

第3章<< 神经计算

神经计算是建立在神经元模型和学习规则基础之上的一种计算范式,由于特殊的拓扑结构和学习方式,产生了多种人工神经网路,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、自适应的动力系统。

3.1 概述

神经计算研究非程序的、适应性的、大脑风格的人工神经网络信息处理的本质和能力^[917]。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,具有学习能力、记忆能力、计算能力以及智能处理功能,并在不同程度和层次上模仿人脑神经系统的信
息处理、存储及检索功能。它是在现代神经科学研究成果的基础上提出的,试图通过模拟大脑神经网络处理、记忆信息的方式设计一种新的机器使之具有人脑那样的信息处理能力。同时,对这种人工神经网络的研究将进一步加深对思维及智能的认识。为了模拟大脑信息处理的机理,人工神经网络具有非线性、非局域性、非定常性、非凸性等特点。人工神经网络是算法和结构统一为
一体的系统,可以看作是硬件和软件的混合体。

现代神经计算开始于麦克洛奇和皮兹的先驱工作^[483]。麦克洛奇是神经学家和解剖学家。他用 20 年的时间考虑神经系统对事件的表示问题。皮兹是数学天才,于 1942 年开始神经计算的研究。1943 年麦克洛奇和皮兹在一个神经建模小组上公布了他们的论文。该小组在芝加哥大学拉谢夫斯基(N Rashevsky)领导下,五年前就很活跃了。

在他们的经典论文里,麦克洛奇和皮兹结合了神经生理学和数理逻辑的研究描述了一个神经网络的逻辑分析。他们的神经元模型假定遵循有-无模型律。如果如此简单的神经元数目足够多,适当设置连接权值并且同步操作,麦克洛奇和皮兹证明这样构成的网络原则上可以计算任何可计算函数。这是一个有重大意义的结果,有了它标志着神经网络和人工智能的诞生。

麦克洛奇和皮兹 1943 年的论文影响了冯·诺依曼,使得他在 EDVAC(Electronic Discrete Variable Automatic Computer,电子离散变元自动计算机)中使用了从麦克洛奇和皮兹的神经元导出的理想化的开关延迟要素,进而改进成为 EDVIC(Electronic Numerical Integrator and Computer,电子数值积分机和计算机)。EDVIC 是第一台通用的电子计算机,于 1943 年到 1946 年在宾夕法尼亚大学电子工程摩尔学院建成。麦克洛奇和皮兹的形式神经网络理论在冯·诺依曼 1949 年在伊利诺伊大学所作的四个报告中的第二个报告中成了显著特色。

1948 年维纳的名著《控制论》(Cybernetics)发表,为控制、通信和统计信号处理描述了

一些重要概念^[810]。1961年该书第二版出版发行,添加了关于学习和自组织的新材料。从第二版的第二章中可以看出,维纳在主体背景下抓住了统计机制的物理意义,但是联合统计机制和学习系统的丰硕成果却留给了霍普菲尔德。

第二个重要发展是1949年,赫布在他的书《行为组织学》(*The Organization of Behavior*)^[311]中第一次清楚说明了突触修正的生理学习规则。特别是赫布提出大脑的连接是随着生物学会不同功能任务而连续地变化的,神经组织就是由这种变化创建起来的。赫布继承了拉莫尼(Ramony)和卡贾尔早期的假设并引入自己的学习假说:两个神经元之间的可变突触被突触两端神经元的重复激活加强了。赫布的书在心理学家中有广泛的影响,但遗憾的是对工程界却影响很少。

赫布的书是学习和自适应系统的计算模型发展的灵感源泉。1956年罗切斯特(Rochester)、霍兰德(Holland)、哈比特(Habt)、杜伯(Duba)的论文也许是用计算机模拟测试公式化的赫布学习假说的神经理论的第一次尝试,论文报告的模拟结果表明必须加上抑制才能实际工作。同一年,乌特列(Uttley)演示了带有可修改的突触的神经网络可以学习分类简单的二值模式集。乌特列引入了所谓漏电求积和点火神经元(Leaky Integrate and Fire Neuron),后来凯恩尼罗(Caianiello)对它进行了形式分析。在更晚一些的工作中,乌特列假设了神经系统可变突触的效果依赖于突触两端波动状态的统计关系,因此和香农的理论有了联系。

