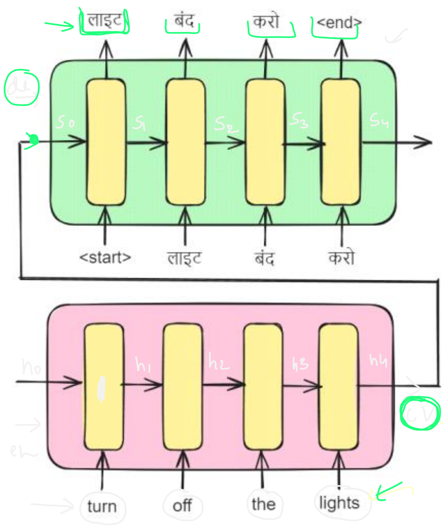
Recap

16 January 2024 16:10



turn off the lights → लाइट बंद करो

[NMT]

encoder-decoder

LSTM

stacked LSTM

set of numbers

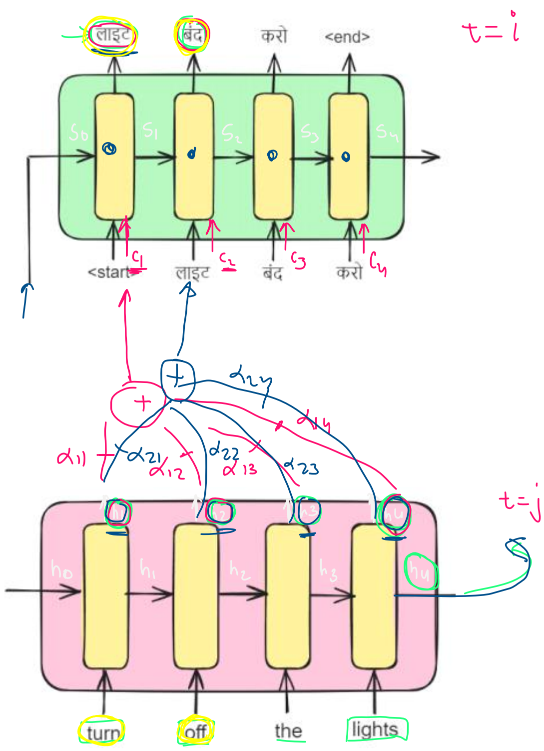
sentence >30 words

paragraph

document

translation

bottleneck → Attention mechanism



t=i

$c_1 \; c_2 \; c_3 \; c_4$

weighted sum

$$\boxed{c_i = \sum_{j=1}^{4} \alpha_{ij} h_j}$$

$\alpha \rightarrow$ alignment score

$4 \times 4 = 16$

$$c_1 = \alpha_{11} h_1 + \alpha_{12} h_2 + \alpha_{13} h_3 + \alpha_{14} h_4$$

$$c_2 = \alpha_{21} h_1 + \alpha_{22} h_2 + \alpha_{23} h_3 + \alpha_{24} h_4$$

$\alpha \rightarrow$ find out

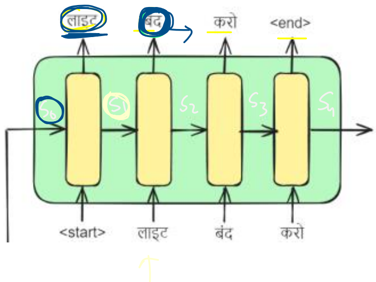
Bahdanau attention

Luong attention

Bahdanau Attention

16 January 2024 16:11

t=1



alignment score

$c_i = \sum \alpha_{ij} h_j$

given

decoder → prev hidden state

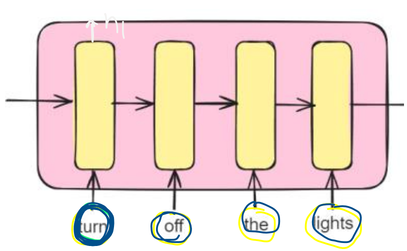
α

alignment

$\alpha_{11} \rightarrow$ लाइट → turn

$\alpha_{12} \rightarrow$ बंद → off

$\alpha_{11} = f(h_1, s_0)$ $\alpha_{21} = f(h_1, s_1)$



t=j

→ $\alpha_{ij} = f(h_j, s_{i-1})$ math func → the encode hidden | decod prev hidden state

