

浅谈人工神经网络

——从神经元到 B-P 网络

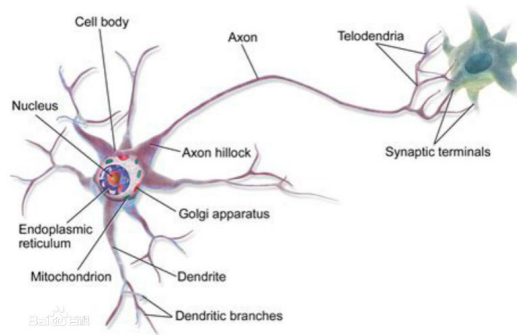
姓名：王之昱 学号：1800011421 指导教师：陈斌

【内容摘要】 人工神经网络 (Artificial Neural Network, ANN)，是通过模拟人脑的结构及其对于外界信息处理的模式和流程，以网络拓扑知识为理论基础，建构起来的一种数学模型。由于其具有强大的并行处理能力、高容错率和自主学习等特点，自诞生起便引起了各领域广泛的关注和研究。本文旨在从此模型的生物界原型——神经元开始，介绍 M-P 模型的产生，从而引出感知器，B-P 网络等技术，在叙述人工神经网络的发展历史的同时，更好地展现其工作原理和性质，并对由此未来进行展望。

【关键词】 人工神经网络 模拟 信息处理 M-P 模型 感知器 B-P 网络

一、M-P 模型的诞生——神经元

20 世纪 40 年代，对神经网络的研究开始逐渐进入人们的视野。生物神经系统所具有的强大的并行计算能力和容错能力，让人们对其结构和工作原理十分感兴趣。而神经系统最基本的砖块——神经元，更是成为了科学家们研究的重点对象。我们的文章，也将从对于神经元的观察与归纳开始。下面是一张神经元的图片¹：



从这里我们可以看出，神经元主要由这几个部分构成：可以接受多个输入的树突，处理信息并转化为输出信号的胞体，对单一目标输出的突触。

我们还知道，神经元具有以下特征：

1. 根据突触性质和强度不同以及递质的种类，会有不同强度（在算法中表现为权重）的兴奋型输入和抑制型输入。
2. 输入需要达到阈值才能产生效果。

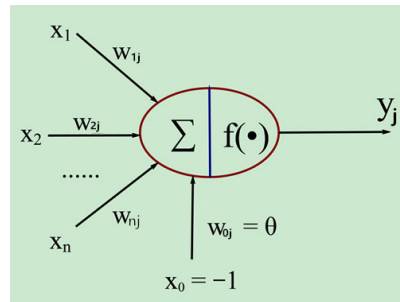
¹ 图片来自于百度百科

3. 输入与输出之间存在时滞。

在上述启发之下，1943 年，美国心理学家麦克洛奇²和数学家皮兹³提出了 M-P 模型。

二、神经网络的砖块——M-P 模型

忽略时间整合作用、不应期等复杂因素，并把神经元的突触时延和强度当成常数之后，我们可以建构如下模型⁴：



用公式表达就是：

$$net'_j(t) = \sum_{i=1}^n \omega_{ij} x_i(t) - \theta_j$$
$$y_j = f(net_j)$$

其中， j 代表神经元； x_i 代表第 i 个输入信号； ω_{ij} 代表接收权重（即对于此信号的灵敏程度），其正负即表示了输出和抑制两种情况； θ_j 表示阈值，只有超过它才能产生反应；这些的和 $net'_j(t)$ 就是净激活，相当于神经元中的膜电位。下面一个公式的 y_j 就代表了神经元 j 的输出； f 称为激活函数，它决定了输出与净激活之间的关系。

这种“阈值加权”的模型就是 M-P 模型，用来作为神经网络的一个处理单元。

三、构建神经网络——单层感知器与多层感知器

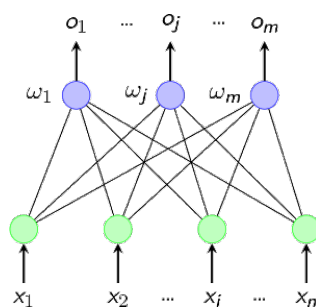
1958 年，美国心理学家 Frank Rosenblatt 提出一种具有单层计算单元的神经网络，称为感知器。它是由数个 M-P 处理单元构成的单层横向结构，如下图⁵：

