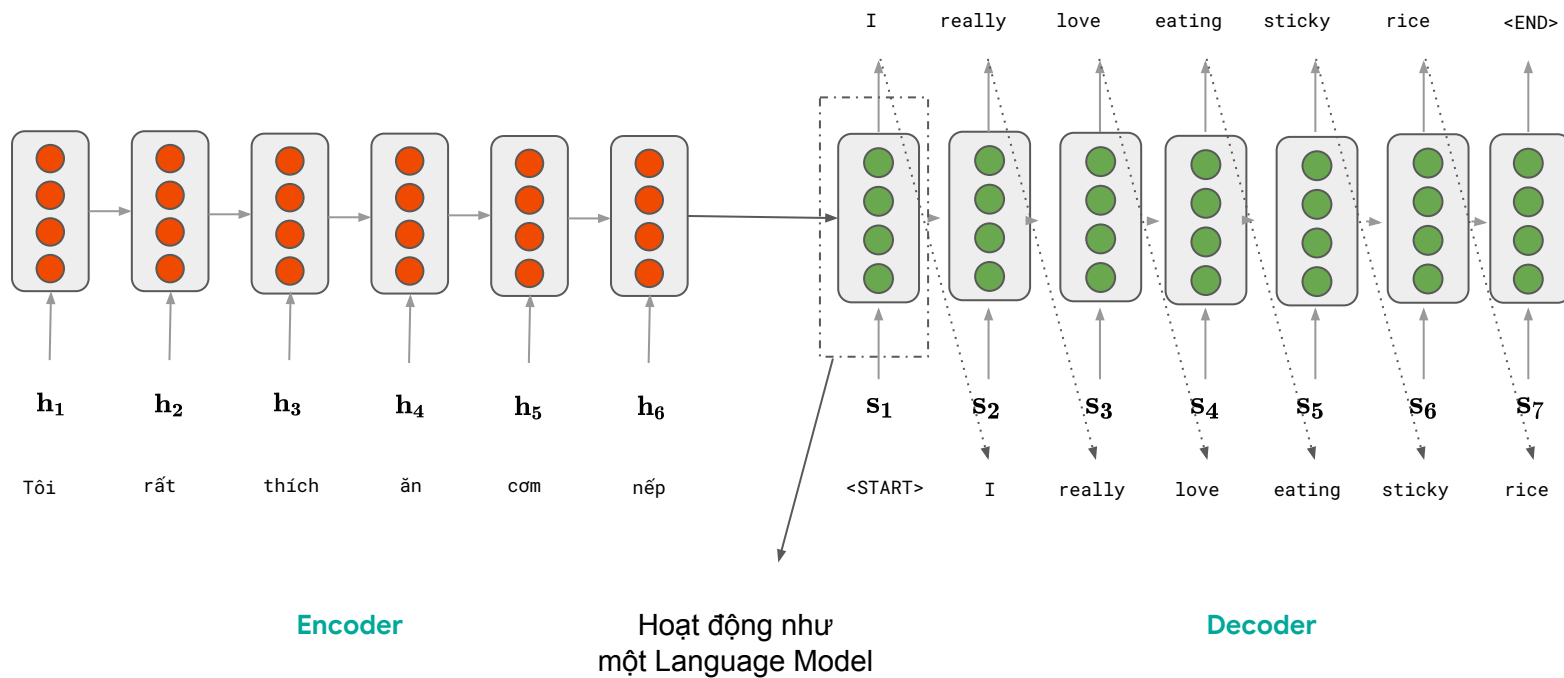


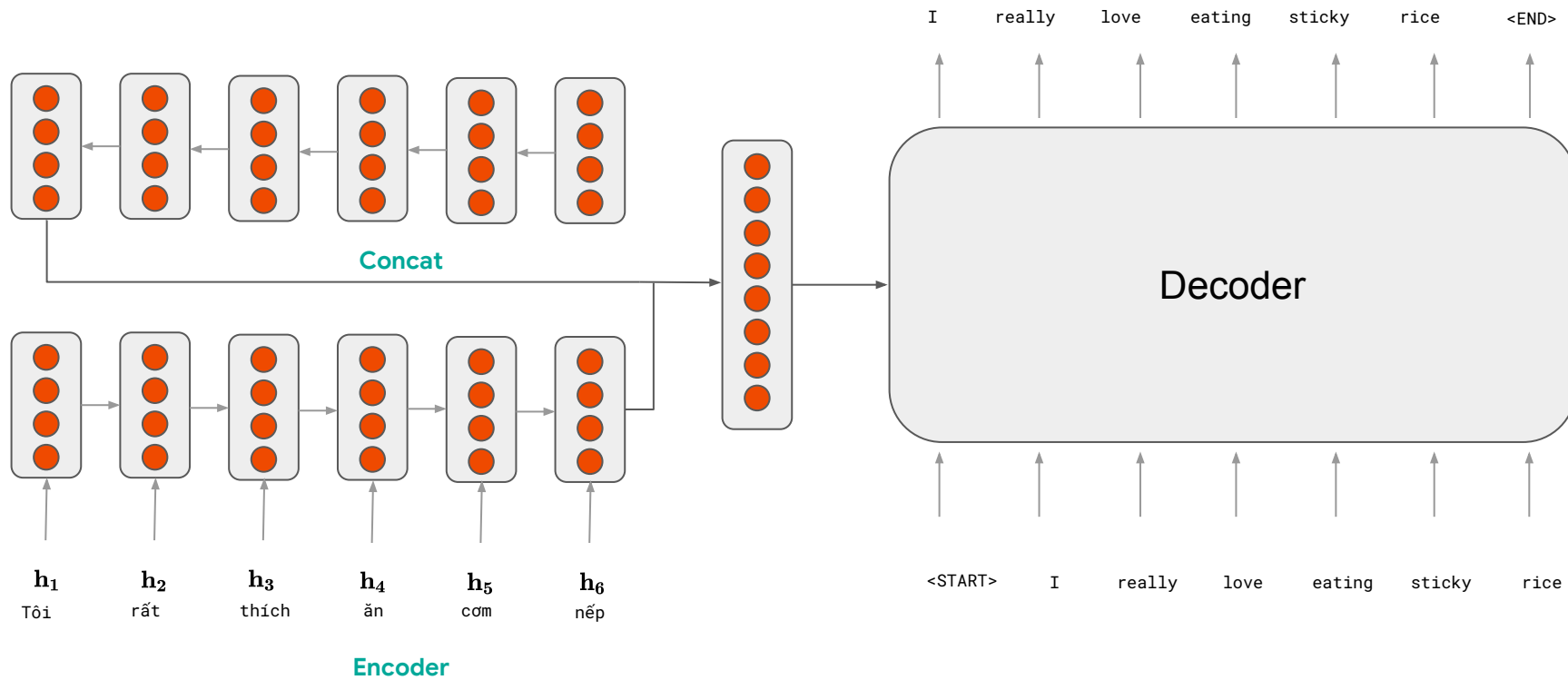
Neural Machine Translation

NMT

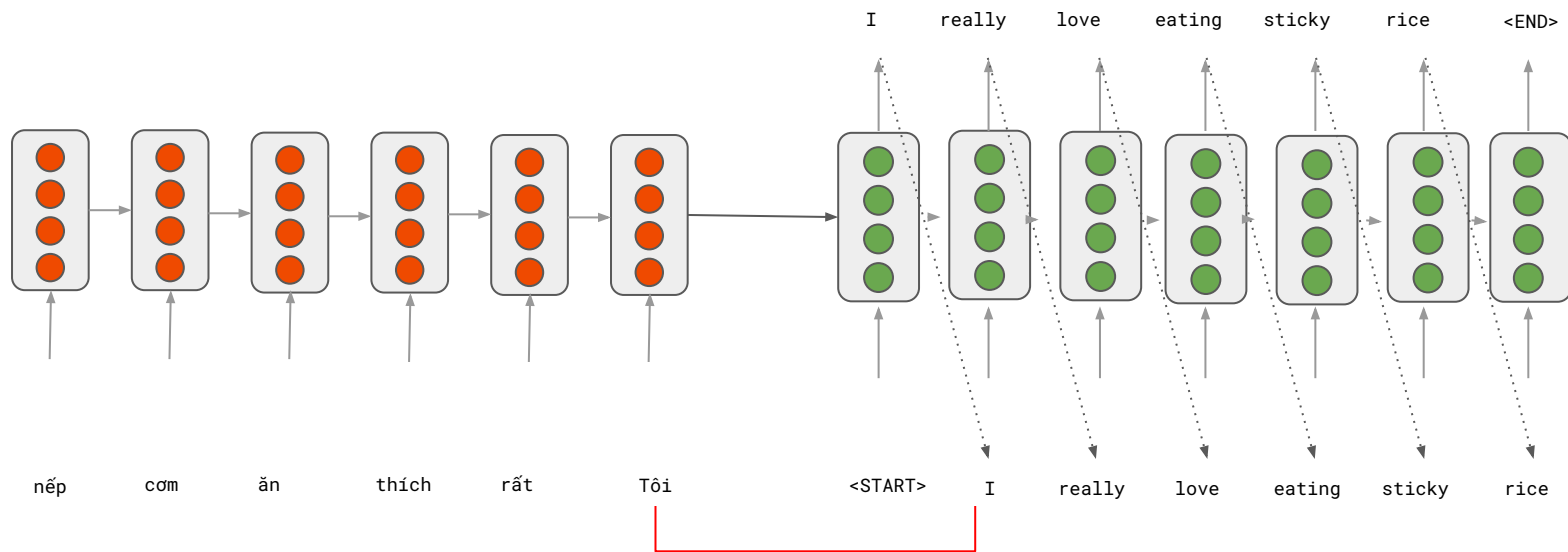
Seq2Seq



Deep RNN



Input sequence in reverse



Từ cuối cùng encoder xử lý sẽ là từ đầu tiên mà decoder sinh ra.

Điều này giúp decoder sinh ra từ đúng hơn.

Attention is efficient

- 1 Cho phép Decoder tập trung vào những phần cụ thể tại câu đầu vào. Hoàn toàn tự động alignment giữa các từ đầu ra và đầu vào
- 2 Giải quyết vấn đề information bottleneck
- 3 Giúp flow gradient tốt hơn → Cải thiện vấn đề Gradient Vanishing

