



翻译小组

手机版

云栖社区 (/) > 翻译小组 (teams/132) > 博客 (teams/132/type\_blog) > 正文

://service.weibo.com/share/share.php?

达人介绍

%E3%80%90%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0%E4%B

<https://www.aliyun.com/section/product/20190309>



【方向】  
阿里云 / 5m4kk67c  
文章 769篇 | 关注

+ 关注

# 【深度学习之美】神经网络不胜数， M-P模型似可寻（入门系列之三）

【方向】 (users/5m4kk67c65aaqq) 2017-05-27 00:38:11 浏览24832 评论3

[云栖社区 \(/tags/type\\_blog-tagid\\_1/\)](#) [深度学习 \(/tags/type\\_blog-tagid\\_16/\)](#)

[神经网络 \(/tags/type\\_blog-tagid\\_13435/\)](#) [激活函数 \(/tags/type\\_blog-tagid\\_17958/\)](#)

[品味大数据 \(/tags/type\\_blog-tagid\\_17970/\)](#) [张玉宏 \(/tags/type\\_blog-tagid\\_18831/\)](#)

[MP模型 \(/tags/type\\_blog-tagid\\_18836/\)](#) [深度学习之美 \(/tags/type\\_blog-tagid\\_28501/\)](#)

摘要：“那些在个人设备里，谦谦卑卑地为我们哼着歌曲的数字仆人，总有一天会成为我们的霸主！”在“忍无可忍，无需再忍”这句俗语背后，也隐藏着神经网络常用的“激活函数”和“卷积”的概念。知其道,用其妙,THIS IS HOW!

## 系列文章：

【深度学习 (https://yq.aliyun.com/articles/86580)之美】一入侯门“深”似海，深度学习深几许（入门系列之一） (https://yq.aliyun.com/articles/86580)

【深度学习 (https://yq.aliyun.com/articles/86580)之美】人工“碳”索意犹未尽，智能“硅”来未可知（入门系列之二） (https://yq.aliyun.com/articles/88300)

“那些在个人设备里，谦谦卑卑地为我们哼着歌曲的数字仆人，总有一天会成为我们的霸主！”

——A.I. winter

在前面的小节中，我们大致了解了机器学习的形式化定义和神经网络的概念，在本小节中，我们将相对深入地探讨一下神经网络中的神经元模型以及深度学习常常用到的激活函数及卷积函数。

### 3.1 M-P神经元模型是什么？

在前一小节中，我们已介绍了人工神经网络（Artificial Neural Networks，ANNs）的定义。简单来说，它是一种模仿动物神经网络行为特征 (http://baike.so.com/doc/152515.html)，进行分布式并行信息处理的算法数学模型 (http://baike.so.com/doc/5513175.html)。

神经网络依靠系统的复杂程度，通过调整内部大量“简单单元”之间相互连接的关系，从而达到处理信息的目的，并具有自学习和自适应的能力。

## 目录

- 3.1 M-P神经元模型是什么？
- 3.2 “M-P神经元模型”背后的那些人和事情
- 3.4 卷积函数又是什么？
- 3.5 小节
- 3.6 请你思考
- 【参考文献】

## 博主其他文章 更多> (/users/5m4kk67)

- 图神经网络(GNN)的简介 (/articles/694432)
- 天桥调参师秘籍：一份深度学习超参微调技 (/articles/694374)
- 数据清理的终极指南 (/articles/694209)
- 100行Python代码理解深度学习关键概念：从肿瘤检测网络 (/articles/694113)
- 活体检测很复杂？仅使用opencv就能实现！ (/articles/694045)
- 2019五个最棒的机器学习课程 (/articles/693932)
- 可应用于实际的14个NLP突破性研究成果 ( /articles/689392)
- 将视觉深度学习模型应用于非视觉领域 (/articles/689232)
- 可应用于实际的14个NLP突破性研究成果 ( /articles/689196)
- 可应用于实际的14个NLP突破性研究成果 ( /articles/689082)

## 相关话题

- 云栖社区微信小程序“云栖365”上线了，帮助碎片时间学习 (/roundtable/495548)
- 三七女生节，邀你一起解密阿里女程序员们 (/roundtable/495243)
- 又双叒删库了，这次是 AI 动的手 (/roundt

