Sequence Model

Jie Wei

Xi'an Jiaotong University

2021-11

Contents

西安交通大學
XI'AN JIAOTONG UNIVERSITY

Recurrent Neural Network

01

Examples of sequence data

Text Sentence XJTU is a C9 League university located in Xi’an.

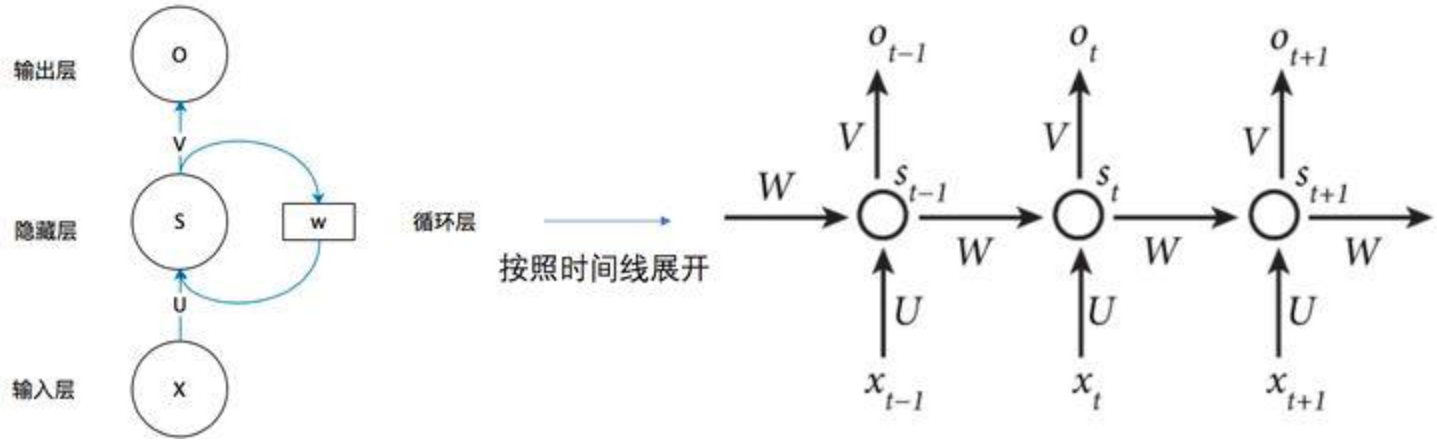
Audio

Video

Why RNN?