Dot Product Attention



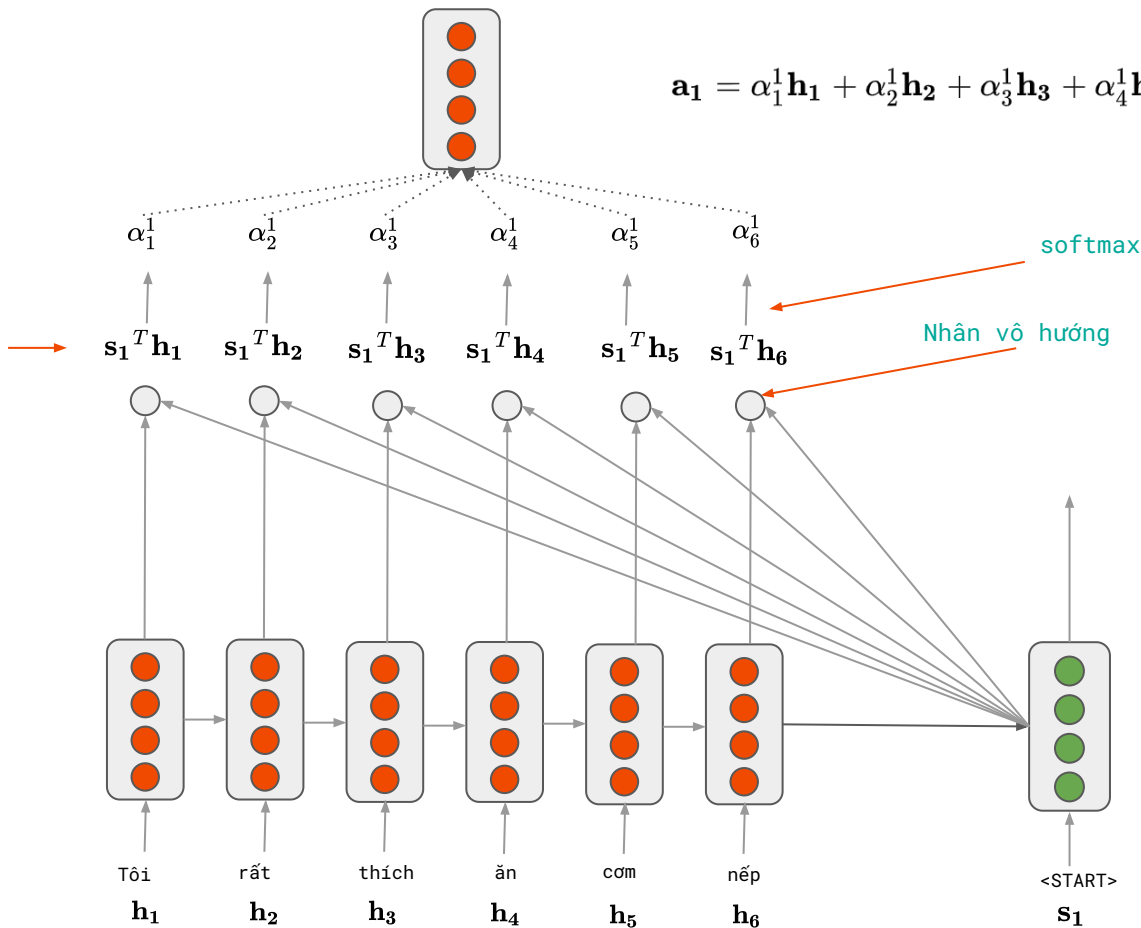
VietAI

Phân phối
Attention

Điểm
Attention

α^1

e^1



softmax

Nhân vô hướng

$$\alpha^1 = \text{softmax}(e^1) \quad \alpha^1 \in \mathbb{R}^6$$

