



神经计算是建立在神经元模型和学习规则基础上的一种计算范式,由于特殊的拓扑结构和学习方式,产生了多种人工神经网络,模仿人脑信息处理的机理。人工神经网络是由大量处理单元组成的非线性大规模自适应动力系统。

3.1 概述

神经计算(Neural Computing, NC)也称作人工神经网络(Artificial Neural Networks, ANN),神经网络(Neural Networks, NN),是对人脑或生物神经网络的抽象和建模,具有从环境学习的能力,以类似生物的交互方式适应环境。

现代神经网络研究开始于麦克洛奇(W. S. McCulloch)和皮兹(W. Pitts)的先驱工作。1943年,他们结合了神经生理学和数理逻辑的研究,提出了M-P神经网络模型,标志着神经网络的诞生。1949年,赫布(D. O. Hebb)的书《行为组织学》第一次清楚说明了突触修正的生理学习规则。

1986年,鲁梅尔哈特和麦克莱伦德(J. L. McClelland)编辑的《并行分布处理:认知微结构的探索(PDP)》一书出版^[50]。这本书对反向传播算法的应用引起重大影响,成为最通用的多层感知器的训练算法。后来证实,有关反向传播学习方法韦勃斯(P. J. Werbos)在1974年8月的博士学位论文中已有描述。

2006年,加拿大多伦多大学的辛顿(G. E. Hinton)及其学生提出了深度学习(deep learning),全世界掀起了深度学习的热潮。2016年3月8~15日,谷歌围棋人工智能AlphaGo与韩国棋手李世石比赛,AlphaGo最终以4:1的战绩取得了人机围棋对决的胜利。2019年3月27日,ACM(国际计算机学会)宣布,有“深度学习三巨头”之称的本吉奥、杨立昆(Yann LeCun)、辛顿共同获得了2018年的图灵奖,以表彰他们为当前人工智能的繁荣发展所奠定的基础。

大脑神经信息处理是由一组相当简单的单元通过相互作用完成的。每个单元向其他单元发送兴奋性信号或抑制性信号。单元表示可能存在的假设,单元之间的相互连接则表示单元之间存在的约束。这些单元的稳定的激活模式就是问题的解。鲁梅尔哈特等提出并行分布处理模型的8个要素:

- (1) 一组处理单元。
- (2) 单元集合的激活状态。
- (3) 各个单元的输出函数。
- (4) 单元之间的连接模式。
- (5) 通过连接网络传送激活模式的传递规则。
- (6) 把单元的输入和它的当前状态结合起来,以产生新激活值的激活规则。
- (7) 通过经验修改连接模式的学习规则。
- (8) 系统运行的环境。

并行分布处理系统的一些基本特点,可以从图 3.1 中看出来。这里有一组用圆图表示的处理单元。在每一时刻,各单元 u_i 都有一个激活值 $a_i(t)$ 。该激活值通过函数 f_i 而产生出一个输出值 $o_i(t)$ 。通过一系列单向连线,该输出值被传送到系统的其他单元。每个连接都有一个叫作连接强度或权值的实数 w_{ij} 与之对应;它表示第 j 个单元对第 i 个单元影响的大小和性质。采用某种运算(通常是加法),把所有的输入结合起来,就得到一个单元的净输入

$$\text{net}_j = w_{ij} o_i(t) \quad (3.1)$$

单元的净输入和当前激活值通过函数 F 的作用,就产生一个新的激活值。图 3.1 下方给出了函数 f 及 F 的具体例子。最后,在内部连接模式并非一成不变的情况下,并行分布处理模型是可塑的;更确切地说,权值作为经验的函数,是可以修改的,因此,系统能演化。单元表达的内容能随经验而变化,因而系统能用各种不同的方式完成计算。

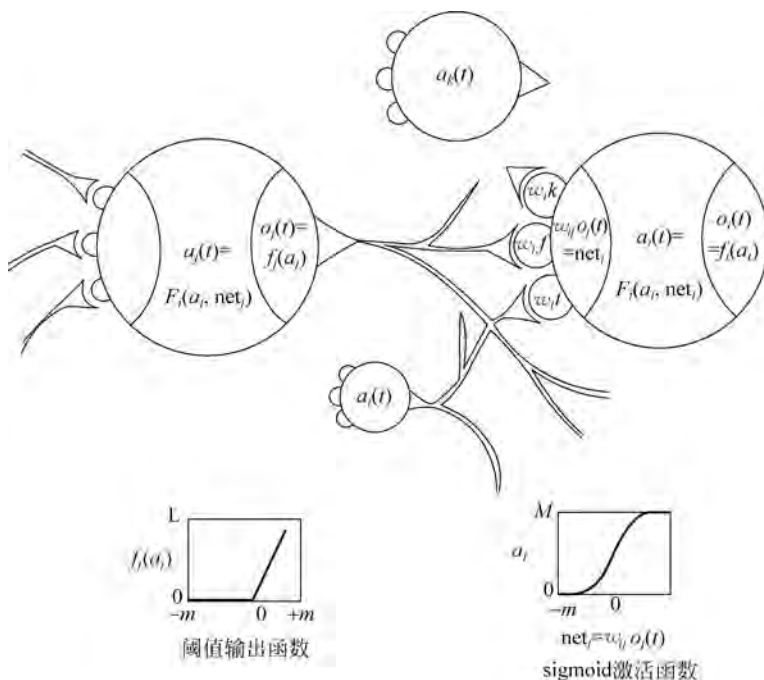


图 3.1 并行分布处理示意图



视频 10
前馈神经网络

3.2 前馈神经网络

一般的前馈网络包括一个输入层和一个输出层,若干隐单元。隐单元可以分层也可以不分层,若分层,则称为多层前馈网络。网络的输入、输出神经元其激活函数一般取为线性函数,而隐单元则为非线性函数。任意的前馈网络,不一定是分层网络或全连接的网络。下面给出它的定义和说明:一个前馈网络可以定义为无圈的有向图 $N=(V, E, W)$, 其中 $V=\{0, 1, \dots, n\}$ 为神经元集合, $E \in V \times V$ 为连接权值集合, $W: E \rightarrow R$ 为对每一连接 $(i, j) \in E$ 赋予一实值的权重 W_{ij} 。对神经元 $i \in V$ 定义它的投射域为 $P_i = \{j: j \in V, (i, j) \in E\}$, 即表示单元 i 的输出经加权后直接作为其净输入的一部分的神经单元; 同样定义神经元的接受域为 $R_i = \{j: j \in V, (i, j) \in E\}$, 即表示其输出经加权后直接作为神经元 i 的净输入的一部分的神经单元。特别地, 对分层前馈网络来说, 每个神经元的接受域和投射域分别是其所在层的前一层神经元和后一层神经元(若它们存在)。神经元集 V 可以被分成无接收域的输入节点集 V_I , 无投射域的输出节点集 V_O 和隐层节点集 V_H 。一般地, 假设一个特殊的偏置节点(在这里其标号为 0), 它的输出恒为 +1, 它和输入节点以外的所有节点相连。