Temporal relationship learning → Contextual information

1.2 RNN Structure



X: input

S: hidden-layer output

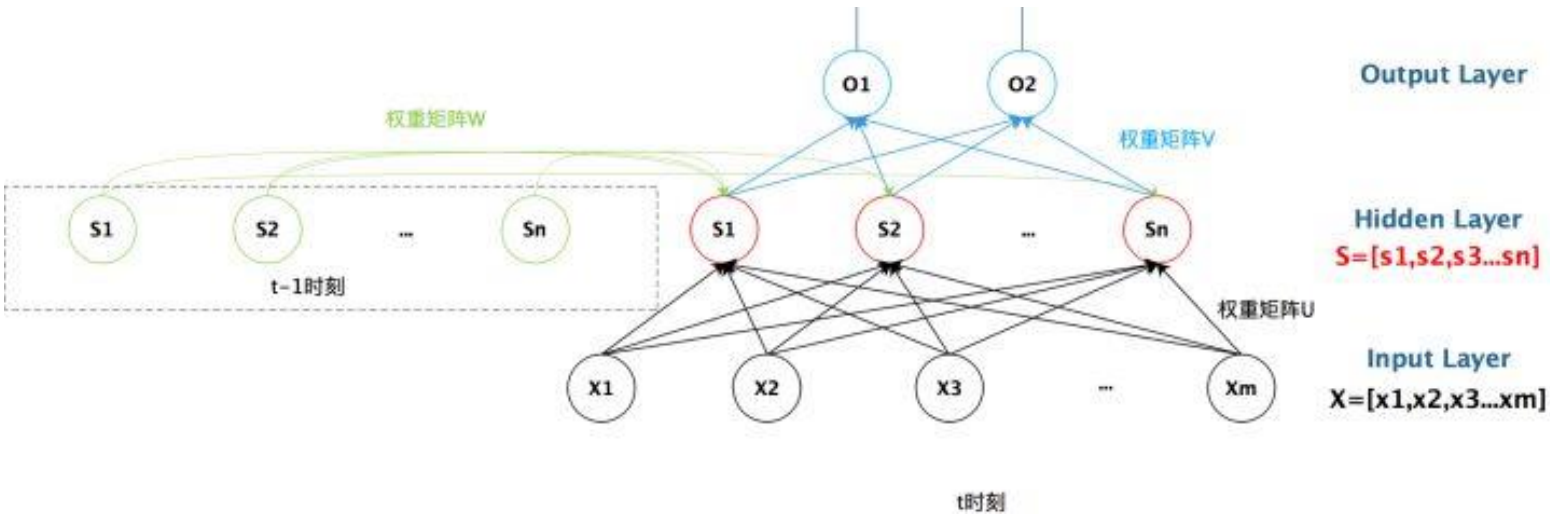
O: output-layer output

U: input → hidden weight matrix

W: t-1 hidden → t hidden weight matrix

V: hidden → output weight matrix

1.2 RNN Structure

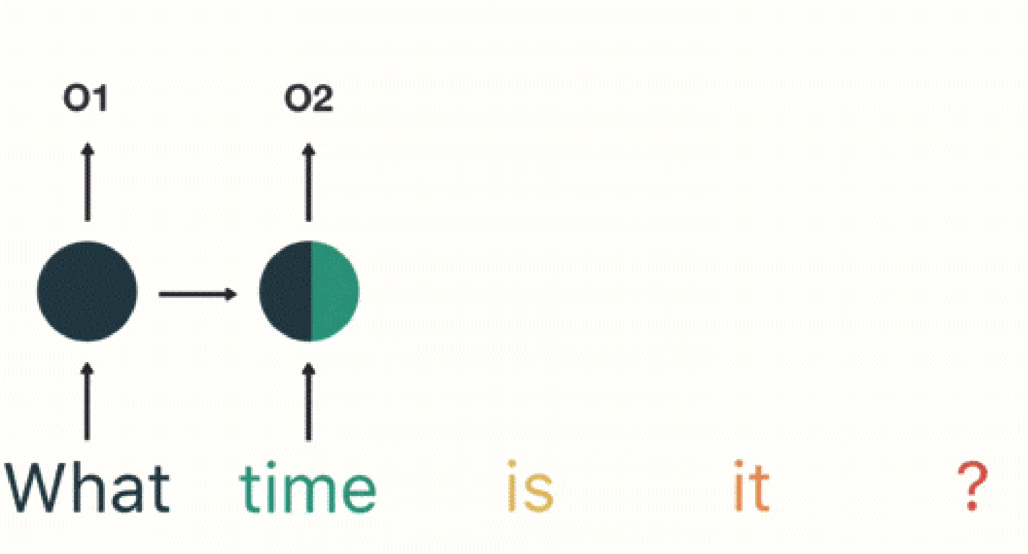


Formula:

$$O_t = g(V \cdot S_t)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1})$$

1.3 Variants



Disadvantages:

Gradient vanishing problems.

It cannot process very lengthy sequences.

LSTM & GRU:

only preserves important and relevant information

西安交通大學
XI'AN JIAOTONG UNIVERSITY

02 Sequence-to-sequence Learning

2.1 Task

Analysis and Recognition

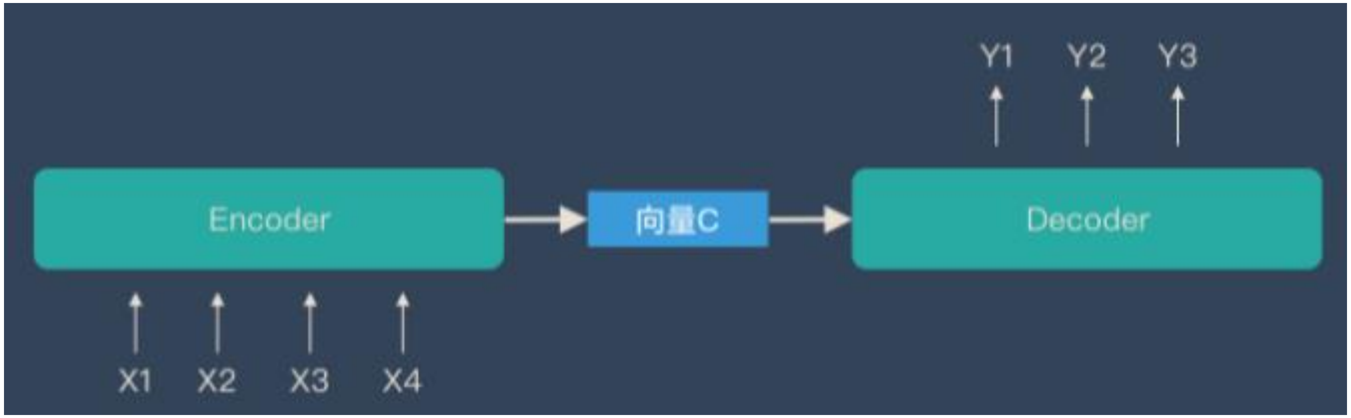
Analysis and Generation

Input / Output	Single	Sequence
Single	\	Image Description Music Generation
Sequence	Sentiment Classification Video Activity Recognition	Speech Recognition Machine Translation

