

离散卷积和卷积神经网络

Liam Huang*

2017 年 8 月 6 日

*liamhuang0205@gmail.com

初识

一问卷积

卷积可以吃吗？

卷积为什么要叫这个名字？

回答：

- 卷的意思是「翻转」；
- 积在这里对应「乘积」；
- 此外，它真的可以卷.

形式定义

- 离散形式： $(x * y)[n] = \sum_{m=-\infty}^{+\infty} x[m] \cdot y[n - m].$
- 连续形式： $(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau) \cdot g(t - \tau) d\tau.$

一维离散卷积

线性时不变系统



- 线性

输入	输出
$x_1(\tau)$	$y_1(\tau)$
$x_2(\tau)$	$y_2(\tau)$
$c_1x_1(\tau) + c_2x_2(\tau)$	$c_1y_1(\tau) + c_2y_2(\tau)$

- 时不变

输入	输出
$x(\tau)$	$y(\tau)$
$x(t_0 + \tau)$	$y(t_0 + \tau)$

冲激和响应

- 脉冲激励序列：输入的脉冲激励信号序列.

$$y[n] = \{\dots, y_{-1} = 0, y_0 = i, y_1 = j, y_2 = k, y_3 = 0, \dots\}.$$

- 冲激响应序列：接收单位脉冲激励的输出序列.

$$x[n] = \{\dots, x_{-1} = 0, x_0 = a, x_1 = b, x_2 = c, x_3 = 0, \dots\}.$$

输出序列

- 输入冲激 \hat{y} 连续给出 $x[n]$ 的响应;
- 连续收到若干个输入.

第 i 时刻的输出

$$y[0] \cdot x[i] + y[1] \cdot x[i-1] + \dots$$

time	0	1	2	3	4
$y_0 = i$	ai	bi	ci	0	0
$y_1 = j$	0	aj	bj	cj	0
$y_2 = k$	0	0	ak	bk	ck

输出序列:

$$\begin{aligned} z[n] = \{ \dots, z_{-1} = 0, z_0 &= ai, z_1 = aj + bi, z_2 = ak + bj + ci, \\ &z_3 = bk + cj, z_4 = ck, z_5 = 0, \dots \}. \end{aligned}$$

离散卷积

对任意时刻 n , 系统的输出是

$$\begin{aligned} z[n] \stackrel{\text{def}}{=} (x * y)[n] &= \sum_{m=-\infty}^{+\infty} x[n-m] \cdot y[m] \\ &= \sum_{m=-\infty}^{+\infty} x[m] \cdot y[n-m]. \end{aligned}$$

以 $\sum_{m=-\infty}^{+\infty} x[m] \cdot y[n-m]$ 讨论, 不难发现, 卷积其实是一种推广的加权平均:

- 以 x 为权, 以 n 为中心;
- 把 y 距离中心 $-m$ 位置的值与 x 距离中心 m 位置的值相乘;
- 最后相加.

定投的例子

对利率不变的定期存款的定投，可视作线性时不变系统.

- 最终收益对投入的资金是线性累加的；
- 利率不变，意味着任何时刻存钱的收益是一致的.

定义响应序列和冲激序列.

- 脉冲激励序列：定期存入的存款.

$$y[n] = \{\dots, y_{-1} = 0, y_0 = 100, y_1 = 100, \dots, y_i = 100, \dots\}.$$

- 冲激响应序列：到期后，本息合计的比例序列.

$$x[n] = \{\dots, x_{-1} = 0, x_0 = 1.05^0, x_1 = 1.05^1, \dots, x_i = 1.05^i, \dots\}.$$

任意时刻的账户余额 $z[n] \stackrel{\text{def}}{=} (x * y)[n]$ 是卷积.

time	0	1	2	3	4
$y_0 = 100$	100	100×1.05^1	100×1.05^2	100×1.05^3	100×1.05^4
$y_1 = 100$	0	100	100×1.05^1	100×1.05^2	100×1.05^3
$y_2 = 100$	0	0	100	100×1.05^1	100×1.05^2
$y_3 = 100$	0	0	0	100	100×1.05^1
$y_4 = 100$	0	0	0	0	100

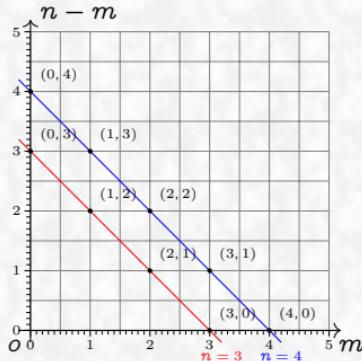
怎样翻转?

观察 $\sum_{m=-\infty}^{+\infty} x[m] \cdot y[n-m]$.

- 以 m 为自变量;
- $y[n-m]$ 相当于:
 - 翻转 $y[m]$ 的图像;
 - 右移 n 个单位;
- 增加 n 可视作 y 沿轴线向右滑动.



怎样「卷」



$$\begin{aligned}z[3] &= \cdots + x[3]y[0] + x[2]y[1] \\&\quad + x[1]y[2] + x[0]y[3] + \cdots \\z[4] &= \cdots + x[4]y[0] + x[3]y[1] \\&\quad + x[2]y[2] + x[1]y[3] + x[0]y[4] + \cdots\end{aligned}$$

二向卷积

卷积有哪些特点？

- 一个冲激影响多个输出；
- 一个输出受多个冲激的影响；
- 冲激和响应共同决定输出；
- 两个序列地位等同可以互换
→ 滑动卷积核.