多层前馈神经网络需要解决的关键问题是学习算法。以鲁梅尔哈特和麦克莱伦德为首的科研小组提出的误差反向传播(error Back Propagation, BP)算法, 为多层前馈神经网络的研究奠定了基础。多层前馈网络能逼近任意非线性函数, 在科学技术领域有广泛的应用。下面介绍多层前馈神经网络的误差反向传播算法, 这是基本的 BP 算法, 基于梯度法极小化二次性能指标函数

$$E = \sum_{k=1}^m E_k \quad (3.2)$$

式中, E_k 为局部误差函数, 即

$$E_k = \sum_{i=1}^{n_0} \phi(e_{i,k}) = \frac{1}{2} \sum_{k=1}^{n_0} (y_{i,k} - \hat{y}_{i,k})^2 = \sum_{i=1}^{n_0} e_{i,k}^2 \quad (3.3)$$

寻求目标函数的极小有两种基本方法, 即逐个处理和成批处理。所谓逐个处理, 即随机依次输入样本, 每输入一个样本都进行连接权的调整。所谓成批处理, 是在所有样本输入后计算其总误差进行的。下面介绍逐个处理。

以具有两个隐含层的多层神经网络为例。对于输出层连接权矩阵 $W^{(0)}$ 第 p 行的调整方程可表示为

$$\begin{aligned} \Delta W_{p,k}^{(0)} &= W_{p,k}^{(0)} - W_{p,k-1}^{(0)} = -\alpha \frac{\partial E_k}{\partial W_{p,k}^{(0)}} \\ &= -\alpha \sum_{i=1}^{n_0} \frac{\partial \phi(e_{i,k})}{\partial \hat{y}_{i,k}} \frac{\partial \sigma_0(\bar{y}_{i,k})}{\partial \bar{y}_{i,k}} \frac{\partial \bar{y}_{i,k}}{\partial W_{p,k}^{(0)}} \end{aligned}$$

因 $\bar{y}_{i,k} = \sum_{j=1}^{n_1} w_{i,j} \hat{h}_{j,k}^{(1)}$, 则

$$\Delta W_{p,k}^{(0)} = \alpha \sum_{i=1}^{n_0} e_{i,k} \sigma_0'(\bar{y}_{i,k}) \hat{h}_{i,k}^{(1)} \delta_{ip}$$

$$\hat{\mathbf{h}}_k^{(1)} = [\hat{h}_{1,k}^{(1)}, \hat{h}_{2,k}^{(1)}, \dots, \hat{h}_{n_1,k}^{(1)}]^T \quad (3.4)$$

$$\delta_{ip} = \begin{cases} 1, & i = p \\ 0, & i \neq p \end{cases} \quad (i = 1, 2, \dots, n_0)$$

考虑 δ_{jp} 的取值, 则

$$\Delta \mathbf{W}_{p,k}^{(0)} = \alpha \epsilon_{p,k}^{(0)} \hat{\mathbf{h}}_k^{(1)} \quad (p = 1, 2, \dots, n_0) \quad (3.5)$$

$$\epsilon_{p,k}^{(0)} = e_{p,k} \sigma'_{0}(\bar{y}_{p,k}) \quad (3.6)$$

对于第一隐含层, 连接权矩阵 $\mathbf{W}^{(1)}$ 第 p 行的调整方程可表示为

$$\begin{aligned} \Delta \mathbf{W}_{p,k}^{(1)} &= \mathbf{W}_{p,k}^{(1)} - \mathbf{W}_{p,k-1}^{(1)} = -\alpha \frac{\partial E_k}{\partial \mathbf{W}_p^{(1)}} \\ &= -\alpha \sum_{i=1}^{n_0} \frac{\partial \phi(e_{i,k})}{\partial \hat{y}_{i,k}} \frac{\partial \hat{y}_{i,k}}{\partial \mathbf{W}_p^{(1)}} \\ &= \alpha \sum_{i=1}^{n_0} e_{i,k} \sigma'_{0}(\bar{y}_{i,k}) \sum_{j=1}^{n_j} \mathbf{W}_{ij}^{(0)} \frac{\partial \sigma_1 \bar{h}_{j,k}^{(1)}}{\partial \bar{h}_{j,k}^{(1)}} \frac{\partial \bar{h}_{j,k}^{(1)}}{\partial \mathbf{W}_p^{(1)}} \\ \Delta \mathbf{W}_{p,k}^{(1)} &= \alpha \sum_{i=1}^{n_0} e_{i,k} \sigma'_{0}(\bar{y}_{i,k}) \sum_{j=1}^{n_j} \mathbf{W}_{ij}^{(0)} \frac{\partial \sigma_1 \bar{h}_{j,k}^{(1)}}{\partial \bar{h}_{j,k}^{(1)}} \frac{\partial \bar{h}_{j,k}^{(1)}}{\partial \mathbf{W}_p^{(1)}} \end{aligned}$$

因为 $\bar{h}_{j,k}^{(1)} = \sum_{i=1}^{n_2} \omega_{i,j} \bar{h}_{j,k}^{(2)}$, 则

$$\begin{aligned} \hat{\mathbf{h}}_k^{(2)} &= [\hat{h}_{1,k}^{(2)}, \hat{h}_{2,k}^{(2)}, \dots, \hat{h}_{n_2,k}^{(2)}]^T \\ \delta_{jp} &= \begin{cases} 1, & j = p \\ 0, & j \neq p \end{cases} \quad (j = 1, 2, \dots, n_1) \end{aligned}$$

考虑到 δ_{jp} 的取值, 则

$$\begin{aligned} \Delta \mathbf{W}_{p,k}^{(1)} &= \alpha \sigma_1(\bar{h}_{p,k}^{(1)}) \sum_{i=1}^{n_0} e_{i,k} \sigma'_{0}(\bar{y}_{i,k}) \mathbf{W}_{i,p}^{(0)} \hat{\mathbf{h}}_k^{(2)} \\ &= \alpha \epsilon_{p,k}^{(1)} \hat{\mathbf{h}}_k^{(2)} \quad (p = 1, 2, \dots, n_1) \end{aligned} \quad (3.7)$$

$$\epsilon_{p,k}^{(1)} = \sigma'_1(\bar{h}_{p,k}^{(1)}) \sum_{i=1}^{n_0} \epsilon_{i,k}^{(0)} \omega_{i,p} \quad (3.8)$$

同理, 对于第二隐含层连接权矩阵 $\mathbf{W}^{(2)}$ 第 p 行的调整方程可表示为

$$\begin{aligned} \Delta \mathbf{W}_{p,k}^{(2)} &= \mathbf{W}_{p,k}^{(2)} - \mathbf{W}_{p,k-1}^{(2)} \\ &= \alpha \epsilon_{p,k}^{(2)} \mathbf{X}_k \quad (p = 1, 2, \dots, n_2) \end{aligned} \quad (3.9)$$

$$\epsilon_{p,k}^{(2)} = \sigma'_2(\bar{h}_{p,k}^{(2)}) \sum_{i=1}^{n_1} \epsilon_{i,k}^{(1)} \omega_{i,p} \quad (3.10)$$