在1952年,阿希比(Ashby)的书《大脑设计:自适应行为的起源》出版^[34],这本书引人注目的是其基本观点,即自适应行为不是与生俱来而是后天习得,通过学习动物(系统)的行为变得更好。这本书强调活的生物如同机器的动态观点和有关稳定性的概念。

1954年明斯基在普林斯顿大学写了关于神经网络的博士论文,题目是“类神经增强式系统的理论及其在大脑建模中的应用”。1961年明斯基出版了早期关于AI的优秀论文“人工智能进展”。后面这篇文章包括了关于现在称为神经网络内容的很大一节。1967年明斯基出版了《计算:有限和无限机器》这本书,第一次以书的形式扩展了麦克洛奇和皮兹的结果,并把它们放在自动理论和计算理论的背景中。

1954年,作为通信理论先驱者之一和全息照相技术的发明者,盖博尔(Gabor)提出了非线性自适应滤波器的思想,并且他与合作者一起建立了这样的机器。随机过程产生的样本以及希望机器产生的目标函数一起提供给机器完成在线学习。

20世纪50年代,泰勒(Taylor)开始研究联想记忆。接着施泰恩布什(Steinbuch)介绍了学习矩阵,这个矩阵由插在成行的“感觉”接受器和“马达”效应器之间的开关平面网络构成。在1969年,威尔肖(Willshaw)、布尼曼(Buneman)和洛埃特-希克斯(Longuet-Higgins)发表了关于非全息照相的联想记忆的优秀论文。这篇文章给出了两类网络模型:实现相关矩阵的简单光学系统和由光学记忆提出的与之相关的神经网络。联想记忆早期发展的重要贡献包括安德森(Anderson)、科霍南(Kohonen)和纳卡诺(Nakano)于1972年的文章,他们在外积学习规则的基础上独立地引入相关矩阵记忆的思想。

冯·诺依曼是20世纪前50年的科学巨匠。为了纪念他,人们把设计数字计算机的基础命名为冯·诺依曼结构。耶鲁大学邀请他在1956年作Silliman报告。他于1957年去世,稍后他未完成的Silliman报告于1958年出版成书:《计算机和大脑》(*The Computer and the Brain*)。这本书很有意义,书中显示出冯·诺依曼开始意识到大脑和计算机的巨大

差异。

神经网络中一个值得特别关心的问题是利用被认为不可靠的神经元分量构建可靠的神经网络。1956年冯·诺依曼利用冗余的思想解决了这个重要问题,基于这种思想维诺格拉德(Winograd)和考恩(Cowan)于1963年建议在神经网络中使用分布冗余表示,这表明了大量的元素怎样集体表示具有鲁棒性和并行性的单个概念。

在麦克洛奇和皮兹的经典论文发表15年以后,罗森勃拉特在他有关感知器的研究著作中提出了模式识别问题的新方法,一种新的监督学习方法^[632]。感知器收敛定理使罗森勃拉特的工作取得圆满的成功。他于1960年提出感知器收敛定理的第一个证明。该定理的证明也出现在Novikoff(1963)和其他人的工作中。威德罗(Widrow)和霍夫(Hoff)介绍了最小平均平方(Least Mean Square,LMS)算法,并用它构成了Adaline(Adaptive Linear Element)。感知器和Adaline的区别在于训练过程。最早的可训练的具有多个自适应元素的分层神经网络之一是由威德罗和他的学生提出的Madaline(Multiple-Adaline)结构^[803]。1967年甘利俊一(Amari)将统计梯度方法用于模式分类。1965年Nilsson出版《学习机》(Learning Machine)一书^[540],迄今为止它仍是关于用超平面区分线性可分模式的最好的作品。在20世纪60年代,感知器神经网络好像可以做任何事。明斯基和佩珀特(Papert)在1969年合写的《感知器》(Perceptrons)一书中利用数学理论证明单层感知器所能计算的本质局限^[508]。在有关多层次感知器的简要介绍一节中,他们认为没有任何理由假定单层感知器的任何局限可以在多层的情况下被克服。在多层次感知器的设计中面临一个重要的问题是置信度问题(即隐藏神经元在网络中的置信度问题)。明斯基于1961年在他的“加强学习系统的置信度问题”中首次使用了“置信度”术语。在之后的60年代里,解决感知器的置信度问题所必需的大多数思想和基本概念已经公式化了,也就是现在称之为霍普菲尔德网络的递归(吸引子)网络所固有的许多思想。然而,直到20世纪80年代这些基本问题求解结果才出现。