RNN exist problem:

the length of output remains the same as input sequence

2.2 Seq2seq



Input: a sequence

Output: a sequence

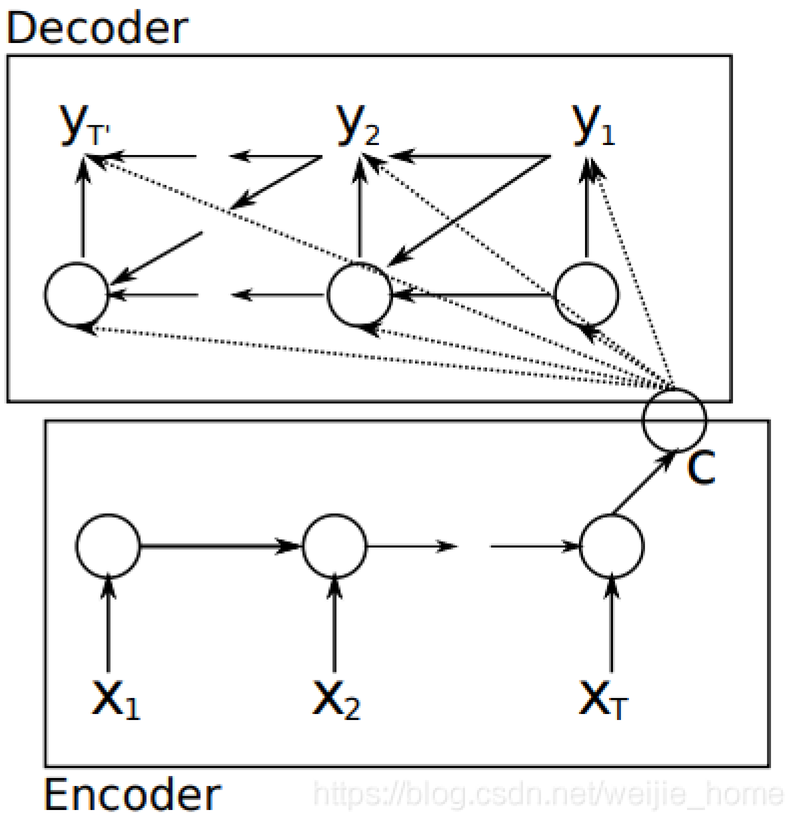
** The length of the input and output sequences is variable

Encoder & Decoder

Encoder: a sequence → context vector

Decoder: context vector → a sequence

Cho RNN Encoder-Decoder [1]



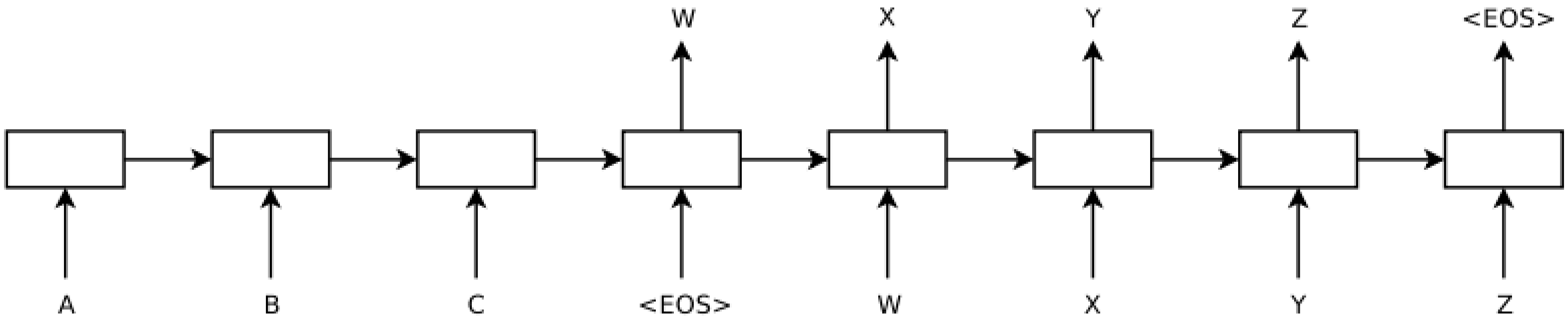
$$h_t = tanh(W[h_{t-1}, y_{t-1}, c] + b)$$
$$o_t = softmax(Vh_t + b)$$

$$h_t = tanh(W[h_{t-1}, x_t] + b)$$
$$o_t = softmax(Vh_t + b)$$
$$c = tanh(Uh_T)$$

[1] https://arxiv.org/pdf/1406.1078.pdf

2.2 Seq2seq

Sutskever Encoder-Decoder [1]



Encoder:

$$h_t = tanh(W[h_{t-1}, x_t] + b)$$
$$o_t = softmax(Vh_t + b)$$
$$c = tanh(Uh_T)$$