对于一般情况, 设隐含层数为 L , 第 r 隐含层连接权矩阵 $\mathbf{W}^{(r)}$ 第 p 行的调整方程为

$$\begin{aligned} \Delta \mathbf{W}_{p,k}^{(r)} &= \mathbf{W}_{p,k}^{(r)} - \mathbf{W}_{p,k-1}^{(r)} \\ &= \alpha \epsilon_{p,k}^{(r)} \hat{\mathbf{h}}_k^{(r+1)} \quad (p = 1, 2, \dots, n_r) \end{aligned} \quad (3.11)$$

$$\epsilon_{p,k}^{(r)} = \sigma'_r(\bar{h}_{p,k}^{(r)}) \sum_{i=1}^{n_{r-1}} \epsilon_{i,k}^{(r-1)} \omega_{i,p} \quad (3.12)$$

当 $r=L$ 时, $\hat{h}_k^{(L+1)} = X_k$ 。

由上面的分析可见,输出的局部误差 $\epsilon_{p,k}^{(0)}$ 决定于输出误差 $e_{p,k}$ 和该层变换函数的偏导数 $\sigma'_0(\cdot)$ 。隐含层局部误差 $\epsilon_{p,k}^{(r)} (r=0,1,2)$ 的计算是以高层的局部误差为基础的。即在计算过程中局部误差是由高层向低层反向传播的。

算法 3.1 误差反向传播 BP 算法。

- (1) 用小的随机数初始化 W 。
- (2) 输入一个样本,用现有的 W 计算网络各神经元的实际输出。
- (3) 根据式(3.6)、式(3.8)和式(3.10)计算局部误差 $\epsilon_{p,k}^{(r)} (r=0,1,2)$ 。
- (4) 根据递推计算公式(3.5)、式(3.7)和式(3.9)计算 $\Delta W_{p,k}^{(r)} (r=0,1,2)$ 。
- (5) 输入另一样本,转步骤(2)。

所有训练样本是随机地输入,直到网络收敛且输出误差小于容许值。

上述的 BP 算法存在如下缺点:

- (1) 为了极小化总误差,学习速率 α 应选得足够小,但是小的 α 学习过程将很慢;
- (2) 大的 α 虽然可以加快学习速度,但又可能导致学习过程的振荡,从而收敛不到期望解;
- (3) 学习过程可能收敛于局部极小点或在误差函数的平稳段停止不前。

针对 BP 算法收敛速度慢的问题,研究工作者提出了很多改进方法。在这些方法中,通过学习速率的调整以提高收敛速度的方法被认为是一种最简单、最有效的方法。

3.3 自适应共振理论 ART 模型



视频 11
自适应共振
理论 ART
模型

自适应共振理论(Adaptive Resonance Theory, ART)模型是美国波士顿大学的格罗斯伯格(S. Grossberg)在 1976 年提出的。ART 是一种自组织神经网络结构,是无监督的学习网络。当神经网络和环境有交互作用时,对环境信息的编码会自发地在神经网中产生,此时认为神经网络在进行自组织活动。ART 就是这样一种能自组织地产生对环境认识编码的神经网络理论模型。

ART 模型是基于下列问题的求解而提出的:

- (1) 对于一个学习系统,要求它有适应性及稳定性,适应性可以响应重要事件,稳定性可以存储重要事件。这种系统的设计问题。
- (2) 学习时,原有的信息和新信息如何处理,保留有用知识,接纳新知识的关系如何解决的问题。
- (3) 对外界信息与原来存储的信息结合并决策的问题。

格罗斯伯格一直对人类的心理和认识活动感兴趣,他长期埋头于这方面的研究并希望用数学来描述人类这项活动,建立人类的心理和认知活动的一种统一的数学模型和理论。ART 就是在这种理论的核心内容基础上经过提高发展得出的。

ART 模型源于 Helmholtz 无意识推理学说的协作-竞争网络交互模型^[26,27]。这个模型如图 3.2 所示。可以看出这个模型由两个协作-竞争模型组成。无意识推理学说认为:原始的感觉信息通过经历过的学习过程不断修改,直到得到一个真实的感知结果为止。从

图 3.2 中协作-竞争网络交互模型可以看出;环境输入信号和自上而下学习期望同时对协作-竞争网络 1 执行输入;而自下而上学习是协作-竞争网络 1 的输出;同时,自下而上学习是协作-竞争网络 2 的输入,而自上而下学习期望则是其输出。真实感知是通过这个协作-竞争网络的学习和匹配产生的。



图 3.2 协作-竞争网络交互模型

环境输入信号对自上而下学习期望进行触发,使协作-竞争网络 1 产生自下而上学习的输出。输出发送到协作-竞争网络 2,则产生自上而下学习期望输出,并送回协作-竞争网络 1。这个过程很明显是自上而下学习和自下而上学习的过程,并且这个过程中不断吸收环境输入信息。经过协作-竞争的匹配,最终取得一致的结果;这就是最终感知或谐振感知。协作-竞争网络交互作用有下列基本要求:

- (1) 交互作用是非局域性的;
- (2) 交互作用是非线性的;
- (3) 自上而下的期望学习是非平稳随机过程。

受到协作-竞争网络交互模型的启发,格罗斯伯格提出了 ART 理论模型。他认为对网络的自适应行为进行分析,可以建立连续非线性网络模型,这种网络可以由短期存储 STM 和长期存储 LTM 作用所实现。STM 是指神经元的激活值,即未由 S 函数处理的输出值, LTM 是指权系数。

格罗斯伯格提出的 ART 理论模型具有如下主要优点:

- (1) 可以进行实时学习,能适应非平稳的环境。
- (2) 对于已经学习过的对象具有稳定的快速识别能力;同时,亦能迅速适应未学习的新对象。
- (3) 具有自归一能力,根据某些特征在整体中所占的比例,有时作为关键特征,有时当作噪声处理。
- (4) 不需要预先知道样本结果,是无监督学习;如果对环境作出错误反应则自动提高“警觉性”,迅速识别对象。
- (5) 容量不受输入通道数的限制,存储对象也不需要是正交的。

ART 的基本结构如图 3.3 所示。它由输入神经元和输出神经元组成。用前向权系数和样本输入来求取神经元的输出,这个输出也就是匹配测度;具有最大匹配测度的神经元的活跃级通过输出神经元之间的横向抑制得到进一步增强,而匹配测度不是最大的神经元的活跃级就会逐渐减弱,从输出神经元到输入神经元之间有反馈连接以进行学习比较。同样,还提供一个用来确定具有最大输出的输出神经元与输入模式进行比较的机制。ART 模型的框如图 3.4 所示。