Dot Product Attention



VietAI

concat

$$\alpha^1 \in \mathbb{R}^6$$

$$\mathbf{a}_1 \in \mathbb{R}^h$$

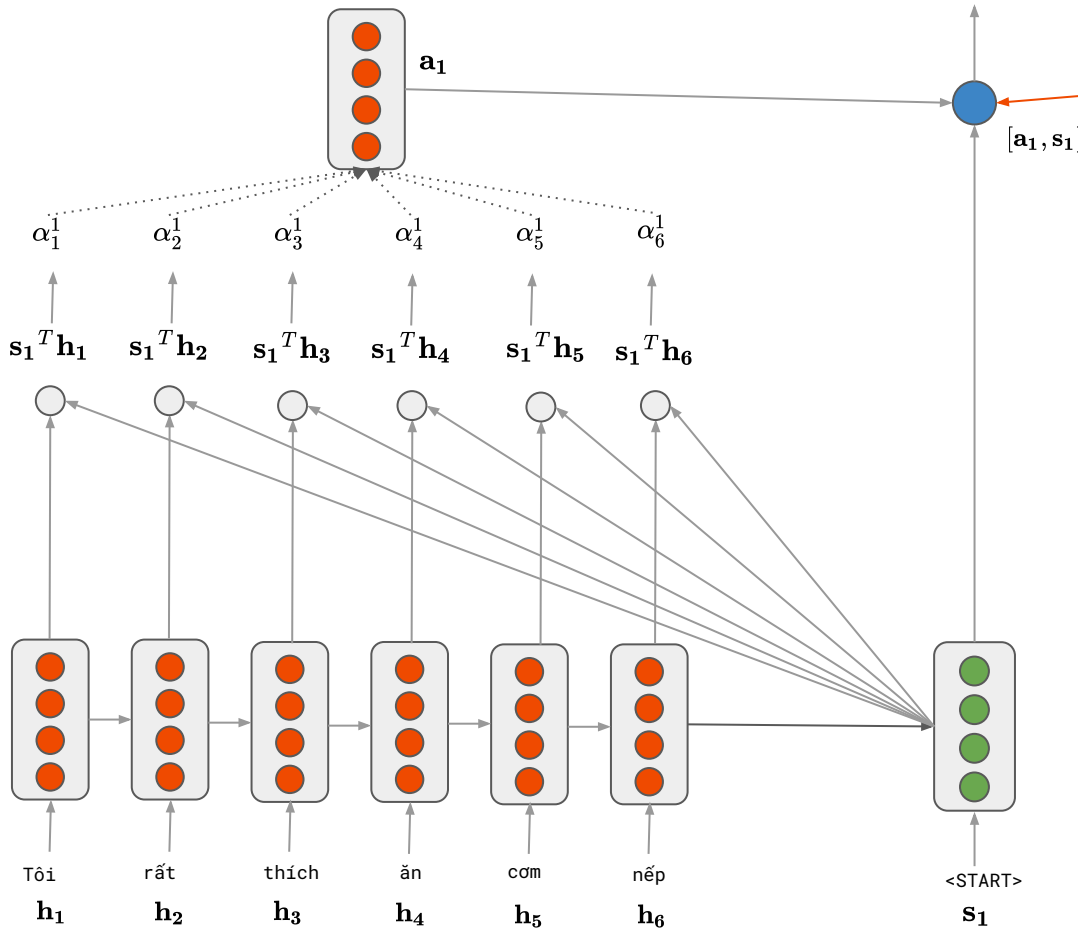
$$[\mathbf{a}_1, \mathbf{s}_1] \in \mathbb{R}^{2h}$$