Decoder:

$$h_t = tanh(W[h_{t-1}, y_{t-1}] + b)$$
$$o_t = softmax(Vh_t + b)$$
$$h_0 = c$$

[1] https://arxiv.org/pdf/1409.3215.pdf

西安交通大學
XI'AN JIAOTONG UNIVERSITY

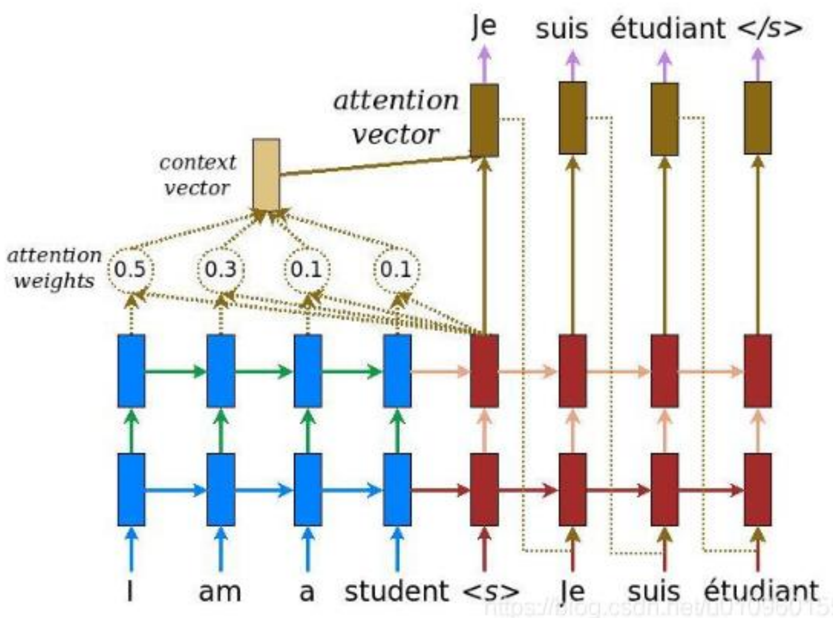
Attention Mechanism

03

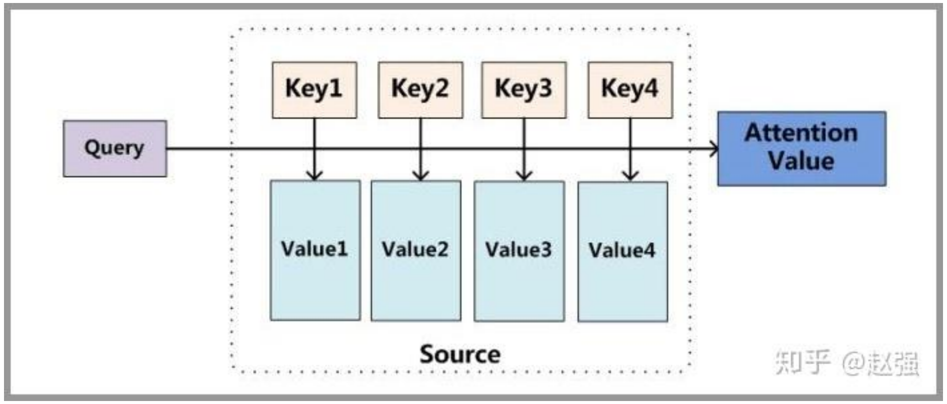
3.1 Motivation

Leverage the complete information from Encoder

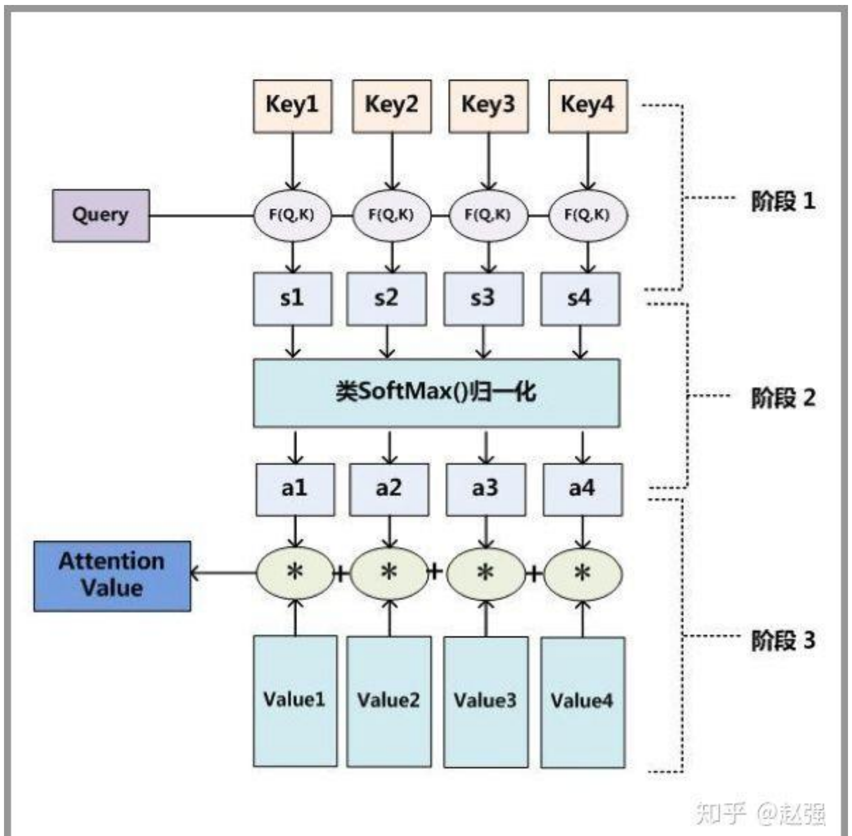
Same scene, different people with different attention