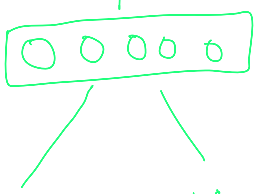
approximate

FFN → ANN → UFA

α_{ij}

softmax

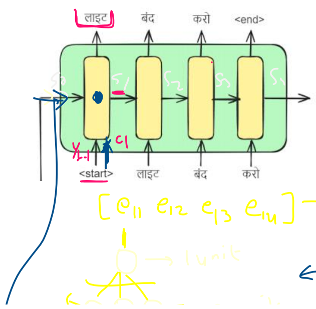
e_{ij}



s_{i-1} h_j

s_{i-1}, h_j → FFN → e_{ij} → softmax → α_{ij} → c_i → y_t

$s_0 \; y_{t-1} \; c_1 \rightarrow$ lstm → y_t (लाइट) $\boxed{s_1}$



i=1 $\boxed{t=2}$

encoder sentence

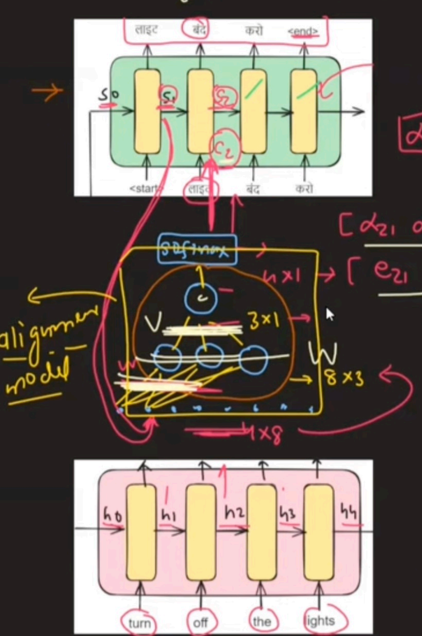
↳ $h_1 \; h_2 \; h_3 \; h_4$

s_{i-1} h_j

$c_1 = \sum \alpha_{ij} h_j$

$[e_{11} \; e_{12} \; e_{13} \; e_{14}] \rightarrow$ softmax $\rightarrow [\alpha_{11} \; \alpha_{12} \; \alpha_{13} \; \alpha_{14}]$

3×1 ← $c_1 = \alpha_{11} h_1 + \alpha_{12} h_2 + \alpha_{13} h_3 + \alpha_{14} h_4$



$c_2 = \alpha_{21} h_1 + \alpha_{22} h_2 + \alpha_{23} h_3 + \alpha_{24} h_4$

$$\begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} & h_{11} & h_{12} & h_{13} & h_{14} \\ s_{11} & s_{12} & s_{13} & s_{14} & h_{21} & h_{22} & h_{23} & h_{24} \\ s_{11} & s_{12} & s_{13} & s_{14} & h_{31} & h_{32} & h_{33} & h_{34} \\ s_{11} & s_{12} & s_{13} & s_{14} & h_{41} & h_{42} & h_{43} & h_{44} \end{bmatrix}$$

4×8

time distributed FNN

$[\alpha_{21}\ \alpha_{22}\ \alpha_{23}\ \alpha_{24}]$

4×1 $[e_{21}\ e_{22}\ e_{23}\ e_{24}]$

4×3

$c_{ij} = \sum \alpha_{ij} h_j$

$\alpha_{ij} = \dfrac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}$

$e_{ij} = V[\tanh(W[s_{i-1} : h_j] + b)]$

additive attention

3×1

8×3

batch operation 4×8 $8\times 3 \rightarrow \boxed{4\times 3}$

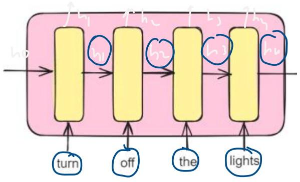
$\tan(4\times 3)$

$C_1 = \alpha_{11}h_1 + \alpha_{12}h_2 + \alpha_{13}h_3 + \alpha_{14}h_4$

$\alpha($ $s_0 = [e\ f\ g\ h]$

$[s_{i-1} : h_j]$

h_1 h_2 h_3 h_4



vector 4dim

$h_0 = [a\ b\ c\ d]$

$(4\times 3)(3\times 1)$

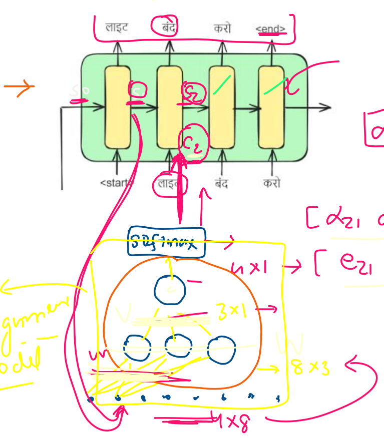
$\left[\dfrac{e^{e_{11}}}{e^{e_{11}} + e^{e_{12}} + e^{e_{13}} + e^{e_{14}}} \right.$ $(4\times 1) \rightarrow$ 4 numbers