在上述定义中提及到的“简单单元”，其实就是神经网络中的最基本元素——神经元（neuron）模型。在生物神经网络中，每个神经元与其它神经元，通过突触联接。神经元之间的“信息”传递，属于化学物质传递的。当它“兴奋（fire）”时，就会向与它相连的神经元发送化学物质（神经递质，neurotransmitter），从而改变这些神经元的电位；如果某些神经元的电位超过了一个“阈值（threshold）”，那么，它就会被“激活（activation）”，也就是“兴奋”起来，接着向其它神经元发送化学物质，犹如涟漪，就这样一层接着一层传播，如图3-1所示。

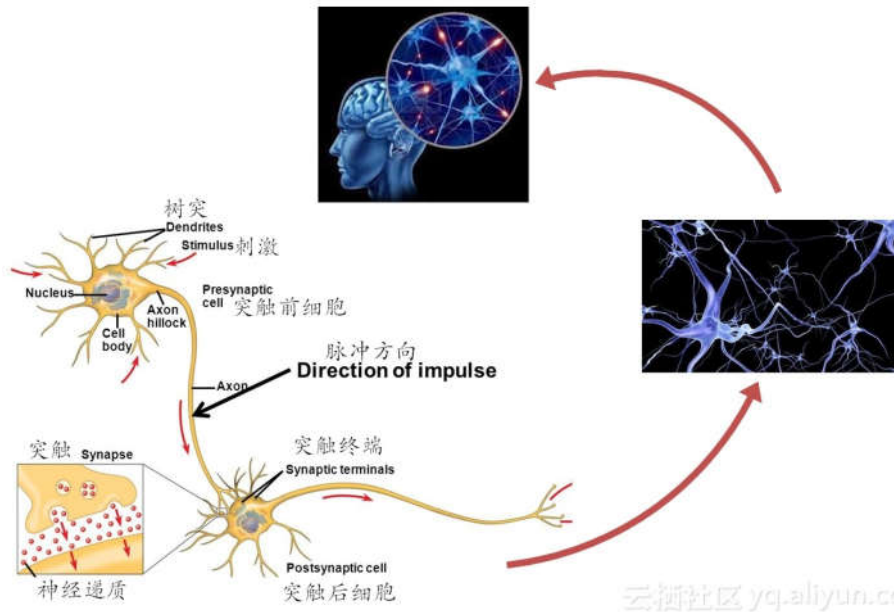


图3-1 大脑神经细胞的工作流程

在人工智能领域，有个好玩的派别叫“飞鸟派”。说的就是，如果我们想要学飞翔，就得向“飞鸟”来学习。简单来说，“飞鸟派”就是“仿生派”，即把进化了几百万年的生物，作为“模仿”对象，搞清楚原理后，再复现这些对象的特征。

其现在所讲的神经网络包括深度学习，都在某种程度上，属于“飞鸟派”——它们在模拟大脑神经元的工作机理。追根溯源，模仿神经元的“飞鸟”实例，就是上世纪40年代提出但一直沿用至今的“M-P神经元模型”。

在这个模型中，神经元接收来自n个其它神经元传递过来的输入信号，这些信号的表达，通常通过神经元之间连接的权重（weight）大小来表示，神经元将接收到的输入值按照某种权重叠加起来，并将当前神经元的阈值进行比较，然后通过“激活函数（activation function）”向外表达输出（这在概念上就叫感知机），如图3-2所示。

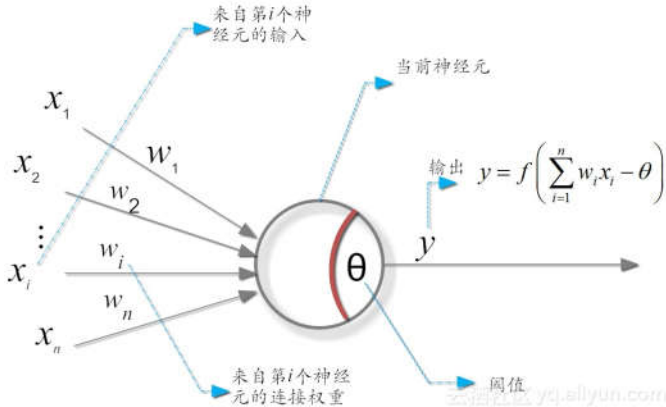


图3-2 M-P神经元模型

简单吧？简单！

可这么简单的道理，你咋能就这么轻易知道咧？

性能测试 PTS 有奖分享，新年大礼包等你拿  
(/roundtable/493729)