3.2 Principle



Key
Value
Query

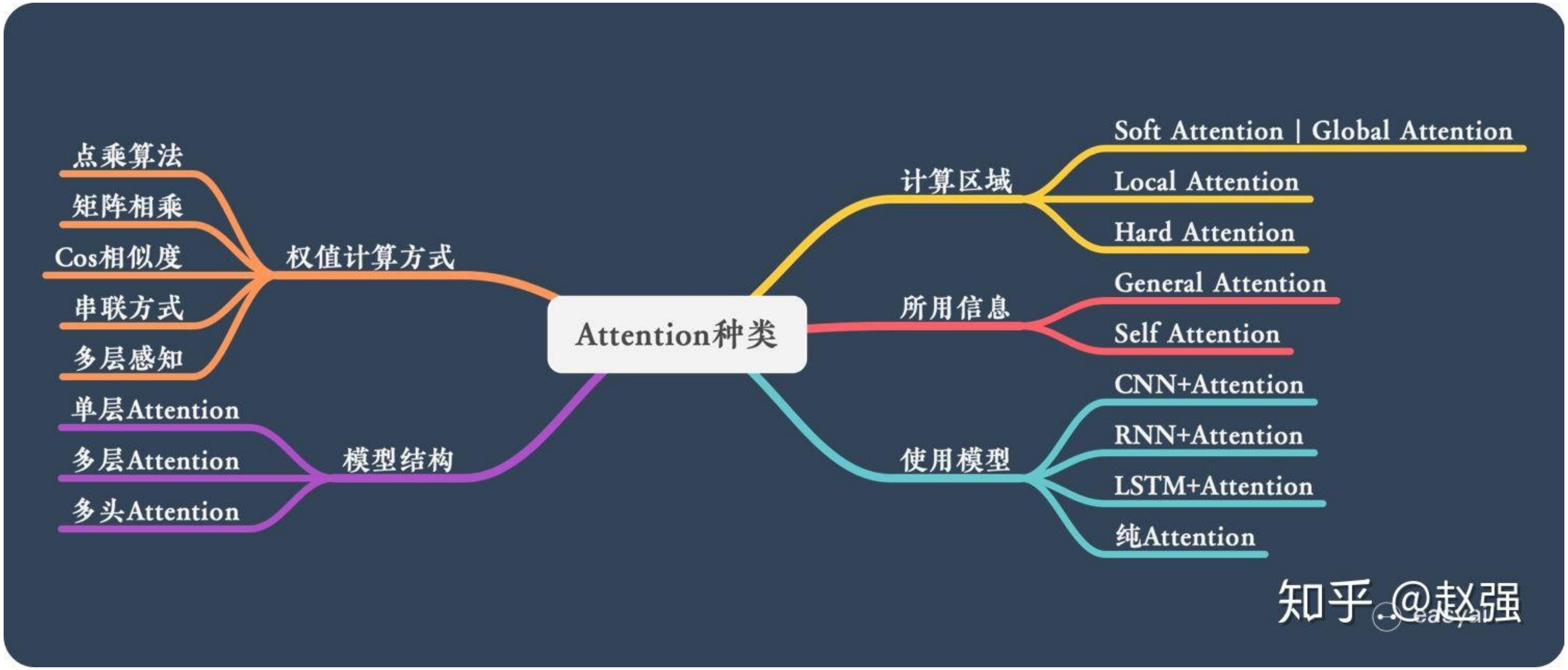


$$s\left(q_t, k_s\right) = W[q_t, k_s]$$

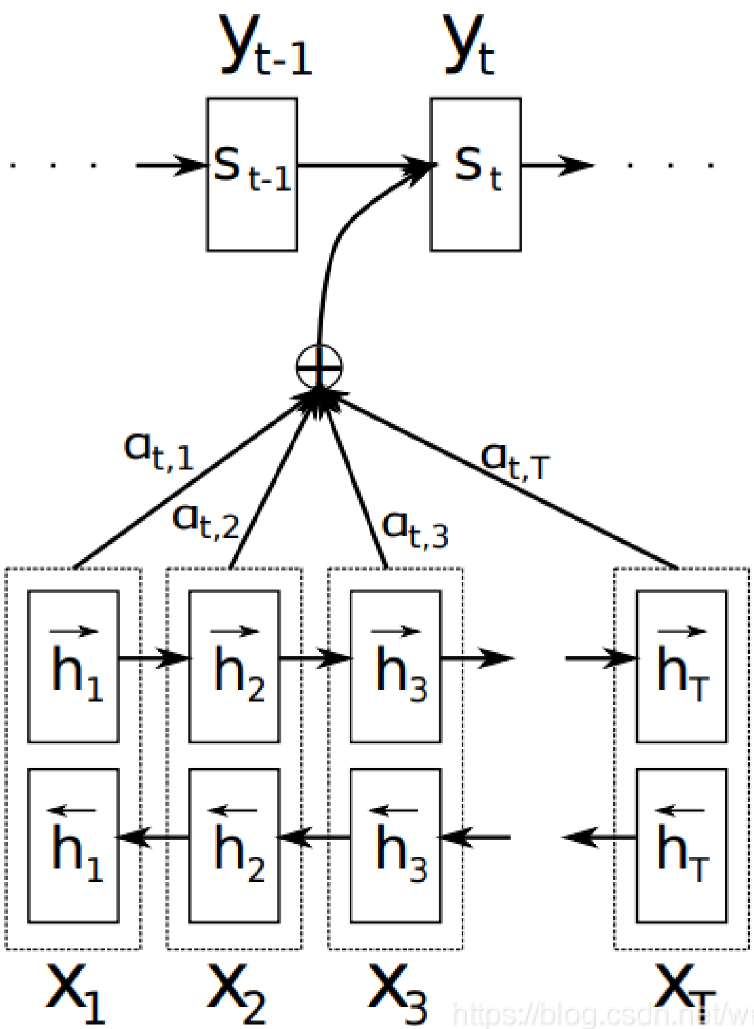
$$a(q_t, k_s) = \frac{exp(s(q_t, k_s))}{\sum_{i=1}^{N} exp(s(q_t, k_i))}$$

$$Attention(q_t, K, V) = \sum_{s=1}^{m} a(q_t, k_s) v_s$$

3.3 Category


Attention种类
权值计算方式
点乘算法
矩阵相乘
Cos相似度
串联方式
多层感知
模型结构
单层Attention
多层Attention
多头Attention
计算区域