二维离散卷积

形式定义

一维离散卷积的形式为

$$(x * y)[n] = \sum_{m=-\infty}^{+\infty} x[m] \cdot y[n - m].$$

二维离散卷积的形式是类似的

$$(x * y)[\vec{n}] \stackrel{\text{def}}{=} \sum_{\vec{m}=(-\infty, -\infty)}^{+\infty, +\infty} x[\vec{m}] \cdot y[\vec{n} - \vec{m}]$$
$$(x * y)[n_1, n_2] \stackrel{\text{def}}{=} \sum_{m_1, m_2=-\infty, -\infty}^{+\infty, +\infty} x[m_1, m_2] \cdot y[n_1 - m_1, n_2 - m_2].$$

卷积与图像滤镜



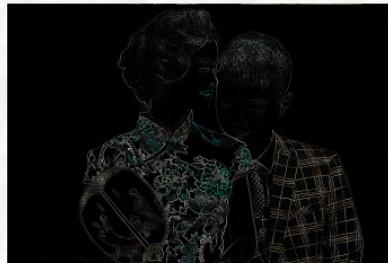
模糊化
 $\begin{bmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}$



锐化
 $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$



边缘检测
 $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$



三问卷积

卷积核有什么作用？

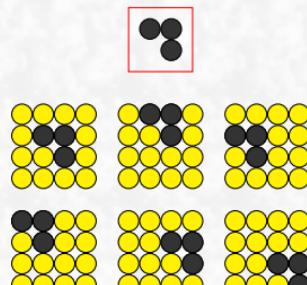
- 先验知识「滤镜」 → 卷积核处理图形 → 生物的视觉处理；
- 适者生存 → 生物进化 → 不同侧重点；
 - 蛇类：不能视物、红外观察；
 - 青蛙：动态视力 max.

卷积	生物视觉	机器学习领域的意义
滑动卷积核	视觉系统处理外界信号	读入并处理结构化特征
卷积核处理结果	脑海中的视觉影像	卷积处理的结果
不同的卷积核	观察世界的不同侧重点	不同的高维特征信号

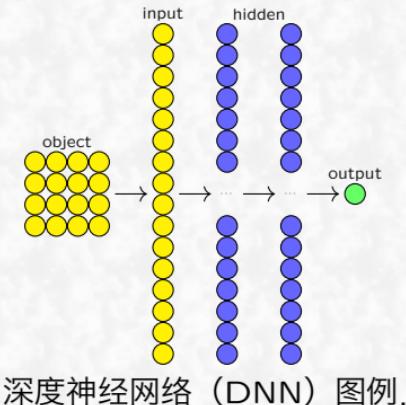
特定的卷积核，能够从**若干相关信号中以特定的方式抽取高维特征**.

卷积神经网络

图片识别任务

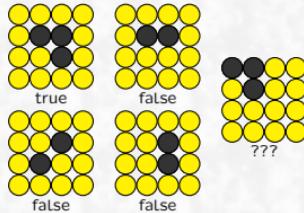


识别「横折」.



深度神经网络 (DNN) 图例.

无法胜任的情况



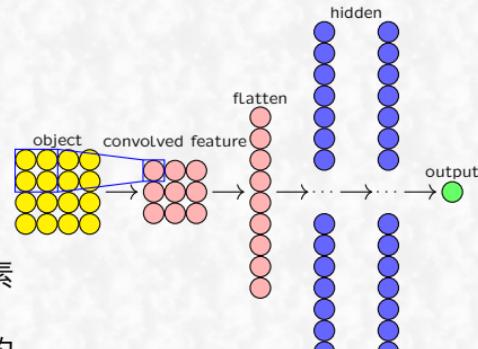
- 左侧为训练数据；
- 右侧为无法识别的验证数据。

解决办法：增加训练数据，覆盖边界情况。

表意的平移不变性

- 平移不变性：有效像素组合，任意平移，表意不变。
- 问题产生的原因
 - 每个像素点是独立的，单独赋予参数；
 - 各个区域各自为政，没有关联。
- 解决办法：卷积。

引入卷积



- 卷积加强关联了相邻的像素点；
- 同一个卷积核应用到图像的不同区域. (引入**先验知识**)

卷积神经网络 (CNN) 图例.

不变性的讨论

平移不变性：平移不影响表意 \leftarrow 卷积核共享.

缩放不变性：缩放大小不影响表意 \leftarrow 同时应用不同大小的卷积核/多次卷积.

旋转不变性：旋转方向不影响表意 \leftarrow 目前只能通过大量数据.

总结

公式

- 一维:

$$(x * y)[n] = \sum_{m=-\infty}^{+\infty} x[m] \cdot y[n - m].$$

- 二维:

$$(x * y)[n_1, n_2] \stackrel{\text{def}}{=} \sum_{m_1, m_2 = -\infty, -\infty}^{+\infty, +\infty} x[m_1, m_2] \cdot y[n_1 - m_1, n_2 - m_2].$$

特点

- 形式上看，「真的可以卷」.
- 关联多个输入， 对应同一输出.
- 输出综合输入和卷积核的信息.
- 卷积核多处复用、共享.

CNN 直觉上的优势

- 适用于相关元素有结构特征的情况；
- 适用于上述结构可能出现在不同位置的情况.

特别地，对于分类问题

- 分割超平面附近，往往存在结构特征；
- 输入的样本，可能位于超平面的不同位置.

因此，人们偏向于认为：「CNN 适合应用在分类问题上」.

UGA