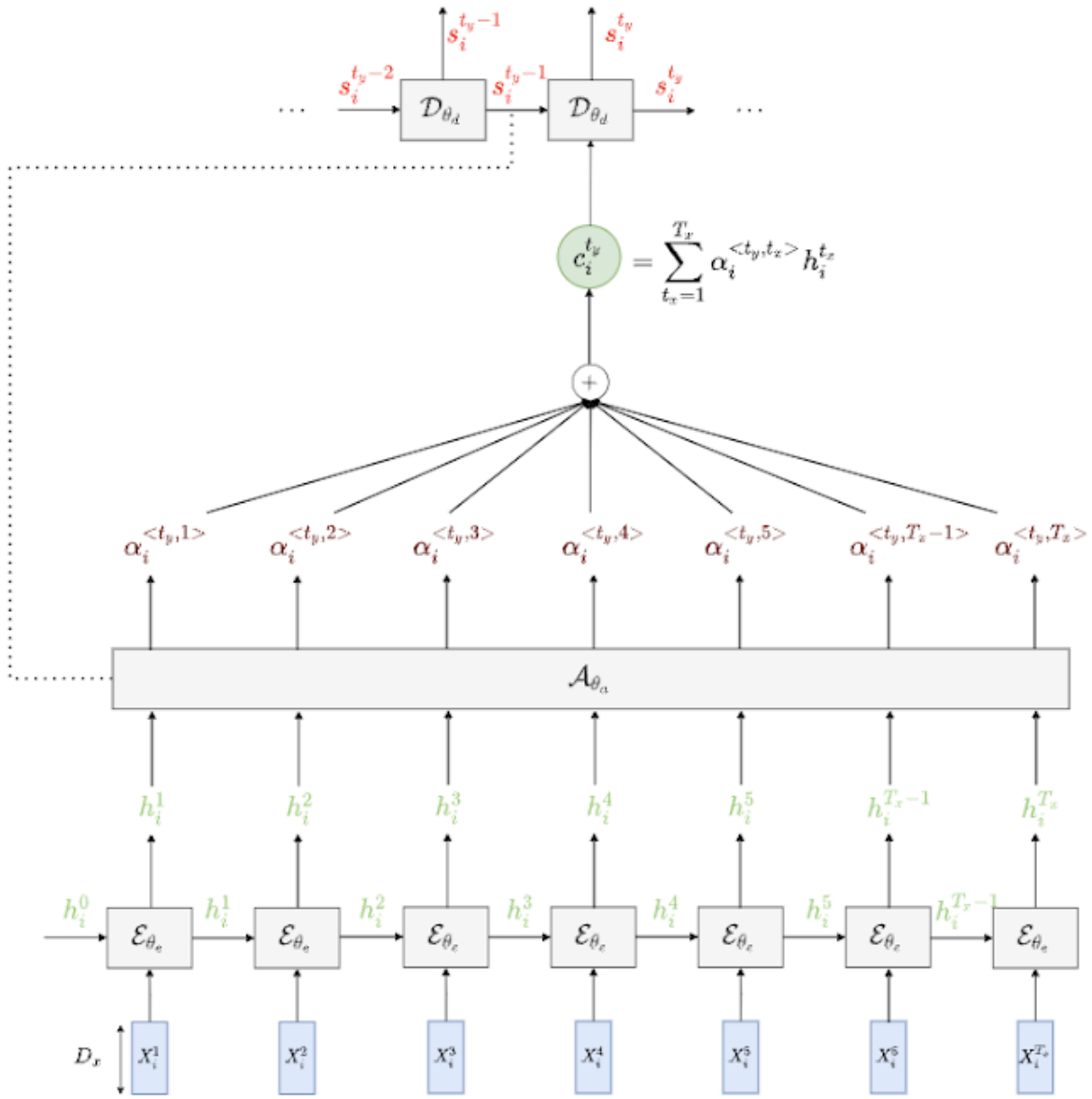
Attention Weights

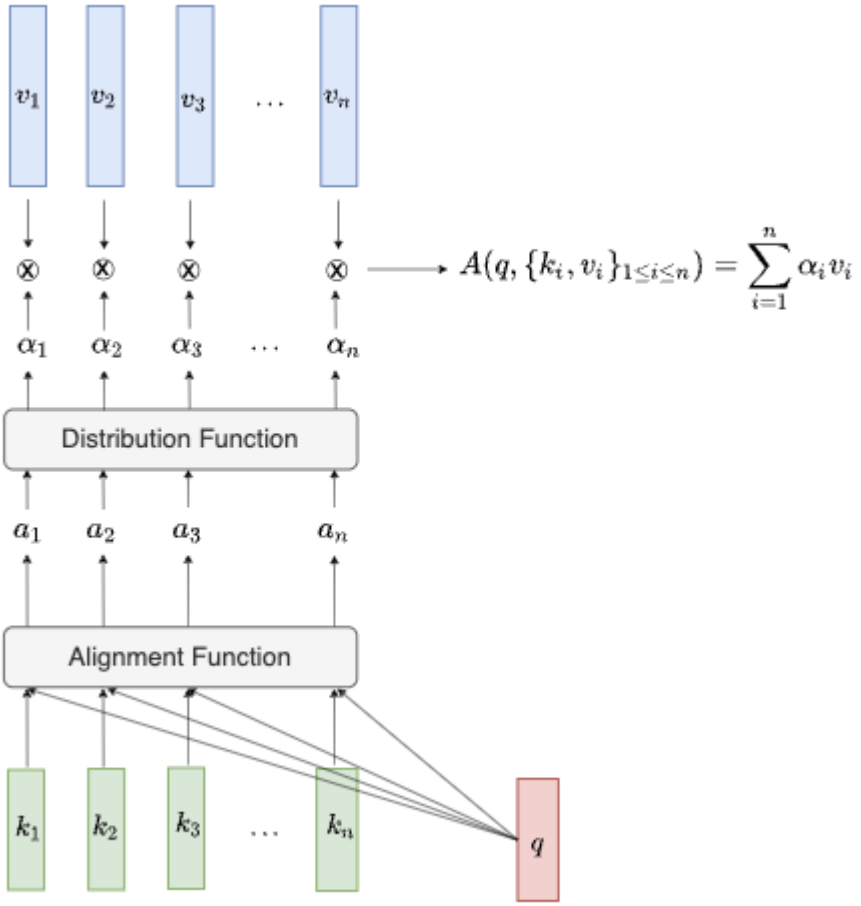
Name

HM

The following figure represents the Sequence to Sequence Model with Attention Mechanisms (S2SWA)



The following figure represents the soft query retrieval problem (SQRP):



Let us compare the context vector in (S2SWA) and the attention vector in (SQRP):

$$A(q, \{k_i, v_i\}_{1 \leq i \leq n}) = \sum_{i=1}^{n} \alpha_i v_i$$

$$c_i^{t_y} = \sum_{t_x=1}^{T_x} \alpha_i^{<t_y, t_x>} h_i^{t_x}$$

Which elements of the architecture (S2SWA) represent the values (v_i)_i in the (SQRP) ? 2 points

(a) $(h_i^{t_x})_{1 \leq t_x \leq T_x}$

(b) $(h_i^{t_x})_{1 \leq i \leq N}$

(c) $(s_i^{t_y})_{1 \leq t_y \leq T_y}$

- [x] (a)
- [ ] (b)
- [ ] (c)

Let us use the dot product as an "alignment function" and the Softmax as a "distribution function" in the (SQRP). Which equation is correct ? 2 points

(a) $$\alpha_i = \frac{q.k_i}{\sum_{j=1}^{n} q.k_j}$$