Soft Attention | Global Attention
Local Attention
Hard Attention
所用信息
General Attention
Self Attention
使用模型
CNN+Attention
RNN+Attention
LSTM+Attention
纯Attention
知乎 @赵强

3.4 Bahdanau Attention



1) Context vector

$$c_t = \sum_{i=1}^{T} \alpha_{ti} h_i$$

$$\alpha_{ti} = \frac{exp(e_{ti})}{\sum_{k=1}^{T} exp(e_{tk})}$$

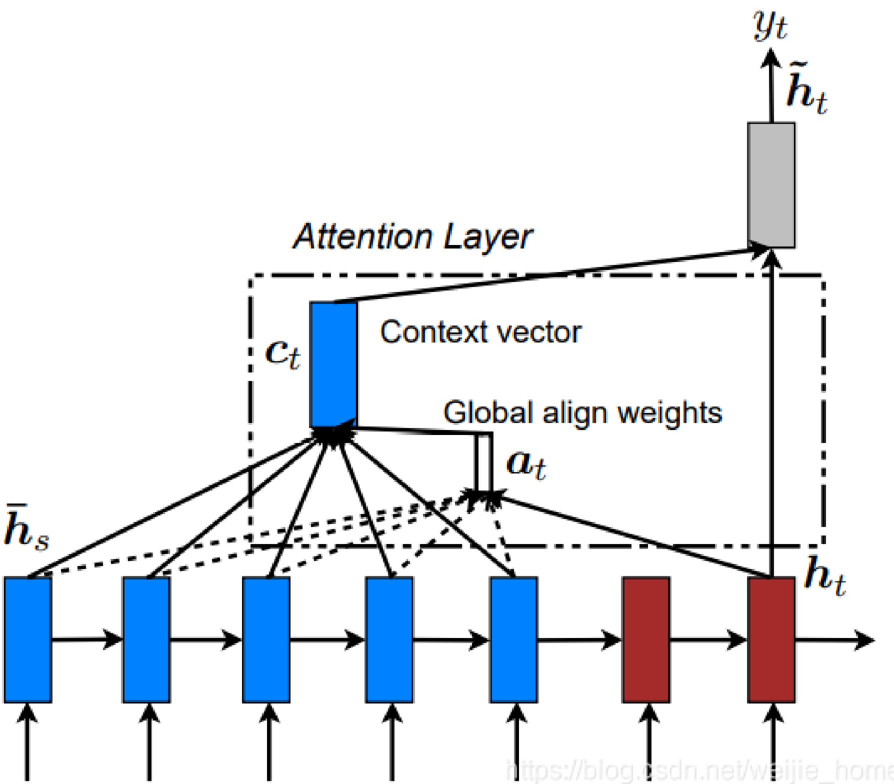
$$e_{ti} = v_a^{\top} tanh(W_a[s_{i-1}, h_i])$$