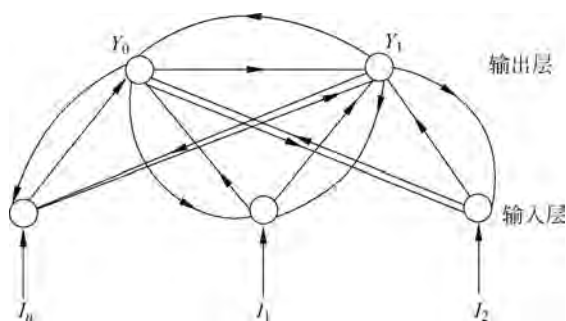


图 3.3 ART 的基本结构

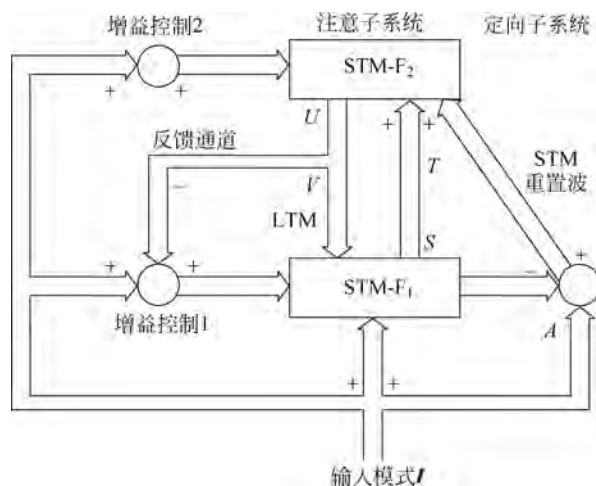


图 3.4 ART 模型的框图

它由两个子系统组成：一个称为注意子系统(attentional subsystem)；一个称为定向子系统(orienting subsystem)，也称调整子系统。这两个子系统是功能互补的子系统。ART模型就是通过这两个子系统和控制机制之间的交互作用来处理熟悉的事件或不熟悉的事件。在注意子系统中，有 F_1 、 F_2 这两个用短时记忆单元组成的部件，即 $STM-F_1$ 和 $STM-F_2$ 。在 F_1 和 F_2 之间的连接通道是长时记忆 LTM。增益控制有两个作用：一个作用是在 F_1 中用于区别自下而上和自上而下的信号；另一个作用是当输入信号进入系统时， F_2 能够对来自 F_1 的信号起阈值作用。调整子系统是由 A 和 STM 重置波通道组成。

注意子系统的作用是对熟悉事件进行处理。在这个子系统中建立熟悉事件对应的内部表示，以便响应有关熟悉事件；这实际上是对 STM 中的激活模式进行编码。同时，在这个子系统中还产生一个从 F_2 到 F_1 的自上而下的期望样本，以帮助稳定已被学习了熟悉事件的编码。

调整子系统的作用是对不熟悉的事件产生响应。在有不熟悉的事件输入时，孤立的一个注意子系统无法对不熟悉的事件建立新的聚类编码；故而设置一个调整子系统，当有不熟悉的事件输入时，调整子系统马上产生重置波对 F_2 进行调整，从而使注意子系统对不熟悉的事件建立新的表达编码。实际上，当自下而上的输入模式和来自 F_2 的自上而下的引发模式，即期望在 F_1 中不匹配时，调整子系统就会发出一个重置波信号到 F_2 ，它重新选择 F_2 。

的激活单元,同时取消 F_2 原来所发出的输出模式。

简言之,注意子系统的功能是完成自下而上的向量的竞争选择,以及完成自下而上向量和自上而下向量的相似度比较。而取向子系统的功能是检验期望向量 V 和输入模式 I 的相似程度;当相似度低于某一给定标准值时,即取消该时的竞争优胜者,转而从其余类别中选取优胜者。

ART 模型就是由注意子系统和调整子系统共同作用,完成自组织过程的。在 ART 模型中,其工作过程采用 2/3 规则。所谓 2/3 规则,就是在 ART 网络中,3 个输入信号中要有 2 个信号起作用才能使神经元产生输出信号。

ART 理论已提出了 3 种模型结构,即 ART1、ART2、ART3。ART1 用于处理二进制输入的信息;ART2 用于处理二进制和模拟信息这两种输入;ART3 用于进行分级搜索。ART 理论可以用于语音、视觉、嗅觉和字符识别等领域。

2008 年,格罗斯伯格等提出了同步匹配自适应共振理论 SMART (synchronous matching adaptive resonance theory) 模型^[28],以反映大脑是怎样协调多级的丘脑和皮质进程来快速学习、稳定记忆外界的重要信息。同步匹配适应共振理论 SMART 模型,展示了自底向上和自顶向下的通路是如何一起工作并通过协调学习、期望、专注、共振和同步这几个进程来完成上述目标的。

3.4 神经网络集成

1990 年,汉森(L. K. Hansen)和萨拉蒙(P. Salamon)提出了神经网络集成(neural network ensemble)方法。他们证明,可以简单地通过训练多个神经网络并将其结果进行拟合,显著地提高神经网络系统的泛化能力。神经网络集成可以定义为用有限个神经网络对同一个问题进行学习,集成在某输入示例下的输出由构成集成的各神经网络在该示例下的输出共同决定。对神经网络集成的理论分析与其实现方法分为两方面,即对结论生成方法以及对网络个体生成方法。

3.4.1 结论生成方法

汉森和萨拉蒙证明,对神经网络分类器来说,采用集成方法能够有效提高系统的泛化能力。假设集成由 N 个独立的神经网络分类器构成,采用绝对多数投票法,再假设每个网络以 $1-p$ 的概率给出正确的分类结果,并且网络之间错误不相关,则该神经网络集成发生错误的概率 p_{err} 为

$$p_{\text{err}} = \sum_{k > N/2}^N \binom{N}{k} p^k (1-p)^{N-k} \quad (3.13)$$

当 $p < 1/2$ 时, p_{err} 随 N 的增大而单调递减。因此,如果每个神经网络的预测精度都高于 50%,并且各网络之间错误不相关,则神经网络集成中的网络数目越多,集成的精度就越高,当 N 趋向于无穷时,集成的错误率趋向于 0。在采用相对多数投票法时,神经网络集成的错误率比式(3.13)复杂得多,但是汉森和萨拉蒙的分析表明,采用相对多数投票法在多数情况下能够得到比绝对多数投票法更好的结果。