4 rows / 8 cols 8cols

α 4 rows

s_{01}	s_{02}	s_{03}	s_{04}	h_{11}	h_{12}	h_{13}	h_{14}
s_{01}	s_{02}	s_{03}	s_{04}	h_{21}	h_{22}	h_{23}	h_{24}
s_{01}	s_{02}	s_{03}	s_{04}	h_{31}	h_{32}	h_{33}	h_{34}
s_{01}	s_{02}	s_{03}	s_{04}	h_{41}	h_{42}	h_{43}	h_{44}

true label



$C_2 = \alpha_{21}h_1 + \alpha_{22}h_2 + \alpha_{23}h_3 + \alpha_{24}h_4$

$\boxed{\alpha}$

$[\alpha_{21}\ \alpha_{22}\ \alpha_{23}\ \alpha_{24}]$

$4\times 1 \rightarrow [e_{21}\ e_{22}\ e_{23}\ e_{24}]$

4×3

s_{11}	s_{12}	s_{13}	s_{14}	h_{11}	h_{12}	h_{13}	h_{14}
s_{11}	s_{12}	s_{13}	s_{14}	h_{21}	h_{22}	h_{23}	h_{24}
s_{11}	s_{12}	s_{13}	s_{14}	h_{31}	h_{32}	h_{33}	h_{34}
s_{11}	s_{12}	s_{13}	s_{14}	h_{41}	h_{42}	h_{43}	h_{44}

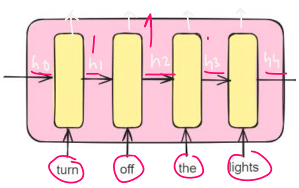
4×8

time distributed FNN

$$c_i = \sum \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1} \exp(e_{ik})}$$

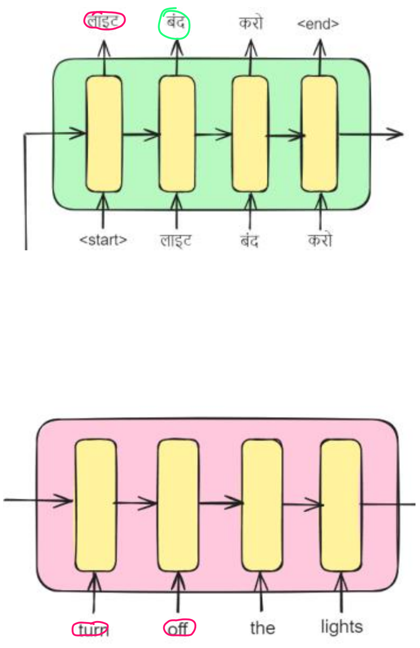
$$e_{ij} = V[\tanh(W[s_{i-1}; h_j] + b)]$$



additive attention

Luong Attention

17 January 2024 00:09



parameters → slow

$c_i = \sum \alpha_{ij} h_j$ → FFN $\{[V \tanh(W[s_{i-1} : h_j] + b)]\}$

$\alpha_{ij} = f(s_{i-1}, h_j)$ ×

$\alpha_{ij} = f(s_i, h_j)$ → $[s_i^T \cdot h_j]$ → dot product

fast

current

① diff

updated info

dynam

adjust

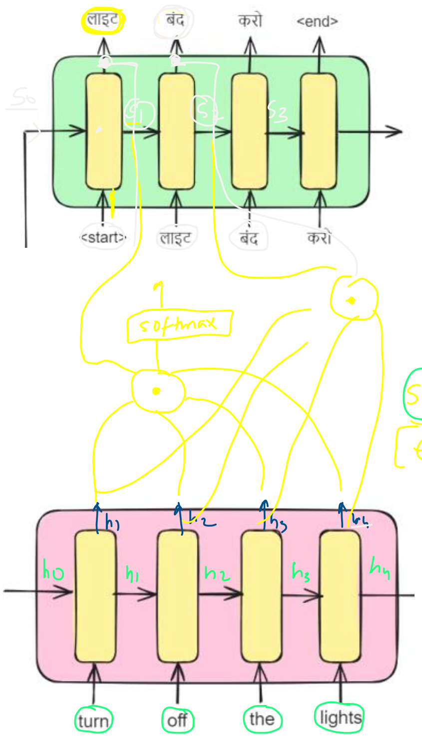
$s_i = [a\ b\ c\ d]$

$h_j = [e\ f\ g\ h]$

$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} [e\ f\ g\ h]$

$[ae + bf + cg + dh]$ → scalar → attention

α_{ij} ← softmax ← $\boxed{e_{ij}}$



output

multiplicative atten

$[e_{21}\ e_{22}\ e_{23}\ e_{24}]$ softm → $\alpha_{21}\ \alpha_{22}\ \alpha_{23}\ \alpha_{24}$ → c_2