(b) $$\alpha_i = \frac{\exp(q.k_i)}{\sum_{j=1}^{n} \exp(q.k_j)}$$

(c) $$\alpha_i = q.k_i$$

- (a)
- (b)
- (c)

In the (S2SWA) architecture 2 points

What is the interpretation of $\alpha_i^{<t_y,t_x>}$?

(a) The weight associated with the hidden state $h_i^{t_x}$ to generate the context vector $c_i^{t_y}$

(b) The weight associated with the hidden state $s_i^{t_y}$ to generate the context vector $c_i^{t_y}$

(c) The weight associated with the hidden state $h_i^{T_x}$ to generate the context vector $c_i^{t_y}$

- [x] (a)
- [ ] (b)
- [ ] (c)

Which element in the (S2SWA) could be a good candidate to represent the query in the (SQRP) ? 2 points

(a) $s_i^{t_y}$

(b) $s_i^{t_y-1}$

(c) $h_i^{t_x}$

- [ ] (a)
- [x] (b)
- [ ] (c)

Which elements in the (S2SWA) could be good candidates to represent the keys in the (SQRP) ? 2 points

(a) $(s_i^{t_y})_{1 \leq t_y \leq T_y}$

(b) $(h_i^{t_x})_{1 \leq t_x \leq T_x}$

(c) $(h_i^{t_x})_{1 \leq i \leq N}$

- [ ] (a)
- [x] (b)
- [ ] (c)