1995年,克罗夫(A. Krogh)和弗德尔斯毕(J. Vedelsby)给出了神经网络集成泛化误差计算公式。假设学习任务是利用 N 个神经网络组成的集成对 $f: \mathbf{R}^n \rightarrow \mathbf{R}$ 进行近似,集成采用加权平均,各网络分别被赋予权值 w_a ,并满足式(3.14)和式(3.15):

$$w_a > 0 \quad (3.14)$$

$$\sum_a w_a = 1 \quad (3.15)$$

再假设训练集按分布 $p(x)$ 随机抽取,网络 a 对输入 X 的输出为 $V^a(X)$,则神经网络集成的输出为

$$\bar{V}(X) = \sum_a w_a V^a(X) \quad (3.16)$$

神经网络 a 的泛化误差 E^a 和神经网络集成的泛化误差 E 分别为

$$E^a = \int dx p(x) (f(x) - V^a(x))^2 \quad (3.17)$$

$$E = \int dx p(x) (f(x) - \bar{V}(x))^2 \quad (3.18)$$

各网络泛化误差的加权平均为

$$\bar{E} = \sum_a w_a E^a \quad (3.19)$$

神经网络 a 的差异度 A^a 和神经网络集成的差异度 $\bar{A}$ 分别为

$$A^a = \int dx p(x) (V^a(x) - \bar{V}(x))^2 \quad (3.20)$$

$$\bar{A} = \sum_a w_a A^a \quad (3.21)$$

则神经网络集成的泛化误差为

$$E = \bar{E} - \bar{A} \quad (3.22)$$

式(3.22)中的 $\bar{A}$ 度量了神经网络集成中各网络的相关程度。若集成是高度偏置的,即对于相同的输入,集成中所有网络都会给出相同或相近的输出,此时集成的差异度接近于0,其泛化误差接近于各网络泛化误差的加权平均。反之,若集成中各网络是相互独立的,则集成的差异度较大,其泛化误差将远小于各网络泛化误差的加权平均。因此,要增强神经网络集成的泛化能力,就应该尽可能地使集成中各网络的误差互不相关。

3.4.2 个体生成方法

1997年,弗洛德(Y. Freund)和沙皮尔(R. E. Schapire)以 AdaBoost 为代表,对 Boosting 类方法进行了分析,并证明此类方法产生的最终预测函数 H 的训练误差满足式(3.23),其中 ϵ_i 为预测函数 h_i 的训练误差, $\gamma_i = 1/2 - \epsilon_i$ 。

$$\begin{aligned} H &= \prod_t [2\sqrt{\epsilon_i(1-\epsilon_i)}] \\ &= \prod_t \sqrt{1-4\gamma_i^2} \leq \exp(-2t \sum_i \gamma_i^2) \end{aligned} \quad (3.23)$$

从式(3.23)可以看出,只要学习算法略好于随机猜测,训练误差将随 t 以指数级下降。

1996年,布雷曼(L. Breiman)对 Bagging 进行了理论分析。他指出,分类问题可达到的最高正确率以及利用 Bagging 可达到的正确率分别如式(3.24)和式(3.25)所示,其中 C 表

示序正确的输入集, C' 为 C 的补集, $I(\cdot)$ 为指示函数(indicator function)。

$$r^* = \int \max_j P(j | x) P_X(x) \quad (3.24)$$

$$r_A = \int_{x \in C} \max_j P(j | x) P_X(x) dx + \int_{x \in C'} \left[\sum_j I(\phi_A(x) = j) P(j | x) \right] P_X(x) \quad (3.25)$$

显然, Bagging 可使序正确集的分类正确率达到最优, 单独的预测函数则无法做到这一点。



视频 12
脉冲耦合神经网络

3.5 脉冲耦合神经网络

近年来, 随着生物神经学的研究和发展, 艾克霍恩(R. Eckhorn)等通过对小型哺乳动物大脑视觉皮层神经系统工作机理的仔细研究, 提出了一种崭新的网络模型——脉冲耦合神经网络(Pulse-Coupled Neural Network, PCNN)模型。PCNN 来源于对哺乳动物猫的视觉皮层神经细胞的研究成果, 具有同步脉冲激发现象、阈值衰减及参数可控性等特性。由于其具有生物学特性的背景以及以空间邻近和亮度相似集群的特点, 因此在数字图像处理等领域具有广阔的应用前景。将 PCNN 的最新理论研究成果与其他新技术相结合, 开发出具有实际应用价值的新算法是当今神经网络研究的主要方向之一。

1952 年, 霍奇金(A. L. Hodgkin)与哈斯利(A. F. Huxley)开始研究神经元电化学特性^[33]。1987 年, 格雷(C. M. Gray)等发现猫的初生视觉皮层有神经激发相关振荡现象^[25]。1989 年, 艾克霍恩和格雷研究了猫的视觉皮层, 提出了具有脉冲同步发放特性的网络模型。1990 年, 艾克霍恩根据猫的大脑皮层同步脉冲发放现象, 提出了展示脉冲发放现象的连接模型。对猴的大脑皮层进行的试验中, 也得到了类似的试验结果。1994 年, 约翰逊(J. L. Johnson)发表论文, 阐述了 PCNN 的周期波动现象及在图像处理中具有旋转、可伸缩、扭曲、强度不变性。通过对艾克霍恩提出的模型进行改进, 形成了脉冲耦合神经网络(PCNN)模型。1999 年, IEEE 神经网络会刊出版了脉冲耦合神经网络专辑; 国内也于 20 世纪 90 年代末开始研究脉冲耦合神经网络。

与传统方法相比, 源自哺乳动物视觉皮层神经元信息传导模型的脉冲耦合神经网络是一种功能强大的图像处理工具, 解决图像处理具体应用问题时能够取得令人满意的性能。

3.5.1 Eckhorn 模型

1990 年, 根据猫的视皮层的同步振荡现象, 艾克霍恩提出了一个脉冲神经网络模型^[18], 如图 3.5 所示。这个模型由许多相互连接的神经元构成, 每个神经元包括两个功能上截然不同的输入部分, 分别是常规的馈接(feeding)输入和起调制作用的链接(linking)输入。而这两部分的关系并非像传统神经元那样是加耦合的关系, 而是乘耦合的关系。

Eckhorn 模型可用如下方程描述:

$$U_{m,k} = F_k(t) [1 + L_k(t)] \quad (3.26)$$

$$F_k(t) = \sum_{i=1}^N [\omega_{ki}^f Y_i(t) + S_k(t) + N_k(t)] \otimes I(V^a, \tau^a, t) \quad (3.27)$$

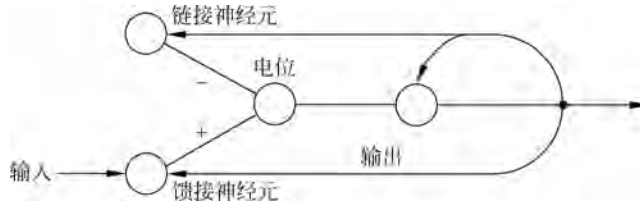


图 3.5 Eckhorn 模型示意图

$$L_k(t) = \sum_{i=1}^N [\omega_{ki}^l Y_i(t) + N_k(t)] \otimes I(V^l, \tau^l, t) \quad (3.28)$$

$$Y_k(t) = \begin{cases} 1, & U_{m,k}(t) \geq \theta_k(t) \\ 0, & \text{其他} \end{cases} \quad (3.29)$$