$s_1 h_1$ $s_1 h_2$ $s_1 h_3$ $s_1 h_4$

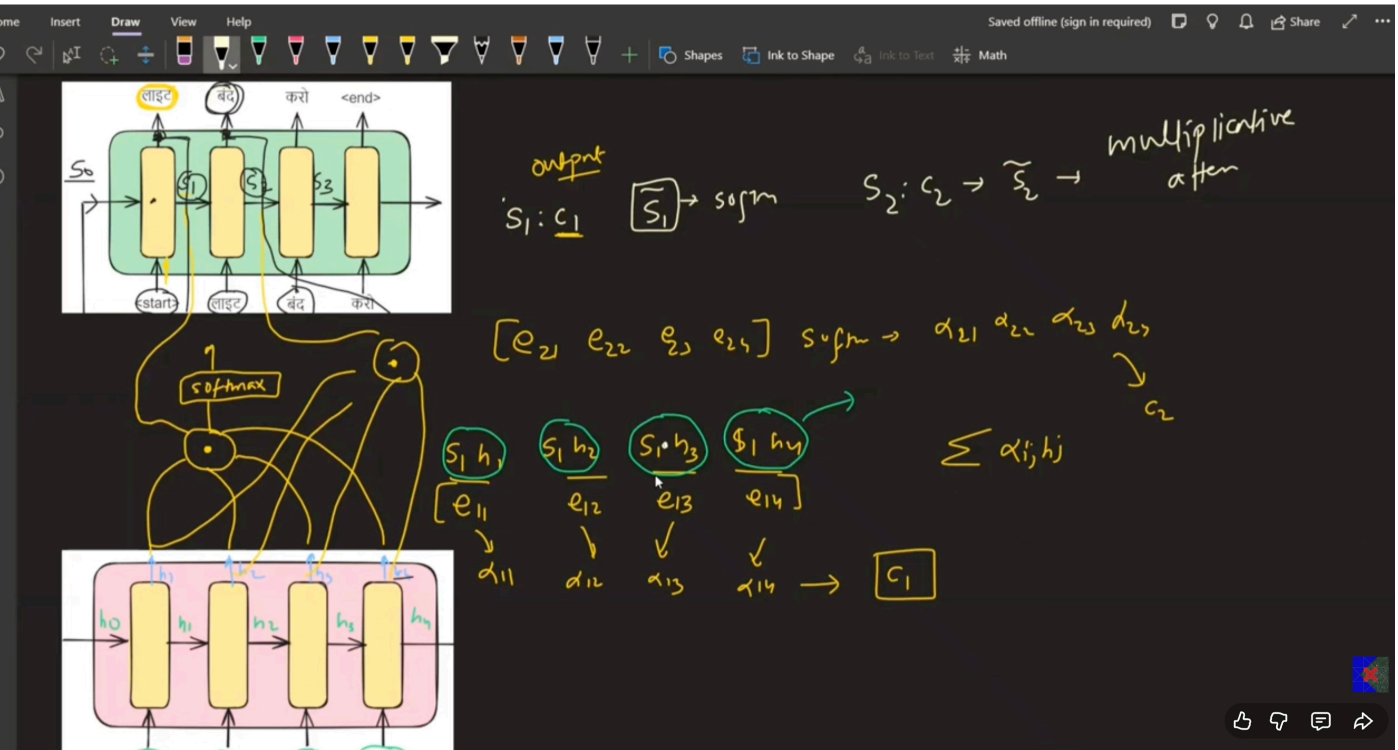
$[e_{11}\ e_{12}\ e_{13}\ e_{14}]$

$\alpha_{11}\ \alpha_{12}\ \alpha_{13}\ \alpha_{14}$ → $\boxed{c_1}$

$\sum \alpha_{ij} h_j$

Here in Luong , we made two changes , instead of providing context vector in input we are providing it in output , also it changes how alignment score is calculated now instead of neural network we are using dot product , also we are using current decoder hidden state and input to calculate (which is more dynamic)

Due to dot product involvement it is also called multiplicative attention.



Insert
Draw
View
Help
Shapes
Ink to Shape
Ink to Text
Math
Saved offline (sign in required)
Share
लाइट
बंद
करो
<end>
<start>
softmax
output
s1 : c1
s1 → softm
s2 : c2 → s2 →
multiplicative atten
[e21 e22 e23 e24] softm → α21 α22 α23 α24
c2
s1 h1
s1 h2
s1·h3
s1 h4
Σ αij hj
[e11 e12 e13 e14]
α11 α12 α13 α14 → c1