【赠送参会码】PostgreSQL 2018 中国技术  
告 (/roundtable/473635)

开年Hi购季

开发者分会

海量学习资源，一站式开发者福利

{/}

(<https://www.aliyun.com/acts/product-section-2019/developer>)

事实上，我们知道，对未知世界（比如说人类大脑）的每一点新认识，其实都是因为有大牛曾经艰辛地为我们站过台。前面提到的“M-P神经元模型”，亦是如此。

### 3.2 “M-P神经元模型”背后的那些人和事情

“M-P神经元模型”，最早源于发表于1943年的一篇开创性论文。论文的两位作者分别是神经生理学家沃伦·麦克洛克（McCulloch）和数学家沃尔特·皮茨（Pitts），论文描述了神经元的时间总和、域值等特征。首次实现用一个简单电路（即感知机），来模拟大脑神经元的行为。

“M-P神经元模型”提出者虽然有两人，但后者皮茨更有声望和传奇色彩。皮茨等人的研究，甚至影响了控制论的诞生和冯·诺伊曼（von Neumann）计算机的设计。

一般来说，每个“牛逼”人的背后，都隐藏着一段“苦逼”的岁月。皮茨同样如此，不过更显“苦涩”而已。1923年3月，沃尔特·皮茨出生于美国密歇根州底特律（Detroit）。小时候，皮茨家境贫寒。但贫穷并没有冷却皮茨那颗爱读书的心。皮茨很早就对数学和哲学表现出浓厚的兴趣，12岁时就开始阅读伯特兰·罗素（Russell）和怀特海（Whitehead）合著的《数学原理》。

年少的皮茨，视罗素为人生偶像，经常就不明觉厉的问题和罗素通信。在书信往来中，罗素发现这个皮茨，人小才不小，于是便邀请他到英国，跟随自己学习逻辑学。

但皮茨家境贫困，连他读高中都供不起，哪还有钱资助他远赴英伦求学。就在皮茨15岁时，皮茨的老爹还强迫他退学，以便在底特律找份工作谋生，补贴家用。

但酷爱学习的皮茨，自然不愿答应，与老爹一言不合，就离家出走，浪荡江湖。从此，他再也没有见过他的家人。

但正可谓“山不转水转，水不转人转”。虽然皮茨无钱从美抵英国求学，但罗素这位学术大家，却有条件从英国抵达美国任教。

在打听到自己偶像罗素要到芝加哥大学任教（访问教授），皮茨兴奋异常，便只身来到芝加哥拜见。精诚所至，金石为开，终于皮茨见到了罗素。罗素一见皮茨，发现他果然骨骼清奇，天赋异禀。于是，罗素就将皮茨留下来，旁听自己的课程。

话说当时在芝加哥大学任教的，还有一个知名的哲学教授，叫鲁道夫·卡尔纳普（Rudolf Carnap）。他是逻辑实证主义的代表性人物，其哲学思想深受罗素影响，和罗素是好基友。

