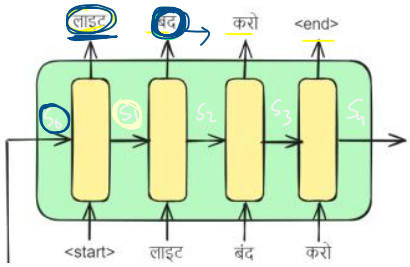
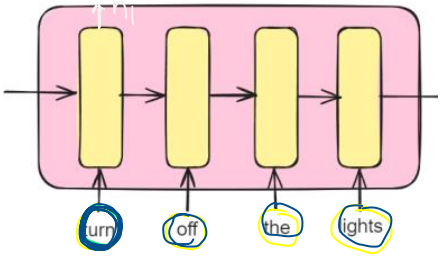
Bahdanau Attention

16 January 2024 16:11





given

alignment score

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

α alignment

α_{11} → लाइट → turn

α_{12} → लाइट → off

decoder → prev hidden state

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

$\alpha_{ij} = f(h_j, s_{i-1})$ → main funct → the encoder hidden

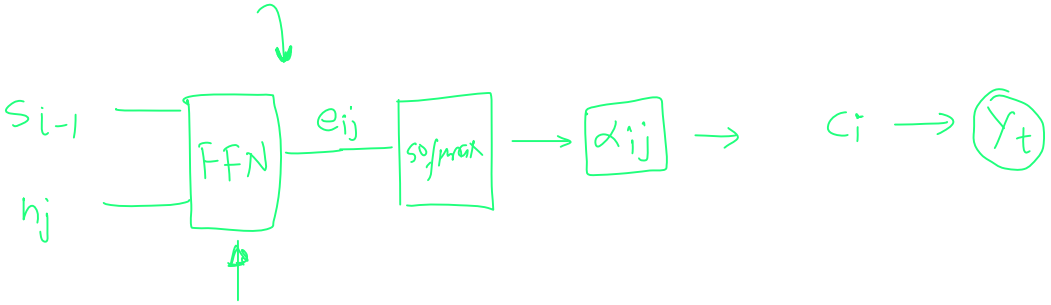
approximate

α_{ij} → FFN → ANN → UFA

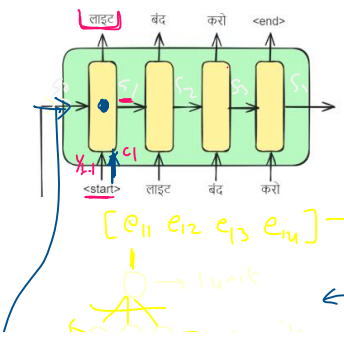
softmax

e_{ij}

s_{i-1} h_j



s_0 y_{t-1} c_1 → dshm → y_t (लाइट) $[s_1]$



$i=1$ $t=2$

encoder sentence

h_1 h_2 h_3 h_4

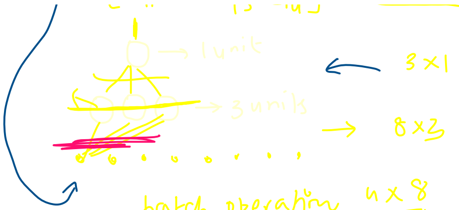
s_{i-1} h_j

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

$c_1 = \sum \alpha_{1j} h_j$

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

3×1



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

3×1

8×3

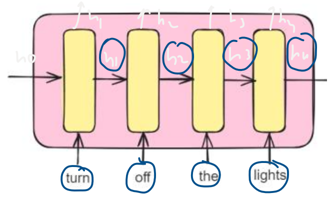
batch operation 4×8 8×3 → 4×3

tan(4×3)

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

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

h_1 h_2 h_3 h_4



vectors 4dim

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

4 rows / 8 cols

8cols

α(4rows

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}

(4×3)(3×1)

$$\frac{e^{e_{11}}}{e^{e_{11}} + e^{e_{12}} + e^{e_{13}} + e^{e_{14}}}$$

(4×1) → 4 numbers

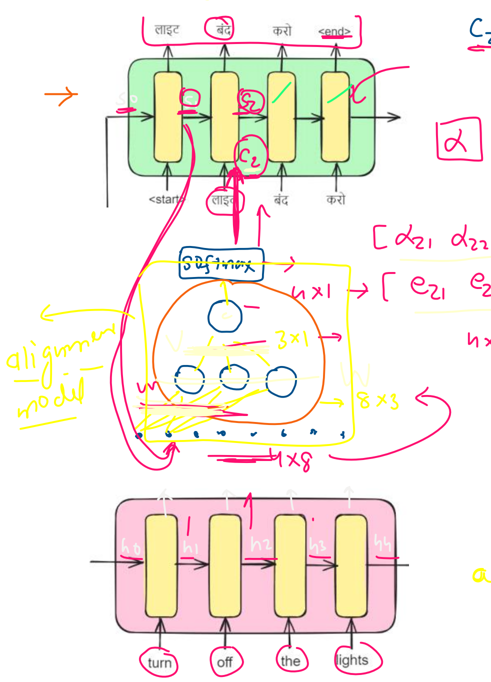
true label

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

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



α

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

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

4×3

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

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

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

additive attention

Luong Attention

17 January 2024 00:09