Phân phối
Attention

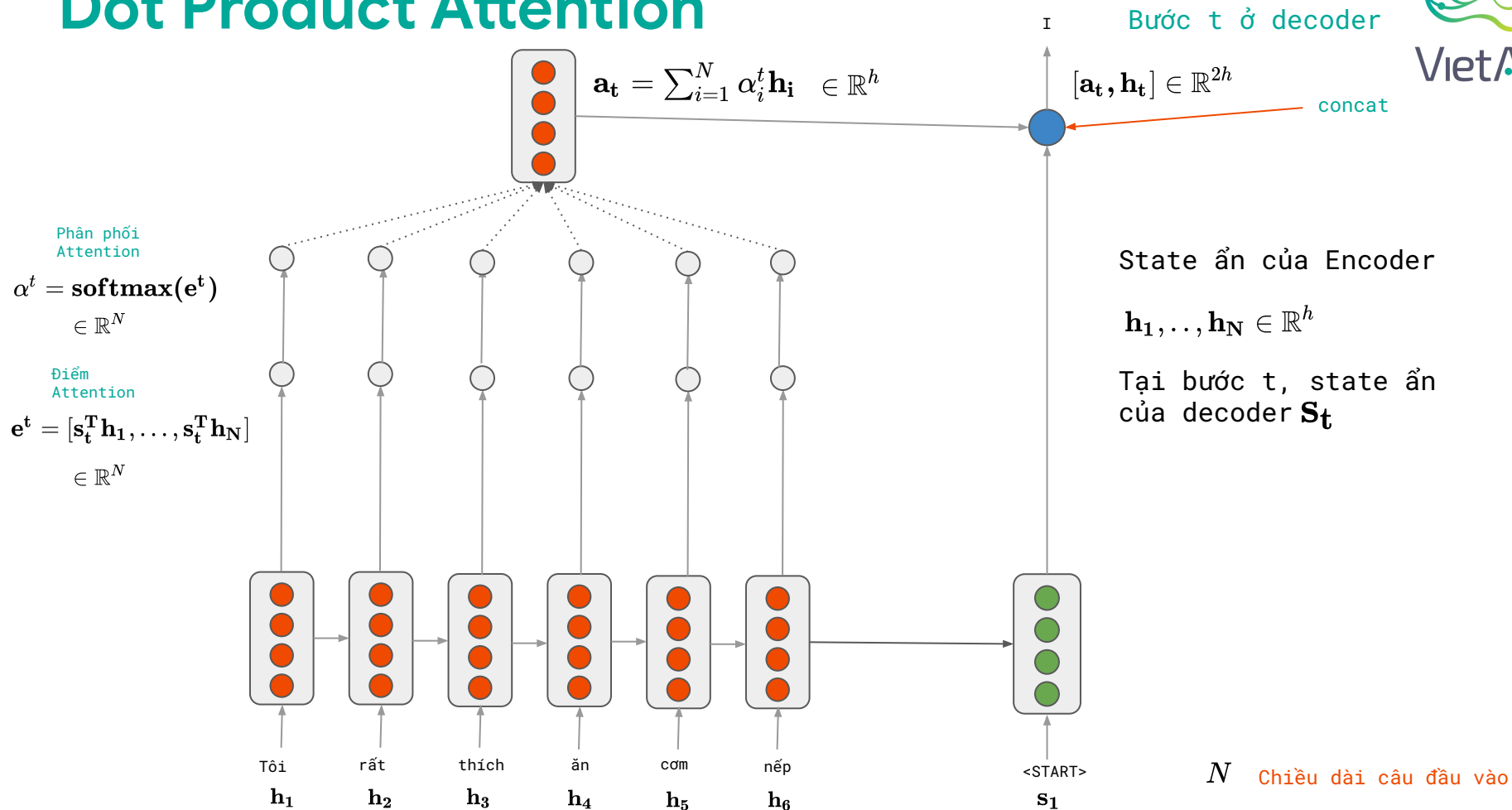
Điểm
Attention

α^1

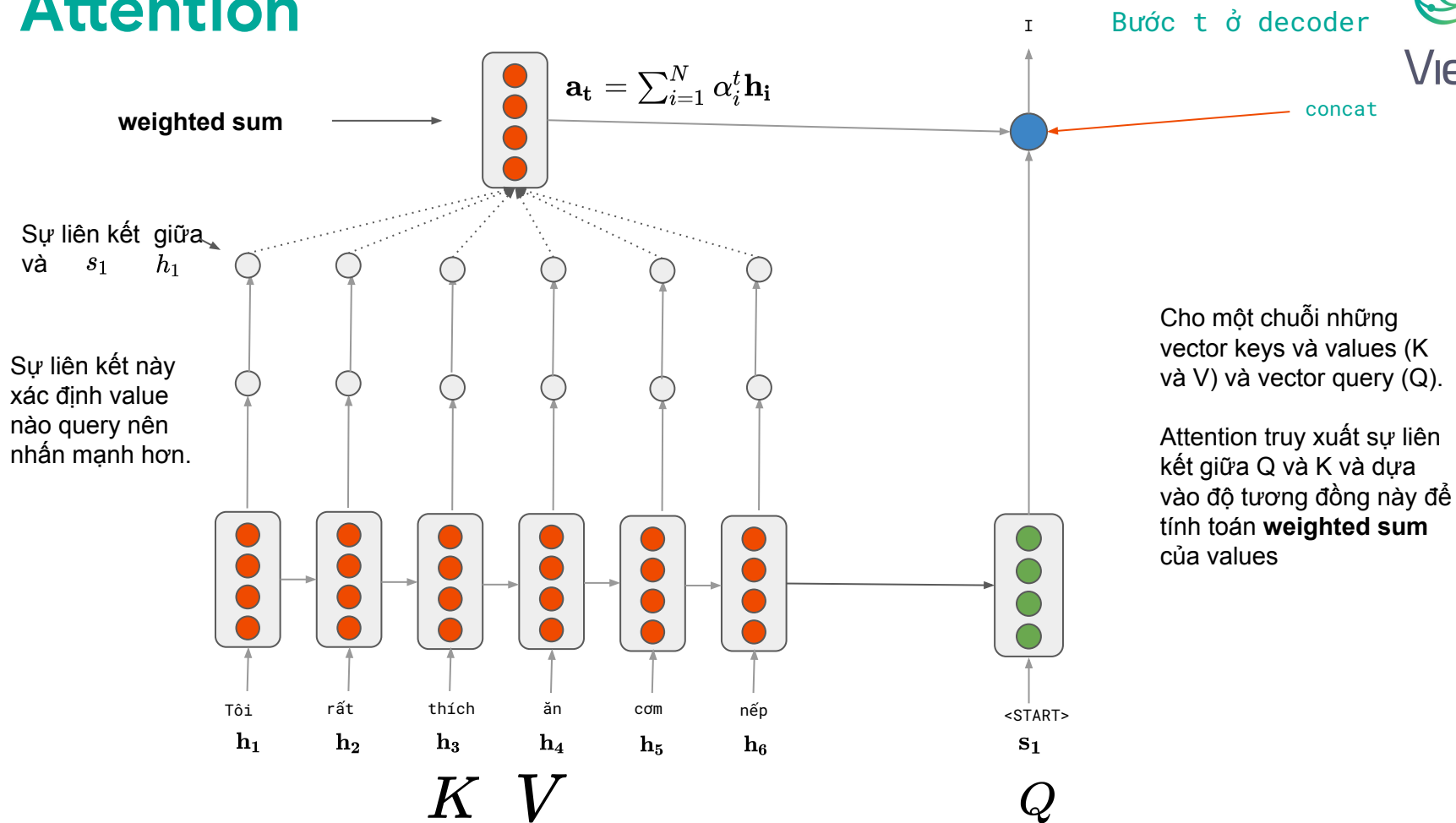
\mathbf{e}^1



Dot Product Attention



Attention



Attention Variants

Encoder Hidden State $h_i \in \mathbb{R}^{d_1}$ Decoder Hidden State $s_j \in \mathbb{R}^{d_2}$

Basic Dot-Product Attention

$$e_i = s_j^T h_i \quad e_i \in \mathbb{R}$$

Yêu cầu: $d_1 = d_2$

Tăng khả năng học sự liên kết

Multiplicative Attention

$$e_i = s_j^T W h_i \quad e_i \in \mathbb{R}$$

$$W \in \mathbb{R}^{d_2 \times d_1}$$

Additive Attention

Neural Network

$$e_i = v^T \tanh(W_1 h_i + W_2 s_j) \quad e_i \in \mathbb{R}$$

$$\begin{aligned} W_1 &\in \mathbb{R}^{d_3 \times d_1} \\ W_2 &\in \mathbb{R}^{d_3 \times d_2} \\ v &\in \mathbb{R}^{d_3} \end{aligned}$$

[1409.0473] Neural Machine Translation by Jointly Learning to Align and Translate (arxiv.org)

1508.04025.pdf (arxiv.org)