² Warren Sturgis McCulloch, 1898~1969, 美国神经生理和控制论科学家。哥伦比亚大学博士，先后执教于 MIT、Yale、芝加哥大学。

³ Warren Sturgis McCulloch, 1898~1969, 美国神经生理和控制论科学家。哥伦比亚大学博士，先后执教于 MIT、Yale、芝加哥大学。

⁴ 图片来自 <https://blog.csdn.net/u013007900/article/details/50066315>

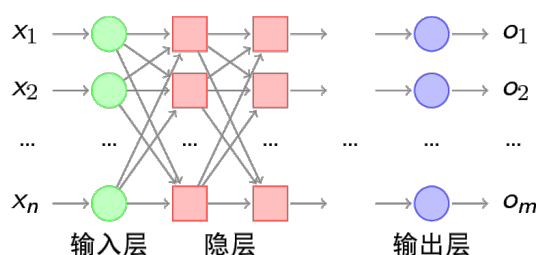
⁵ 同上



其中 o_j 即为节点 j 的输出。我们取 f 为阶梯函数，即

$$o_j = \begin{cases} 1, & \sum_{i=1}^n \omega_{ij}x_i - \theta_j > 0 \\ -1, & \sum_{i=1}^n \omega_{ij}x_i - \theta_j < 0 \end{cases}$$

由于输出函数取值的分界线为一个线性方程，即 $\sum_{i=1}^n \omega_{ij}x_i - \theta_j = 0$ 决定，所以单层感知器只能解决线性问题（如逻辑‘与’、‘或’），所以当需要处理非线性问题时，我们就需要引入多层感知器，如下图⁶：



从图中可以看出，多层感知器比单层感知器多了数个隐层。这些隐层的加入使得其可以解决单层感知器无能为力的非线性问题（如逻辑‘异或’）。根据 Kolmogorov 理论，双隐层感知器就足以解决任何复杂的分类问题。

但是随着层数的增多，问题也随之而来：各个隐层的权值，即接收权重，应当怎样确定或者训练呢？由于隐层位于中间，所以无法计算其输出的期望值，同时也就无法通过一般感知器的学习规则来训练多层感知器。此时的感知器研究无疑已经碰到了瓶颈⁷，直到 1982 年 Hopfield 网络的提出与《并行分布式处理》的发表，才使得神经网络方面的研究重获新生。下面要介绍的 B-P 网络便是一种解决的方案。

四、多层感知器的改进——B-P（Back Propagation）网络

⁶ 图片来源同上

⁷ 1966 年，Minisky 和 Papert 在他们的《感知器》一书中提出了上述的感知器的研究瓶颈

B-P 网络，本质上还是多层感知器网络，只不过在训练过程中使用的是 B-P 学习算法⁸，故简称为 B-P 网络。正如它的名字一样，B-P 算法利用其反向的误差传播来作为修正各个单元权值的依据。其拓扑图⁹和具体计算方法如下（以三层感知器为例）：

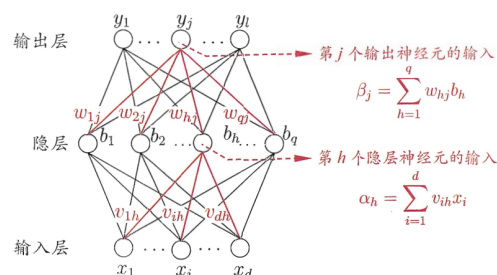


图 5.7 BP 网络及算法中的变量符号^{et/antkillerfarm}

注：由于图源不同，此图及后文中的 y 相当于前文的 o ， b 相当于前文的 y 。

设有 l 个输出端，输出信号为 $y_1, y_2 \dots y_l$ ，各端口输出信号的期望值为 $d_1, d_2 \dots d_l$ ，则由此定义输出误差 $E \equiv \frac{1}{2} \sum_{h=1}^l (d_h - y_h)^2$ 。

将以上误差定义中的输出可以改写成（此式中将阈值以 b_0, x_0 形式表示，并放进求和符号中）：

$$E = \frac{1}{2} \sum_{k=1}^l \left[d_k - f \left(\sum_{h=0}^q \omega_{hj} b_h \right) \right]^2 = \frac{1}{2} \sum_{k=1}^l \left[d_k - f \left(\sum_{h=0}^q \omega_{hj} f \left(\sum_{i=0}^d v_{ih} x_i \right) \right) \right]^2$$