这里,一般表示为

$$X(t) = Z(t) \otimes I(v, \tau, t) \quad (3.30)$$

即

$$X[n] = X[n-1]e^{-t/\tau} + VZ[n], \quad n = 1, 2, \dots, N \quad (3.31)$$

其中, N 为神经元的个数, ω 为突触加权系数。当外部激励为 S 型时, Y 为二值输出。

3.5.2 脉冲耦合神经网络模型

由于 Eckhorn 模型提供了一个简单有效的方法来研究脉冲神经网络中的动态同步振荡活动, Eckhorn 模型的最大创新在于它引入了第二个感受野(secondary receptive field), 即链接域(linking field)。如果去掉链接输入部分, Eckhorn 模型中的神经元模型与常规的神经元模型没什么不同, 而正是链接输入的引入, 使我们神经元如何整合输入有了更深入的认识。通过对模型中神经元的电路进行分析, 研究人员证明了: 神经元的不同输入之间的关系不仅有加耦合的关系, 而且有乘耦合的关系。它很快被应用到图像处理领域, 而它的许多变种模型被一起称为脉冲耦合神经网络(PCNN)。

图 3.6 给出了脉冲耦合神经元的示意图。神经元主要由两个功能单元构成: 链接输入域和链接输入域, 分别通过突触连接权值 $\mathbf{M}$ 和 $\mathbf{K}$ 来与其邻近的神经元相连。两功能单元都要进行迭代运算, 迭代过程中按指数规律衰减。链接输入域多加一个外部激励 S 。可以用如下数学公式描述两个功能单元:

$$F_{ij}[n] = e^{aF\delta^n} F_{ij}[n-1] + S_{ij} + V_F \sum_{kl} \mathbf{M}_{ijkl} Y_{kl}[n-1] \quad (3.32)$$

$$L_{ij}[n] = e^{aL\delta^n} L_{ij}[n-1] + V_L \sum_{kl} \mathbf{K}_{ijkl} Y_{kl}[n-1] \quad (3.33)$$

式中, F_{ij} 是第 (i, j) 个神经元的链接, L_{ij} 是耦合连接, Y_{kl} 是 $(n-1)$ 次迭代时神经元的输出。两功能单元都要进行迭代运算, 迭代过程按指数规律衰减。 V_F 和 V_L 分别为 F_{ij} 、 L_{ij} 的固有电位。这里 $\mathbf{M}$ 和 $\mathbf{K}$ 为连接权值系数矩阵, 表示中心神经元受周围神经元影响的大小, 反映邻近神经元对中心神经元传递信息的强弱, $\mathbf{M}$ 和 $\mathbf{K}$ 有多种取值选择方式, 但选择要合适, 一般不宜过大。

神经元内部活动项由这两个功能单元按非线性相乘方式共同组成, β 为突触之间的连

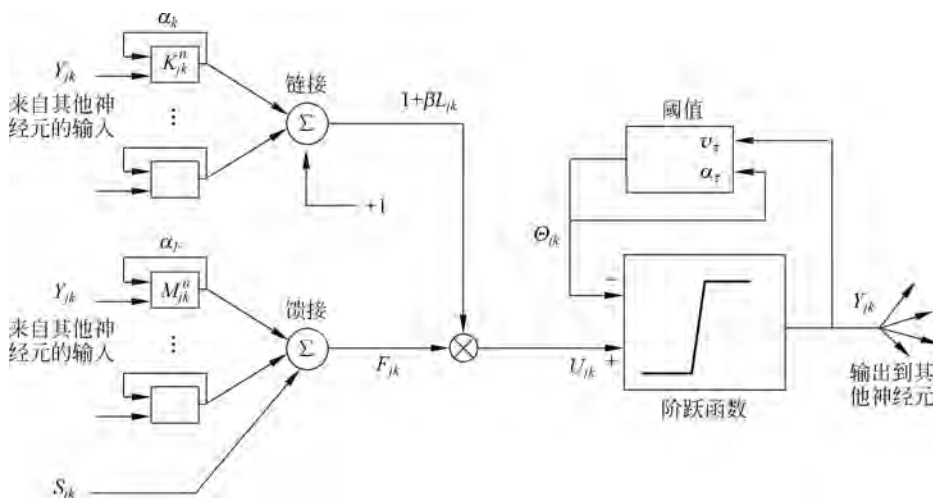


图 3.6 脉冲耦合神经元示意图

接强度系数。神经元内部活动项的数学表达式如下：

$$U_{ij}[n] = F_{ij}[n] \{1 + \beta L_{ij}[n]\} \quad (3.34)$$

当神经元内部活动项大于动态阈值 Θ 时，产生输出时序脉冲序列 Y ，即下式所示：

$$Y_{ij}[n] = \begin{cases} 1, & U_{ij}[n] > \Theta_{ij}[n] \\ 0, & \text{其他} \end{cases} \quad (3.35)$$

动态阈值在迭代过程中衰减，当神经元激发兴奋 ($U > \Theta$) 时，动态阈值立刻增大，然后又按指数规律逐渐衰减，直到神经元再次激发兴奋。这个过程可描述为

$$\Theta_{ij}[n] = e^{a\Theta_{ij}} \Theta_{ij}[n-1] + V_{\Theta} Y_{ij}[n] \quad (3.36)$$

式中， Θ 一般取一个比较大的值，相比 U 的均值还大一个数量级。

PCNN 由这些神经元排列(通常是矩阵)而成。 M 和 K 在神经元间传递信息通常是局部的，并符合高斯正态分布，但不必严格要求这样。矩阵 F 、 L 、 U 、 Y 初始化时，设其所有矩阵元素为零。 Θ 元素的初始值可以是 0，也可以根据实际需要设为某些更大值。任何有激励的神经元都将在第一次循环中激发兴奋，结果将生成一个很大的阈值。接下来需要经过几次循环才能使阈值衰减到足以使神经元再次激发兴奋。后者的情况趋向于围绕这些信息量小的初始循环。

本算法循环计算式(3.32)~式(3.36)，直到用户决定停止。目前 PCNN 本身还没有自动停止的机制。

与传统神经网络相比，PCNN 具有自己鲜明的特色，它具有如下特性：

(1) 变阈值特性。PCNN 中各神经元之所以能动态发放脉冲，是因为它内部的变阈值函数作用的结果。由式(3.36)可见，它是随时间按指数规律衰减的。当神经元的内部行为 U 大于当前的阈值输出值时就发放。对于无连接耦合的 PCNN 来说，每一时刻的发放图对应于该阈值下的二值图像帧。对于存在连接耦合的 PCNN 来说，每一时刻的发放图对应于该阈值下带有捕获功能的二值图像帧。