罗素的课，皮茨自然可以去随便蹭，但活下去，却是皮茨自己的事。在机缘巧合下，卡尔纳普认识了皮茨，并发现皮茨有惊为天人的才华。

卡尔纳普惜才如命，便心生怜悯，利用自己的人际关系，帮皮茨在芝加哥大学谋了打扫卫生的差事。就这样，皮茨这位当代的“扫地僧”，总算是在这所著名大学安顿下来。

再后来，18岁的皮茨在芝加哥的结识了神经学家、哲学家沃伦·麦克洛克（Warren McCulloch）。正是这位麦克洛克，彻底改变了皮茨的人生轨迹。

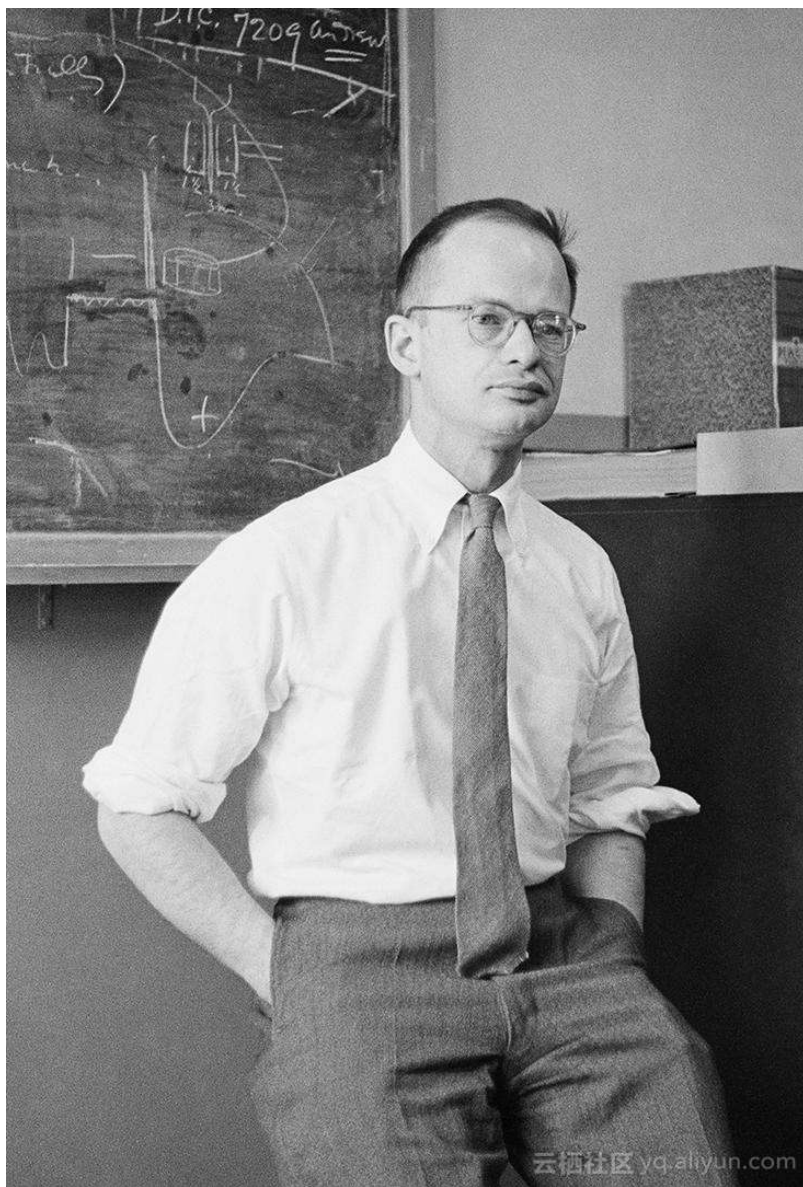


图3-3 “M-P模型”的创立者——沃尔特·皮茨（Walter Pitts）

他们合作的最重要的成就之一，就是那篇“神作”——神经网络的天下第一文：《神经活动中思想内在性的逻辑演算》（A Logical Calculus of Ideas Immanent in Nervous Activity）[1]，这篇文章发表在《数学生物物理期刊》上，它也成为控制论的思想源泉之一。

信号在大脑中到底是怎样的一种传输，确切来说，依然一个谜。麦克洛克和皮茨提出的“M-P神经元模型”，是对生物大脑的过度简化，但却成功地给我们提供了基本原理的证明。但对我们而言，重要的是可以把它视为与计算机一样的存在，利用一系列的0和1来进行操作。也就是说，大脑的神经细胞也只有两种状态：兴奋（fire）和不兴奋（即抑制）。

这样一来，神经元的工作形式，类似于数字电路中的逻辑门，它接受多个输入，然后产生单一的输出。通过改变神经元的激发阈值，就可完成“与（AND）”、“或（OR）”及“非（NOT）”等三个状态转换功能，如图3-4所示。