BLEU Score

$$P_n = \frac{\sum Count_{clip}(n-gram)}{\sum Count(n-gram)}$$

We compute the brevity penalty BP,

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}.$$

Then,

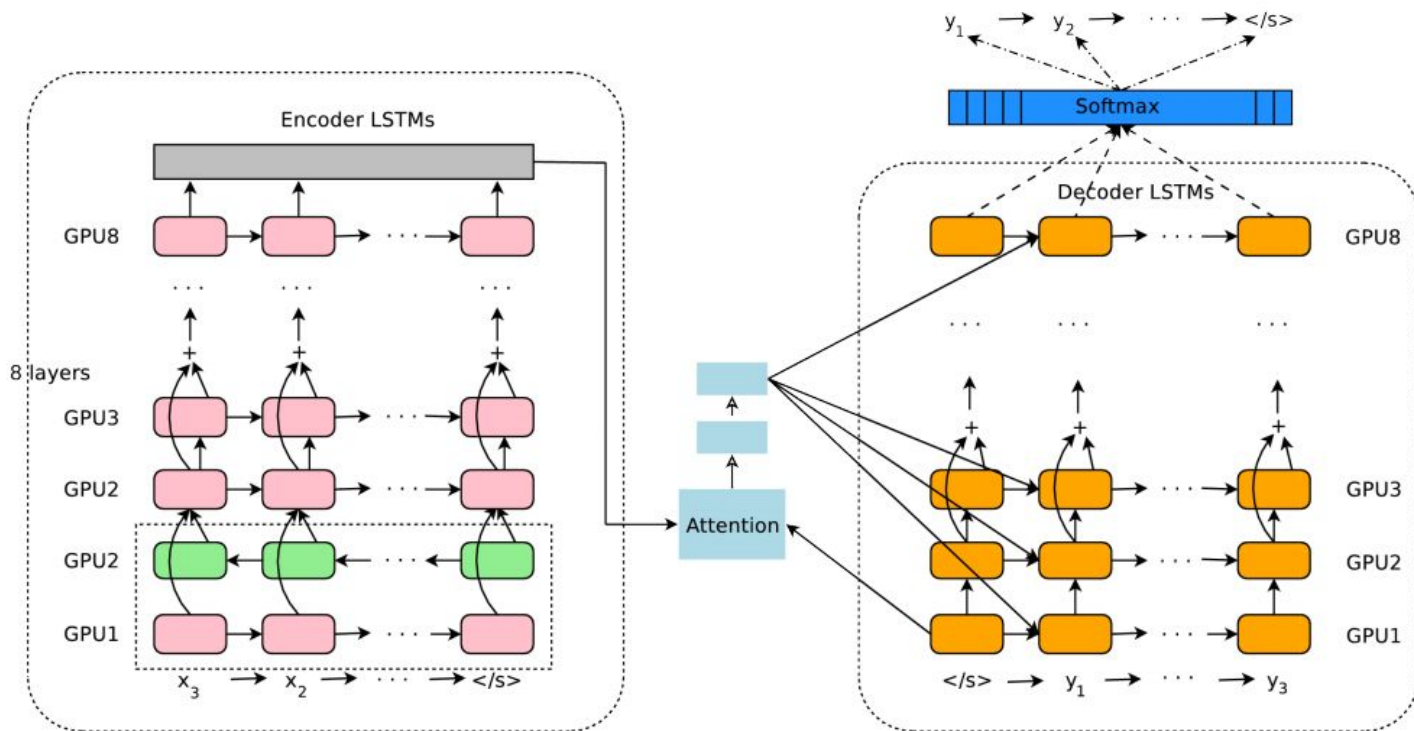
$$BLEU = BP \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right).$$

The ranking behavior is more immediately apparent in the log domain,

$$\log BLEU = \min(1 - \frac{r}{c}, 0) + \sum_{n=1}^N w_n \log p_n.$$

In our baseline, we use $N = 4$ and uniform weights $w_n = 1/N$.

Google Neural Machine Translation 2016



Google chuyển từ SMT (Statistical machine translation) sang NMT từ năm 2016.

<https://arxiv.org/pdf/1609.08144v2.pdf>

NMT Advantages

- 1 End-to-end models, giải quyết được vấn đề với câu đầu vào có độ dài lớn.
- 2 Tiết kiệm nguồn lực con người trong quá trình xây dựng, bỏ qua feature engineering
- 3 Một công thức cho tất cả các cặp ngôn ngữ khác nhau

NMT Problems

- 1 Tốn kém trong việc training: thời gian + sức mạnh máy tính
- 2 Gặp vấn đề trong việc dịch các từ hiếm (**rare words**)
- 3 Đôi khi NMT dịch chỉ một phần đầu vào, và có những kết quả không như mong muốn, khó debug.