从上式可以得到，误差是 ω_{hj}, v_{ih} 的函数。那么每次调整过后，应该使误差不断减少。记调整的量为 $\Delta \omega_{hj}, \Delta v_{ih}$ ，学习率为 η ，则有：

$$\Delta \omega_{hj} = -\eta \frac{\partial E}{\partial \omega_{hj}} \quad \Delta v_{ih} = -\eta \frac{\partial E}{\partial v_{ih}}$$

我们不难看出，每次权值改变量和学习率、本层输出的误差信号（即偏导计算的内容），以及本层的输入信号有关。

明确这些规则后，剩下的就是训练了。一般来说，B-P 神经网络的训练过程分为两个部分：

1. 前向传输，即逐层向后传递输出值。
2. 逆向反馈，即反向逐层调整权重和偏置。

先初始化权重和偏置，之后经过预期次数的迭代，或者达到预测的准确率时，即证明神经网络已经训练完成。

⁸ 即误差逆传播（error BackPropagation）算法，最早由 Paul J. Werbos 于 1974 年提出

⁹ 图片来源：https://blog.csdn.net/antkillerfarm/article/details/74187428?depth_1-utm_source=distribute.pc_relevant.none-task&utm_source=distribute.pc_relevant.none-task

五、B-P 模型的不足与对未来的展望

B-P 网络出色的非线性映射能力、泛化能力和容错能力,使其成为了迄今为止使用最为广泛的人工神经网络。但世界上没有完美的事物,B-P 网络自然也不例外。由于其自身算法的不足,导致了它存在诸多局限性和缺点,主要体现为¹⁰:

1. BP 算法学习过程收敛速度慢;
2. BP 算法所得到的网络性能差;
3. 因为误差平方和函数可能有局部极小点出现,故 BP 算法不完备;
4. BP 网络学习率不稳定。

这些缺陷也使得科学家们不满足于现状,不断的研究和探索可以使其改进的方法和更加先进的算法。比如 1988 年 Chua 和 Yang 提出的细胞神经网络(CNN)模型, Kosko 建立的具有非监督学习能力的双向联想存储模型(BAM)等等,还有近年来备受关注的深度学习(Deep Learning)

深度学习,由 Hinton 等人于 2006 年提出。作为机器学习(Machine Learning, ML)的一个新领域,它的引入使得机器学习更加接近于其最初的目标——人工智能。深度学习在语言和图像识别方面取得的成果,远远超过了先前的相关技术。通过模仿视听和思考等人累的活动,它解决了很多复杂的模式难题,使得相关技术取得了巨大的进步,同时也被各大互联网公司广泛应用。2011 年,微软公司推出了基于深度神经网络的语音识别系统,这一成果将语音识别领域已有的技术框架完全改变。国际上,谷歌也使用深层神经网络对声音进行建模,是最早在深度神经网络的工业化应用领域取得突破的企业之一。

现在这个时代,可以说是神经网络研究的黄金时代。互联网技术的高速发展,使得这方面的研究具有广泛的应用前景和巨大的潜在价值,现有算法理论的种种局限性和延展性也给了我们充分的研究空间和方向。相信在不远的将来,还会涌现更多的新算法,新技术,让我们的生活变得更加智能和便利,让科幻小说中的未来不再是未来。

¹⁰ 余本国. BP 神经网络局限性及其改进的研究[J]. 山西农业大学学报(自然科学版), 2009, 29(1):89-93.

参考:

- <https://blog.csdn.net/fengbingchun/article/details/50087005>
- <https://blog.csdn.net/htrdchh/article/details/87903720>
- <https://blog.csdn.net/u013007900/article/details/50118945>
- https://blog.csdn.net/antkillerfarm/article/details/74187428?depth_1-utm_source=distribute.pc_relevant.none-task&utm_source=distribute.pc_relevant.none-task
- <https://blog.csdn.net/u013007900/article/details/50066315>
- <https://wenku.baidu.com/view/9e0860dfec3a87c24028c45a.html>
- 余本国. BP 神经网络局限性及其改进的研究[J]. 山西农业大学学报(自然科学版), 2009, 29(1):89-93.