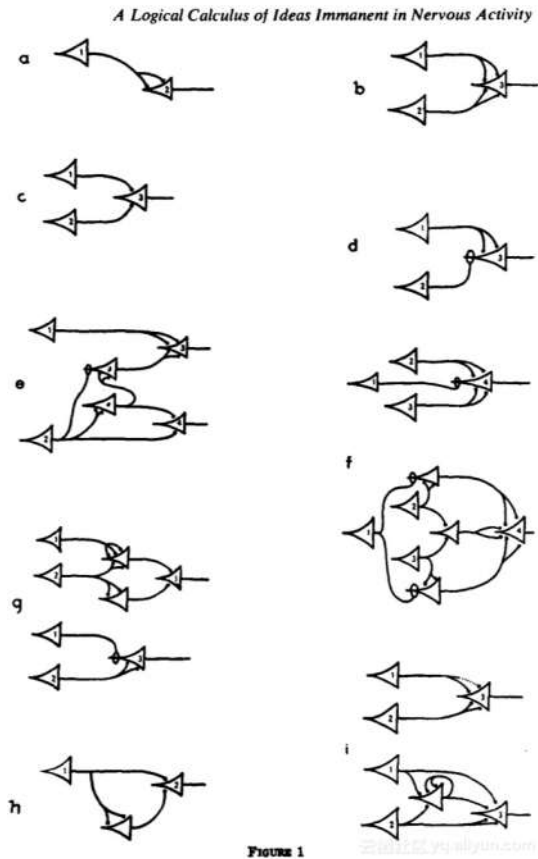


图3-4 神经网络天下第一文《神经活动中思想内在的逻辑演算》中的M-P神经元（发表于1943年）

这里需要说明的是，“感知机”作为一个专业术语，是皮茨等人发表论文15年之后，在1958年，由康内尔大学心理学教授弗兰克·罗森布拉特(Frank Rosenblatt)提出来的，这是一个两层的人工神经网络，成为后来许多神经网络的基础，但它的理论基础依然还是皮茨等人提出来的“M-P”神经元模型。

不知道读者朋友发现了没有？皮茨(Pitts)等人提出的M-P模型，居然不能实现常用的“异或(XOR)”功能，这也成为人工智能泰斗马文·明斯基(Marvin Lee Minsky)评判“神经网络”的重要“罪证”之一。

1969年，马文·明斯基和西摩尔·派普特(Seymour Papert)在出版了《感知机：计算几何简介》一书[2]，书中论述了感知机模型存在的两个关键问题：

- (1) 单层的神经网络无法解决不可线性分割的问题，典型例子如异或门电路(XOR Circuit)；
- (2) 更为严重的问题是，即使使用当时最先进的计算机，也没有足够计算能力，完成神经网络模型所需要的超大的计算量(比如调整网络中的权重参数)。

鉴于明斯基的江湖地位(1969年刚刚获得大名鼎鼎的图灵奖)，他老人家一发话不要紧，直接把人工智能的研究，送进一个长达近二十年的低潮，史称“人工智能冬天(AI Winter)”。

现在，距离明斯基出版《感知机》，已经近50年过去了。我们看到了，以深度学习为代表的神经网络学习，又一番风生水起，好一个热闹了得！

这让我想起了一本心理学名著《改变：问题形成和解决的原则》[3]。在这本书中，三位斯坦福大学医学院精神病与行为科学系临床教授提出一个有趣的观点，“一个问题尚未解决，虽然令人生厌，但其本身就常是一种解决方式”。

我们通常想一次性地把所有问题都解决掉，即存在“问题洁癖”状态。但这种“洁癖”状态，通常并不能解决问题，反而很可能会带来更多、更麻烦的问题。因此，带着问题生活(或研究)，是可以接受的，并且很可能是最好的结果。

从最初级的“M-P”神经元模型，发展到今天的“深度学习”，还有没有问题呢？当然有！比如说深度学习的解释性低，对数据依赖度高等(想一想，现在的小娃娃识别一个猫狗的图片，仅仅需要几张图片训练就够了，可深度学习算法呢？可能需要数以百万张)，诸如此类问题，这并没有妨碍深度学习在各个领域的大放异彩。

著名物理学家理查德·费曼（Richard Feynman，1965年诺贝尔物理学奖得主）也曾说过的，“物理的进步，其实源于它对寻求根源的让步。”举例说来，牛顿描绘出了如何用万有引力（how），但到死他也没有弄明白为什么会存在引力（why），但这并没有妨碍我们应用它们，解决我们宏观世界中的各类经典物理问题（一个题外话，很苦闷的牛顿，最后把why问题归结为上帝所为，所以他成为一个极度虔诚的基督教徒）。

针对目前深度学习存在的问题，有解决之道吗？虽然方向并不十分明朗，但未来一定会有，比如说在未来，我们可以把深度学习、强化学习和迁移学习相结合，可以实现几个突破——反馈可以延迟，通用的模型可以个性化，可以解决冷启动的问题等[4]。

### 3.3 激活函数是怎样的一种存在？

言归正传。前面我们提到了神经元的工作模型存在“激活（1）”和“抑制（0）”等两种状态的跳变，那么理想型的激活函数（activation functions）就应该是如图3-5-a所示的阶跃函数，但事实上，在实际使用中，这种函数具有不光滑、不连续等众多不“友好”的特性。为什么说它“不友好”呢，这是因为在训练网络权重时，通常依赖对某个权重求偏导、寻极值，而不光滑、不连续等通常意味着该函数无法“连续可导”。