(2) 捕获特性。PCNN 的捕获过程就是使亮度强度相似的输入神经元能够同步发放脉冲，而同步的结果就好像把低亮度强度提升至先发放的那个神经元对应输入的亮度强度。

这就意味着因捕获可使得某一神经元的先发放,而激励或带动邻近其他神经元提前点火。PCNN 神经元间存在链接但不一定存在影响,存在影响但不一定存在链接,这一现象更加突出了 PCNN 对突发事件的处理能力,表现在由于某种原因(如噪声)使得网络原本已经组织好的有序状态,因某个或某些神经元点火状态的改变而被打破时,网络可自动地适应新的变化,实现对信息的重新组织,进而达到一个新的有序状态。

(3) 动态脉冲发放特性。PCNN 动态神经元的变阈值特性是其动态脉冲发放的根源,如果将由输入信号与突触通道的卷积和所产生的信号称为该神经元的(内部)作用信号,则当作用信号超过阈值时,该神经元被激活而产生高电平,又由于阈值受神经元输出控制,因此该神经元输出的高电平又反过来控制阈值的提高,从而作用信号在阈值以下,神经元又恢复为原来的抑制状态(即低电平)。这一过程在神经元输出上明显地形成一个脉冲发放。

(4) 同步脉冲发放特性。PCNN 每个神经元有一个输入,并与其他神经元的输出有链接。当一个神经元发放时,它会将其信号的一部分送至与其相邻的神经元上。从而这一链接会引起邻近神经元比其原来更快地点火,这就导致了在图像的一个大的区域内产生同步振荡:以相似性集群产生同步脉冲发放,这一性质对于图像平滑、分割、图像自动目标识别、融合等具有重要的应用意义。

(5) PCNN 时间序列。在点火捕获及脉冲传播特性的基础上,PCNN 能够由二进制图像生成一维向量信息: $G[n] = \sum Y_{ij}[n]$ 。对时间序列信号进行分析,可以达到识别图像的目的。

3.5.3 贝叶斯连接域神经网络模型

与 Eckhorn 模型类似,我们提出的贝叶斯连接域神经网络(Bayesian Linking Field Network, BLFN)模型也是一个由众多神经元构成的网络模型,而且模型中的神经元都包含两类输入:一类是馈接输入,另一类是链接输入,两类输入之间的耦合关系是相乘。与 Eckhorn 模型不同的是:为了解决特征捆绑的问题,我们还引入了噪声神经元模型的思想、贝叶斯方法和竞争机制。

图 3.7 给出了模型中的一个神经元输入耦合方式的示意^[90]。由于模型中神经元的输出是发放概率,所以输入的耦合实际上是各个传入神经元的发放概率的耦合。

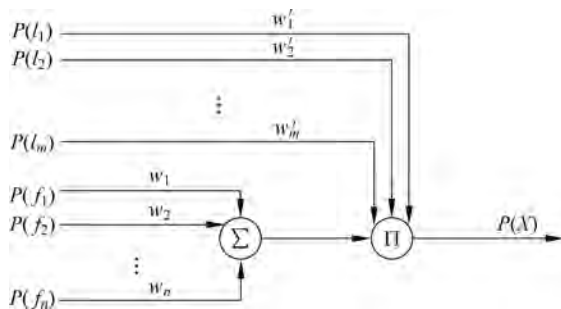


图 3.7 BLFN 模型中神经元输入耦合示意

BLFN 模型是一个由神经元构成的网络,它具有如下特点:

(1) 它采用噪声神经元模型,即每个神经元的输入和输出都是发放概率,而不是脉

冲值。

- (2) 每个神经元可以包含两部分输入：分别是馈接输入和链接输入。
- (3) 神经元之间的连接权反映了它们之间的统计相关性,是通过学习得到的。
- (4) 神经元的输出除了受输入影响,还受到竞争的制约。

3.6 超限学习机

单隐层前馈神经网络(Single-hidden Layer Feedforward Neural network, SLFN)之所以能够在很多领域得到广泛应用,是因为它具有如下优点:

- (1) 具有很强的学习能力,能够逼近复杂非线性函数;
- (2) 能够解决传统参数方法无法解决的问题。

另一方面,它缺乏快速学习方法,也使其很多时候无法满足实际需要。产生这种情况的主要原因是:

- (1) 传统的误差反向传播方法主要基于梯度下降的思想,需要多次迭代;
- (2) 网络的所有参数都需要在训练过程中迭代确定。

黄广斌等研究了有限集情况下 SLFN 的学习能力,只和隐层节点的数目有关,而和输入层的权值无关。在此基础上,黄广斌提出了超限学习机(Extreme Learning Machine, ELM)^[36],设置合适的隐层节点数,为输入权和隐层偏差进行随机赋值,然后输出层权值通过最小二乘法得到。整个过程一次完成,无须迭代,与 BP 相比速度显著提高。超限学习机的结构如图 3.8 所示。

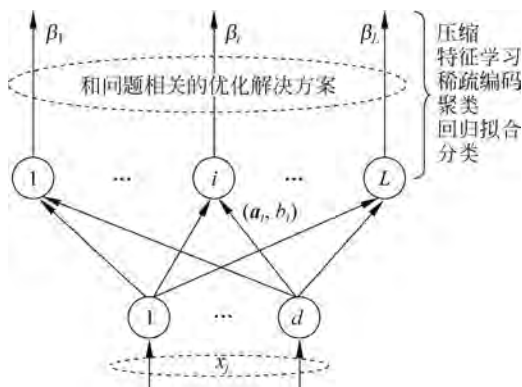


图 3.8 超限学习机的结构

在图 3.8 中,超限学习机的网络输出:

$$f_L(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x) \quad (3.37)$$

超限学习机的特征映射:

$$\mathbf{h}(x) = [G(a_1, b_1, x), \dots, G(a_L, b_L, x)] \quad (3.38)$$

隐层节点的输出函数

$$\text{Sigmoid: } G(a_i, b_i, x) = g(a_i \cdot x + b_i)$$

$$\text{RBF: } G(\mathbf{a}_i, b_i, \mathbf{x}) = g(b_i \|\mathbf{x} - \mathbf{a}_i\|) \quad (3.39)$$

$$\text{Fourier Series: } G(\mathbf{a}_i, b_i, \mathbf{x}) = \cos(\mathbf{a}_i \cdot \mathbf{x} + b_i)$$

3.7 功能柱神经网络模型

自1957年莫特卡斯勒(V. B. Mountcastle)发现功能柱结构以来,已有许多研究结果表明,在不同物种(鼠、猫、兔、猴和人等)的视皮层、听皮层、体感皮层、运动皮层以及其他联合皮层中都存在功能柱结构。这些结果表明,功能柱是皮层中一种普遍存在的结构,是结构和生理上的基本单元,这些柱的活动构成了整个大脑皮层活动的基础。

