Day 12, 9th Jan 2024

NLP Live Class - by mentor Dr Ayan Debnath
IIT Delhi + Harvard University Alumni

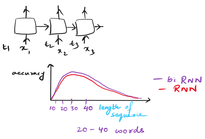
Recap: RNN (padding, embedding)
Practical implementation (Coding)
Bi-directional RNN

To connect:
LinkedIN:
https://www.linkedin.com/in/ayan-debnath/

RNN → disadvantage → ① length of input sequence ↑ tends to forget

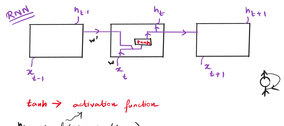
200
300
500

→ not able to memorize long term sequence
→ suffers from long term memory dependencies
↳ Goyal



LSTM (Long Short term Dependencies)

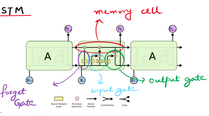
capable of learning long term dependencies
1997 by Schmidhuber



tanh → activation function

$h_t \quad y = f(w x_t + w' h_{t-1})$

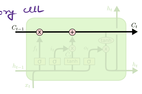
LSTM



① memory cell
② Forget Gate
③ Input gate
④ Output gate
} gate

① MS. Dhoni was a cricket Indian Captain & he was good wicket Keeper.
② Now Rohit Sharma is captain & he is good Batsman.

① memory cell



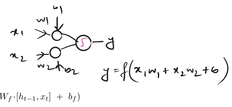
C_t → Current cell/state
C_{t-1} → previous cell

② what is the main function of memory cell
Ans Remembers
↳ relevant/imp info
Forgets → irrelevant/unimp
Context of information.

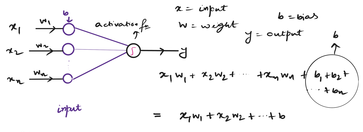
⊗ multiplication
⊕ addition

Pointwise operation

$$\begin{bmatrix}10\\20\\30\\50\\86\\90\end{bmatrix} \otimes \begin{bmatrix}1\\0\\1\\0\\1\\1\end{bmatrix} = \begin{bmatrix}10\\0\\30\\0\\86\\90\end{bmatrix}$$



$y = f(x_1 w_1 + x_2 w_2 + b)$



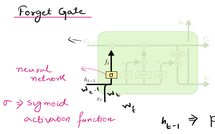
x = input
w = weight
y = output
b = bias

$x_1 w_1 + x_2 w_2 + \cdots + x_n w_n + b_1 + b_2 + \cdots + b_n$

$= x_1 w_1 + x_2 w_2 + \cdots + b$

$y = f(x_1 w_1 + x_2 w_2 + b)$

Forget Gate



σ → sigmoid activation function

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

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

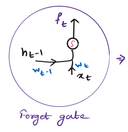
h_{t-1} → previous output
x_t → present input

h_{t-1} & x_t are similar → f_t $\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$ → remembering

$C_{t-1} \begin{bmatrix}10\\20\\30\\\vdots\end{bmatrix} \times \begin{bmatrix}1\\1\\1\\\vdots\end{bmatrix} \rightarrow$ —

h_{t-1} & x_t → context change → f_t $\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}$

$C_{t-1} \begin{bmatrix}10\\20\\30\\\vdots\end{bmatrix} \times \begin{bmatrix}0\\0\\1\\0\\0\\0\end{bmatrix} \rightarrow \begin{bmatrix}\\\\\\\\\end{bmatrix}$ → forgetting

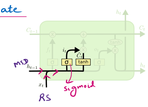


Forget gate

→ $f_t = \sigma(x_t w_t + h_{t-1} w_{t-1}) + b$

$f_t = \sigma[W(x_t, h_{t-1}) + b]$

Input gate



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

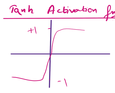
MSD —
RS —

Adding information to memory cell
↳ context change → new info
↳ relevant info

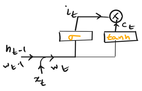
h_{t-1} & x_t → similar → no change in context $\begin{bmatrix}1\\1\\1\\1\\0\end{bmatrix}$

h_{t-1} & x_t → change in context → $\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}$

Tanh Activation fn



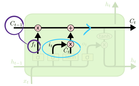
$\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}$ i_t $\begin{bmatrix}1\\1\\1\\1\\-1\\1\end{bmatrix}$ C_t → added to memory cell



$i_t = \sigma\left[(x_t w_t + h_{t-1} w_{t-1}) + b_i\right]$ -- (I)

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

$C_t = \tanh\left[W_c \cdot [h_{t-1}, x_t] + b_c\right]$ -- (II)

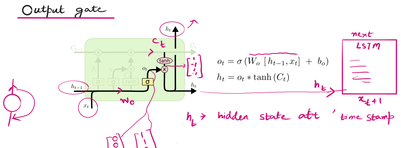


$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

forget gate
input gate

$C_t = f_t * C_{t-1} + i_t * C_t$

Output gate



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

h_t → hidden state at 'time stamp'

$\begin{bmatrix}0\\0\\1\\0\\1\end{bmatrix} \times \begin{bmatrix}1\\-1\\1\\1\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}$ → output from current state passed to next

o_t C_t