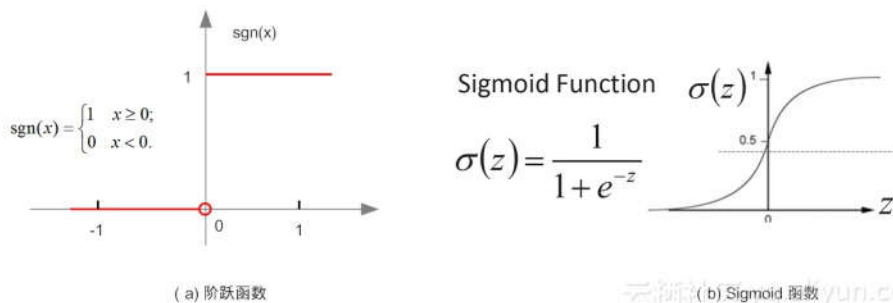


图3-5 典型的神经元激活函数

因此，我们通常用Sigmoid函数（如图3-5-b所示）来代替阶跃函数。这个函数可以把较大变化范围内输入值（x）挤压输出在（0,1）范围之内，故此这个函数又称为“挤压函数（Squashing function）”。

那么，我们应该怎样理解激活函数呢？实际上，我们还是能从生活中找到相似的影子（理论来源人们对生活的抽象嘛）。比如说，如果你的“野蛮女友”打你耳光，当她打你第一个耳光时，你想了很多，考虑她的脸蛋（权重w1）、身材（权重w2）、学历（权重w3）、性格（权重w4）、你爱她的程度（权重w5）以及娶媳妇的难度（权重w6）等等因素（这些因素，在机器学习领域，就是研究对象的特征，它们组合在一起，就构成了对象的特征空间），综合衡量后，你决定忍了，或者说这一切并没有超出你忍耐的阈值（这期间，你也给自己找了一个优雅的台阶：打是亲、骂是爱啊），如果把你的忍耐“功能”看作一个函数（回顾一下我们前面的文章可以了解，“功能”和“函数”，本来就是一个概念——Function），那么在这个场景，相当于你的函数输出为0，目前还处于没有被激活状态。

当她打你第二个耳光时，你又想了很多，依然忍了，但上述因素的权重都开始调整了，比如性格权重下降，爱她的程度权重下降等。

……

当她打你第n个耳光时，你终于忍不了，“忍无可忍，无须再忍”，这时函数输出超出了阈值，你可能“发飙”扬长而去，也可能哭着喊“要打，也别老打脸啊”。

### 3.4 卷积函数又是什么？

说到“忍无可忍，无须再忍”这个场面，其实，我们还可以提前把“卷积（Convolution）”讲讲（在后续卷积神经网络章节中，我们还会给出它形式化的定义，细细讲解）。

假设你的承受能力是一个在时间维度上的函数f，而你“野蛮女友”的打脸操作为函数g，那么卷积的概念，就是重新定义出来一个新的函数h（比如说，h用来刻画你的崩溃指数）： $h = f * g$ 。

这是什么意思呢，通俗来讲，所谓卷积，就是一个功能（如刻画你的承受能力）和另一个功能（比如说描述你女友的打脸）在时间的维度上的“叠加”作用。就是这么简单！



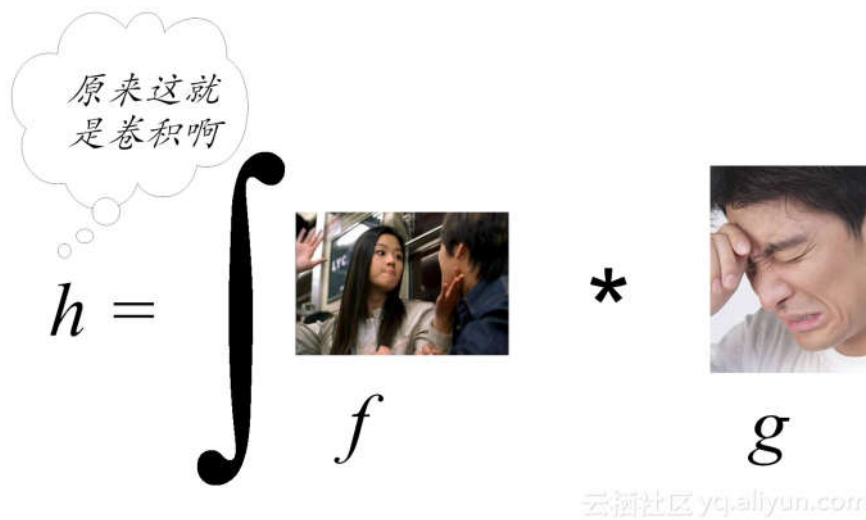


图3-6 恋爱生活中的“卷积”