为了深刻地理解功能柱的生物学意义和在信息加工中所起的作用,研究者开展了许多数学建模研究。模型研究中最常见的是采用Wilson-Cowan方程来描述功能柱,例如,舒斯特(H. G. Shuster)等模拟视皮层中发现的同步振荡现象;詹森(B. H. Jansen)等提出了耦合功能柱模型,产生了类EEG波形和诱发电位;富凯(T. Fukai)设计了功能柱式的网络模型来模拟视觉图样的获取等等。还有一些功能柱模型是描述功能柱振荡活动的相位模型。只有少数模型是基于单神经元的,如:弗朗森(E. Fransén)等把传统网络中的单细胞代换成多细胞构成的功能柱,构建了一个吸引子网络,来模拟工作记忆;汉塞勒(D. Hansel)等根据视皮层朝向柱的结构构建了一个超柱模型,研究其中的同步性和混沌特性,并对朝向选择性的功能柱机理做出解释。

2005年,瑞士洛桑理工学院的科学家马克拉姆(H. Markram)与IBM公司合作开展蓝脑工程研究^[45],希望复制人类大脑,以达到治疗阿尔茨海默氏症和帕金森氏症的目的。在2006年年底,蓝脑工程已经创建了大脑皮质功能柱的基本单元模型。2008年,IBM公司使用蓝色基因巨型计算机,模拟具有5500万个神经元和5000亿个突触的老鼠大脑。

3.8 神经元集群的编码和解码

脑的信息编码的研究由来已久。20世纪20年代,阿德里安(Adrian)提出神经动作电位的概念,他在20世纪30年代进行的实验工作,为揭示大脑信息处理提供了一些基本线索。从1949年赫布(Hebb)提出的经典细胞群假设^[31],到1972年巴洛(Barlow)的单个神经元的编码假设,以及1996年藤井(Fujii)等提出的动态神经元的集群时空编码假设^[23],不同观点间的争论仍在继续。其中重要问题是:是单个神经元还是神经元集群编码刺激信息?是神经元动作电位出现的明确时间还是电位脉冲的平均发放速率携带信息?由于神经系统的高度复杂性,利用现有的实验手段还不能彻底解决神经元信息编码原理。但是现在已有越来越多的实验结果提示我们,神经系统中信息的编码与处理在很大程度上是在特定的发放频率与发放模式的框架下,通过大量神经元构成的集群编码活动完成的。在神经元集群中,每个神经元的活动特性都有其自身的特点,因而存在一定的差异性。然而,它们通过短暂的相关性活动与其他神经元进行相互协调,以神经元群体的整体活动或神经元活动的动态相关关系为特征,来实现对多种信息的并行处理和传递。

目前集群编码作为大脑信息处理的一种通用模型,主要是基于单个神经元对刺激的反

应是充满噪声的并且对刺激值的变化缺乏灵敏性这样的实验事实,因此具有代表性的单个神经元所携带的信息是非常低的。大脑要克服这种局限,就必须将信息分配给拥有大量数目的神经元集群来共同携带关于刺激的精确信息。集群编码的一个关键特性在于其鲁棒性和可塑性,由于信息的编码是在许多神经元共同活动的基础上得以完成的,因此单个神经元的损伤不至于在太大程度上影响编码过程。集群编码还具有其他一些优点,例如可以降低噪声水平,并有助于短时程信息存储的形成等;同时这种编码方式也具有复杂性和非线性等特性。

神经元集群编码的一种方式经典放电率模型意义下的群体编码。在早期的研究工作中,人们通过单位时间内动作电位的放电次数,对给定刺激作用下神经元的响应进行描述。这个测量值称为放电率,它一般由刺激诱导的放电率的平均响应(典型情况下呈钟形分布)和叠加于其上的噪声部分组成,噪声在每次测量时都有变化。早期人们的注意力主要集中在放电率上,因为该参量较为简单,易于测量且易于理解。虽然不能包含其所代表的各种各样的神经信息,比如刺激强度的大小;虽然仍没有完全了解神经信息是如何通过动作电位来编码的,但是动作电位作为神经信息编码的基本单位是确定的。当然响应的其他方面的特性,譬如动作电位发生的精确时间关系,即放电序列模式对信息编码来说同样具有重要的意义。

考虑在不同噪声水平和神经元相关性的影响下,通过给定刺激条件下观察记录到的神经元活动,建立描述外界刺激与神经元响应间的对应关系的概率模型已成为研究集群编码的普遍方法。基于这种共识,产生了大量分析集群编码与解码的研究。贝叶斯推理法则是研究神经元集群编码与解码的关键,是量化编码与解码行为的重要方法。早在1998年,泽梅尔(R. S. Zemel)就给出了贝叶斯原理框架下神经元集群编码与解码活动的概率解释,比较了在外界刺激诱导条件下神经元放电活动的泊松模型、KDE(Kernel Density Estimation)模型与扩展泊松(Extended Poisson)模型的性能,包括编码、解码、似然度与误差分析比较。近年来的理论研究表明大脑中包括编码与解码的神经计算过程类似于贝叶斯推理过程^[69];目前贝叶斯方法已被成功用于感知与感觉控制的神经计算理论,并且心理物理学上不断涌现的证据也表明大脑的感知计算是贝叶斯最优的,这也导致了尼尔(D. C. Kniil)等将之称为贝叶斯编码假说。从记录到的神经元放电活动中重构外界刺激或刺激的某些特性,贝叶斯推理为揭示这样的解码过程行为提供了可能。葛杨和蒋文新探讨了采用逻辑回归混合模型的贝叶斯推断的一致性。

神经元集群编码与解码是神经信息处理的关键问题,是揭示大脑工作机理的理论框架。它的发展能够促进人们对脑的总体功能的认识,为研究更为复杂的高级认知功能提供基本理论与指导方法。基于贝叶斯原理的编码与解码方法能够从总体上大致揭示神经系统信息处理过程的特性,对脑的工作机理作出客观合理的数学解释。

3.9 小结

神经计算是通过对人脑工作机理的简单模仿,是建立在简化的神经元模型和学习规则基础之上的一种计算范式,特殊的拓扑结构和学习方式产生了多种神经网络模型。本章重点介绍前馈神经网络、自适应共振理论、神经网络集成、脉冲耦合神经网络、超限学习机等。

具有相同感受野并具有相同功能的视皮层神经元,在垂直于皮层表面的方向上呈柱状分布,只对某一种视觉特征发生反应,从而形成了该种视觉特征的基本功能单位。蓝脑工程创建了大脑皮层功能柱的基本单元模型。

神经元集群编码与解码是神经信息处理的关键问题,是揭示大脑工作机理的理论框架。它的发展能够促进人们对脑的总体功能的认识,为研究更为复杂的高级认知功能提供基本理论与方法。

思考题

1. 试述并行分布处理模型的 8 个要素。
2. 描述误差反向传播算法的步骤。
3. 请给出自适应共振理论 ART 模型的框图,并说明定向子系统、注意子系统的功能。
4. 脉冲耦合神经网络(PCNN)模型中的连接域是什么?
5. 什么是功能柱? 视皮层中有哪些功能柱?
6. 神经元集群编码与解码有哪些方法?