

# 注意力机制及相关说明

DemonPan

2019-12-19

当用神经网络来处理大量的输入信息时，也可以借鉴人脑的注意力机制，只选择一些关键的信息输入进行处理，来提高神经网络的效率。按照认知神经学中的注意力，可以总体上分为两类：[1]

聚焦式 (focus) 注意力：自上而下的有意识的注意力，主动注意——是指有预定目的、依赖任务的、主动有意识地聚焦于某一对象的注意力；

显著性 (saliency-based) 注意力：自下而上的有意识的注意力，被动注意——基于显著性的注意力是由外界刺激驱动的注意，不需要主动干预，也和任务无关；可以将 max-pooling 和门控 (gating) 机制来近似地看作是自下而上的基于显著性的注意力机制。

在人工神经网络中，注意力机制一般就特指聚焦式注意力。

注意力机制的理解。

在给定输入  $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  的情况下，用  $\mathbf{q}$  表示当前任务，用  $Attention(\mathbf{q}, \mathbf{X})$  表示  $\mathbf{q}$  对  $\mathbf{X}$  各元素的注意力，则

$$Attention(\mathbf{q}, \mathbf{X}) = softmax(score(\mathbf{q}, \mathbf{X})) = [\alpha_1, \alpha_2, \dots, \alpha_N] \quad (1)$$

$score$  的计算方式有多种，如：

$$score(\mathbf{q}, \mathbf{x}_i) = \mathbf{q} \cdot \mathbf{x}_i \quad (2)$$

1

$$score(\mathbf{q}, \mathbf{x}_i) = \frac{\mathbf{q} \cdot \mathbf{x}_i}{\sqrt{d_x}} \quad (3)$$

2

$$score(\mathbf{q}, \mathbf{x}_i) = \mathbf{W}\mathbf{q} + \mathbf{U}\mathbf{x}_i \quad (4)$$

---

<sup>1</sup>点积模型，点积模型和加法模型效果差不多，但点积模型可以更好的利用矩阵的乘积，从而是计算效率更高

<sup>2</sup>缩放点积模型：维度较大的点积会使得方差较大，从而导致 softmax 函数梯度比较小不利于优化

<sup>3</sup> 用 softmax 求 *score* 的值，是为了让 score 值都为正值，并且使其分数之和为 1 的概率分布。

想象一种注意力场景。比如雪地上有一颗树，眼睛看到了雪地也看到了树，但我们更多的注意到了树（当然雪地也看到了），可是眼睛只能在同一时间将一种信号送给大脑，那么这种信号该如何表示呢？

有了任务  $\mathbf{q}$  对  $\mathbf{X}$  不同的注意力，现在要做的是把对不同注意力的影响，融合成一种最终注意到的结果。用  $r(\mathbf{X}, \mathbf{q})$  表示最终结果，则

$$r(\mathbf{X}, \mathbf{q}) = \sum_{i=1}^N \alpha_i \mathbf{x}_i \quad (5)$$

## 参考文献

- [1] 邱锡鹏. 神经网络与深度学习. <https://nndl.github.io/>, 2019.

---

<sup>3</sup>加法模型：与点击模型效果区别不大