你可别忘了，在前面我们反复提到，函数（function）就是功能（function），功能就是函数！函数有一定的功能，才有其存在的意义！但孤立的函数并不好玩，叠加才更有意义。说学术点，由卷积得到的函数 $h$ 一般要比 $f$ 和 $g$ 都光滑。利用这一性质，对于任意的可积函数 $f$ ，都可简单地构造出一列逼近于 $f$ 的光滑函数列，这种方法被称之为函数的光滑化或正则化。

在时间的维度上的“叠加作用”，如果函数是离散的，就用求累积和来刻画。如果函数是连续的，就求积分来表达。

虽然第 $n$ 个耳光才让你发飙或跪地求饶，那在时间维度上，第一个耳光、第二个耳光已经过去了……那它们有没有作用呢？当然有！它们的影响早已隐“忍”其内啦！

好了，刚才的例子可能有点得罪女性读者了（我们有那么凶吗？），那我们再举一个稍微温馨的例子好了。当你在向女友求婚时，总是不得其果，然后就一直对女友各种好，直到有一天，你给女朋友洗了一双袜子，哇，不得了，你的女友一下子就感动到崩溃，答应嫁给你了。

这时你要想明白啊，你的女友可没有那么便宜，你洗一双袜子就把人家搞定了？想得美！实际上，正解应该是，你一直对女友的各种好（函数 $f$ ）和你女友的心理期许（函数 $g$ ），一直在时间维度上进行不断地叠加耦合（积分求和），最终超出了女友的阈值，然后她输出了你想要的结果“Yes, I Do”。那么，这个函数 $f$ 和函数 $g$ 一起“卷积”出来的，到底是个什么样的函数呢，我劝你也别猜，“猜来猜去也猜不明白”，反正它很复杂就是了，而你要做的，就是持续不断地对她好就对了。

网络上，有句话说得挺好玩：“爱是积累出来的，不爱也是。曾经无话不说，现在无话，不说。”这一切都是“卷积”卷出来的啊！

### 3.5 小节

以俺读书破万卷，下笔如神经的风格，虽然讲着神经网络的理论问题，却又常常离题万里，似乎非常应景（好神经啊），就此打住吧，下面我们总结一下本节的主要内容。

在本小节，我们主要讲了皮茨等人提出“M-P”感知机模型。简单来说，感知机模型，就是一个由两层神经元构成的网络结构，输入层接收外界的输入，通过激活函数（阈值）变换，把信号传送到输出层，因此它也称之为“阈值逻辑单元（threshold logic unit）”，正是这种简单的逻辑单元，慢慢演进，越来越复杂，就构成了我们目前研究的热点——深度学习网络。

在下小节，我们将聊聊机器学习的三种方式，请你关注。

### 3.6 请你思考

(1) 在生物神经网络中，神经元之间的信息传递，是一种非常局部化的化学物质传递。试想一下，如果每个神经都接收传递物质，那么上亿的神经一起工作，那种能量的消耗是不可想象的。而现在的人工神经网络（深度学习），是依靠大型计算设备（如大规模集群、GPU等）来海量遍历调整网络中的参数，所以不耗能巨大才怪呢，因此你觉得深度学习还能从生物神经网络中学习什么吗？你觉得人工智能领域中的“飞鸟派”还有市场吗？为什么？

(2) 其实，“卷积”在我们日常生活中比比皆是，你还能列举去其它的“卷积”函数吗？

### 【参考文献】

- [1] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.
- [2] Marvin Minsky and Seymour Papert, 1972 (2nd edition with corrections, first edition 1969) Perceptrons: An Introduction to Computational Geometry, The MIT Press, Cambridge MA, ISBN 0-262-63022-2.
- [3] (美) 瓦茨拉维克等, 夏林清等译. 改变：问题形成和解决的原则. 教育科学出版社, 2007-11