2) Hidden layer parameters

$$s_t = tanh(W[s_{t-1}, y_{t-1}, c_t])$$

$$o_t = softmax(V s_t)$$

3.5 Luong Attention



1) Hidden layer parameters

$$s_t = tanh(W[s_{t-1}, y_{t-1}])$$

2) Context vector

$$c_t = \sum_{i=1}^{T} \alpha_{ti} h_i$$

$$\alpha_{ti} = \frac{exp(e_{ti})}{\sum_{k=1}^{T} exp(e_{tk})}$$

$$e_{ti} = s_t^{\top} W_a h_i$$

3) Hidden layer parameters

$$\tilde{s}_t = tanh(W_c[s_t, c_t])$$

$$o_t = softmax(V\tilde{s}_t)$$

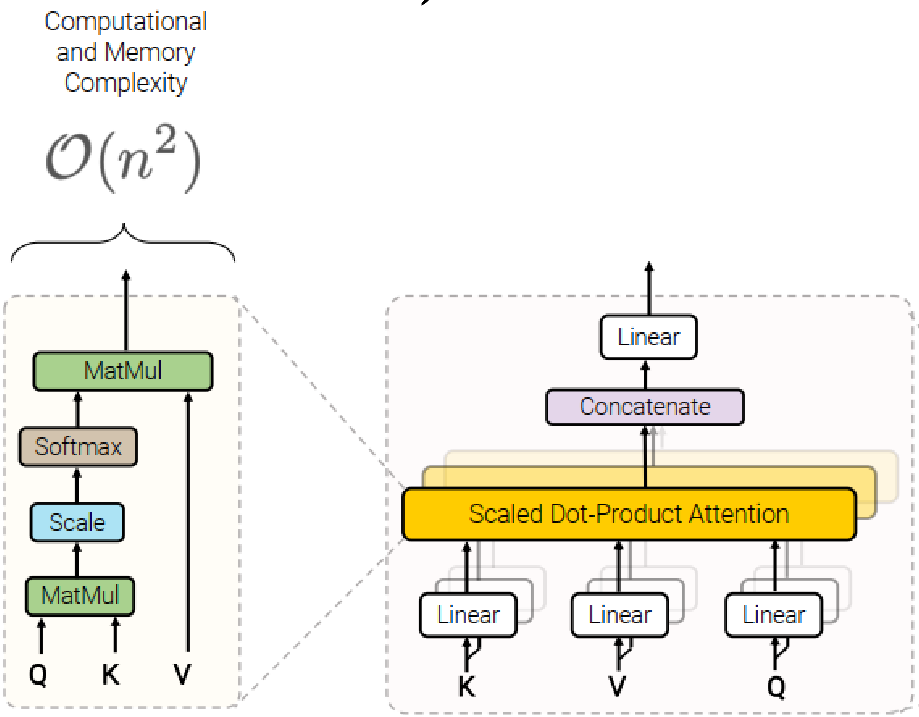
西安交通大學
XI'AN JIAOTONG UNIVERSITY

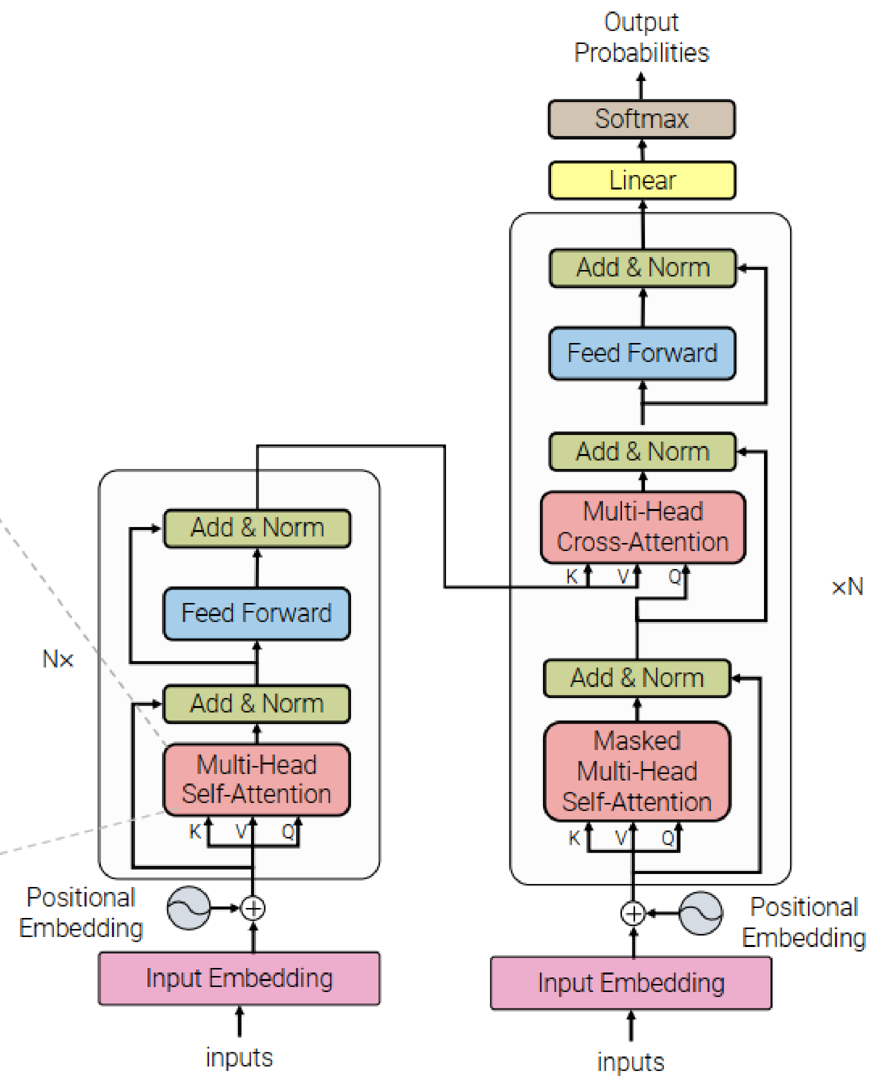
04

Transformer

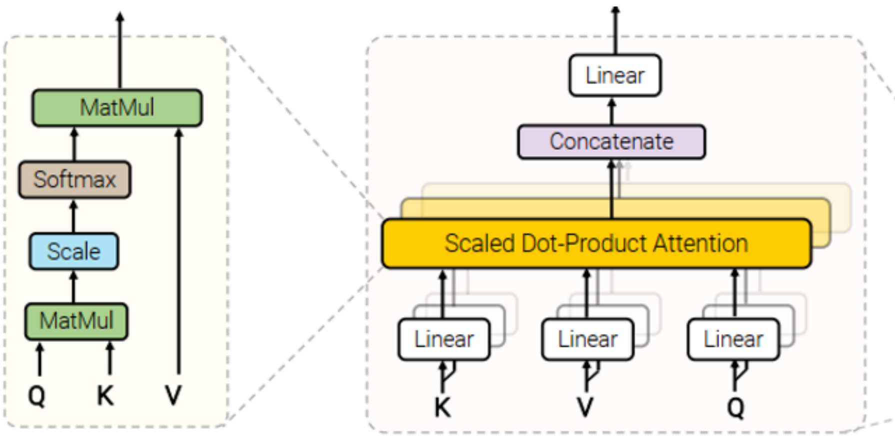
4.1 Structure

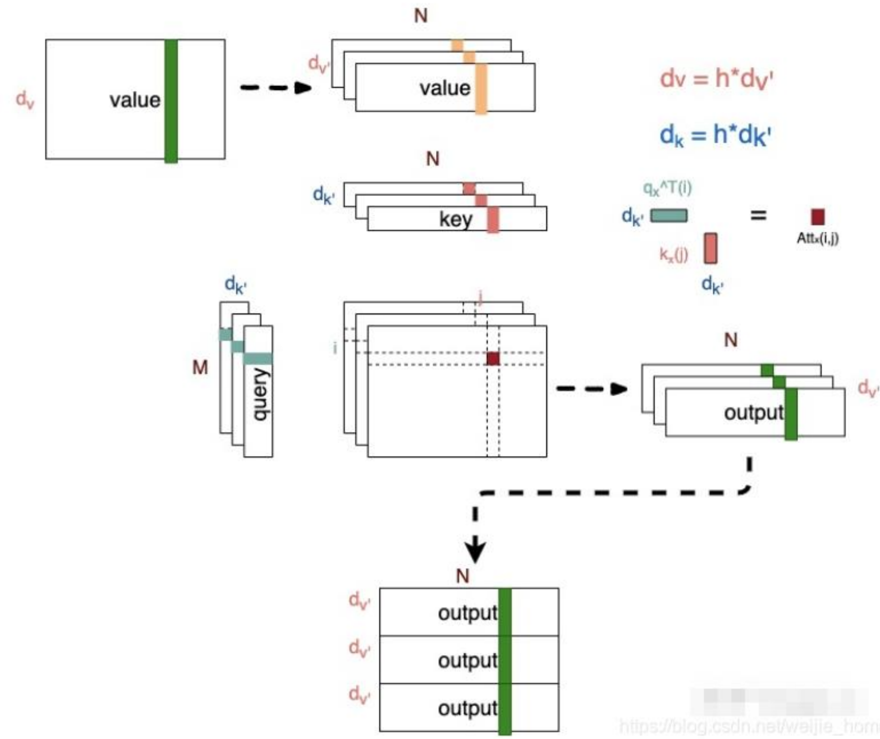
1) **Self-Attention**
2) **Cross-Attention**





4.2 Multi-Head Attention





$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions.

西安交通大學
XI'AN JIAOTONG UNIVERSITY

05 References

5 References

Paper:

[1] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.

[2] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems. 2014: 3104-3112.

[3] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.

[4] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv:1508.04025, 2015.

[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.

Blog:

https://zhuanlan.zhihu.com/p/30844905

https://easyai.tech/ai-definition/rnn/

https://blog.csdn.net/weijie_home/article/details/116407137

Github:

https://github.com/pprp/awesome-attention-mechanism-in-cv

Thank you for watching!

weijie_xjtu@stu.xjtu.edu.cn