Attention – Formal Notation

Inputs:

decoder state s_i

encoder states $h_j = \left[\overrightarrow{h_j}; \overleftarrow{h_j}\right] \quad \forall i = 1 \dots T_x$

where $\overrightarrow{h_j} = \mathsf{RNN}_{\mathsf{enc}}(h_{j-1}, x_j) = \tanh(U_e \overrightarrow{h_{j-1}} + W_e E_e x_j + b_e)$

Attention energies: $e_{ij} = v_a^\top \tanh\left(W_a s_{i-1} + U_a h_j + b_a\right)$

Attention distribution: $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$

Context vector: $c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$