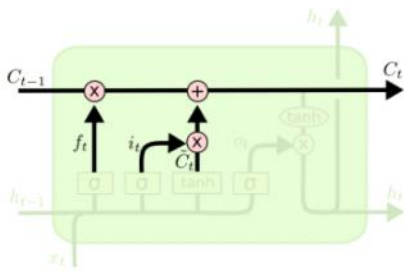


$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

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

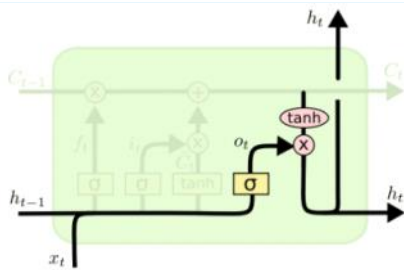
(input gate layer) How much info should we add.

The sigmoid part of it decides how much of a value should be added to the new cell state and tanh is the value that should be added (information of current cell)



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

this is the new cell state



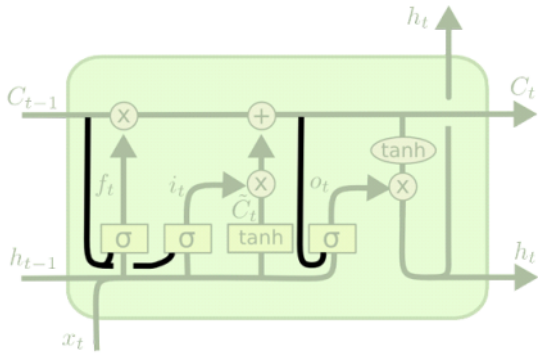
$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$

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

In the last gate we decide what we are going to output which is based on sigmoid that decides how much of the cell state should be the output but the cell state is passed through tanh and multiplied by sigmoid. The resulting number is output of cell.

⇒ versions of LSTM

① LSI m with peephole connection



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

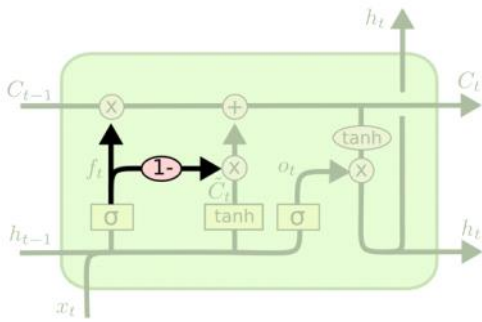
$$i_t = \sigma (W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i)$$

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

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

Qn the above diagram peep holes are added to every gate and these holes come from cell gate.

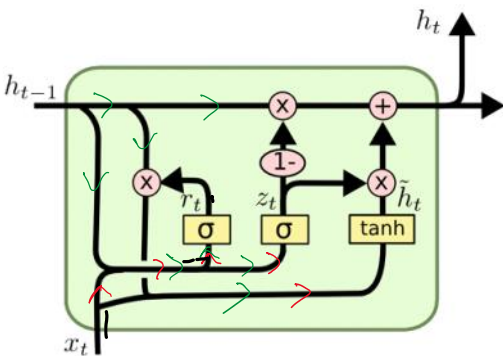
② coupled forget & input gate



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

In this the input and forget gate are coupled. In this cell we only forget when we input something or we only input something when we forget something.

③ Rated Recurrent unit :-



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

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

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

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

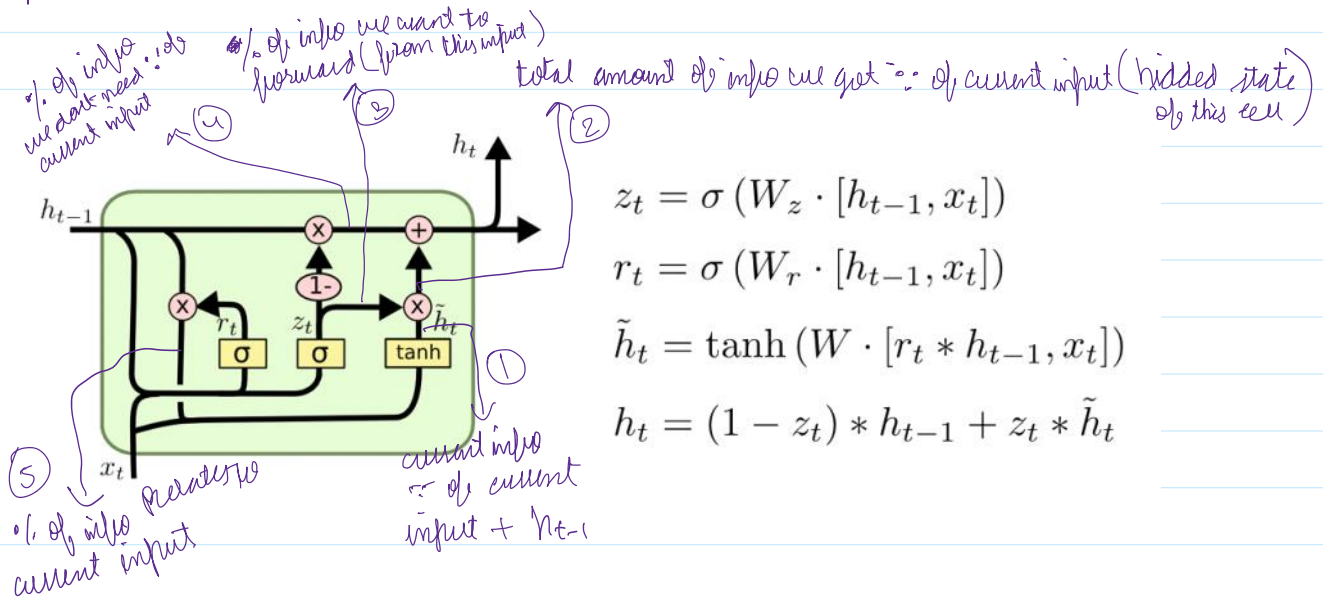
\Rightarrow It combines the input and input rate into single

⇒ It combines the forget and input gate into single gate called "update gate".

It also merges the cell state and hidden state

* The resulting model is simpler than LSTM.

personal analysis



Transformers

Monday, 5. August 2024

09:49