在20世纪70年代这些因素以这种或那种方式阻碍了人们进一步研究神经网络。除了一些心理学和神经学方面的专家之外,许多研究人员在那个时期都改变了研究领域,只有屈指可数的早期开创者继续神经网络研究。从工程学的角度,可将20世纪70年代追溯为神经网络的潜伏期。

在20世纪70年代一个重要的成果就是出现了利用竞争学习的自组织理论。马尔斯伯格(von der Malsburg)1973年完成的计算机模拟工作也许是第一次演示了自组织。在人脑中拓扑映射启发下,1976年威尔肖和马尔斯伯格发表了第一篇关于自组织映射形成的论文。

在20世纪80年代关于神经网络的理论和设计方面取得了进展,随之神经网络的研究工作进入了恢复期。格罗斯伯格(S Grossberg)基于他的竞争学习理论的早期工作^[292,293,294],建立了一个新的自组织原则,就是著名的自适应共振理论(Adaptive Resonance Theory,ART)。基本上说,这个理论包括一个由底向上的识别层和一个由顶向下的产生层。如果输入形式和已学习的反馈形式匹配,一个叫做自适应共振的不定状态(即神经活动的放大和延长)发生了。这个前向或反向映射原已由其他的研究者在不同的条件下重新发现。

在1982年,霍普菲尔德基于能量函数的思想提出了一种对称连接的递归网络计算的新

方法。并且他在这种递归网络和统计物理中使用的 Ising 模型之间建立了同构。这个类推为一连串的物理理论进入神经元模型奠定了基础,因此改变了神经网络的研究方法。这类具有反馈的特殊神经网络在 80 年代引起了大量的关注,在那个时期产生了著名的 Hopfield 网络。尽管 Hopfield 网络不可能是真正的神经生物系统模型,然而它们包涵的原理,即在动态的稳定网络中存储信息的原理,是极深刻的。事实上,这个原理可以追溯到许多其他研究者的早期工作。

格雷盖(Gragg)和汤姆泼勒(Tamperley)分别于 1954 年和 1955 年观察得出正是由于神经元能被“点火”(激活)或“不点火”(静止),所以在一个网格中的原子可以用它们自旋指向“上”或“下”。

1967 年考恩引入了“sigmoid”的激活特征和一个神经元基于 Logist 函数的平滑激活条件。

格罗斯伯格于 1967—1968 年引入了一个神经元的相加模型,涉及非线性差分/微分方程,并且探测了以短期记忆为基础的模型用途。

1972 年甘利俊一独立地引入了神经元的相加模型,并用它研究随机连接的类神经元的元素的动态行为。

威尔森(Wilson)和考恩于 1972 年推导了包括兴奋和抑制模型神经元的空间局部化的群体动力学耦合非线性微分方程。

利特勒(Little)和肖于 1975 年描述了神经元激活或不激活概率模型,并用它发展了短期记忆理论。

安德森、西弗施泰(J W Siverstein)、里兹和琼斯(R S Jones)于 1977 年提出盒中脑(Brain-State-in-Box, BSB)模型,由一个非线性动力学耦合的简单联想网络组成。

因此,毫不奇怪 1982 年霍普菲尔德的论文发表后引起了很大争论^[330]。但是,该文第一次使在动态的稳定网络中存储信息的原理清楚了。霍普菲尔德表明了他对从统计力学自旋玻璃体检验具有对称连接的特殊递归网络富有洞察力,对称性设计可以保证收敛到一个稳定的条件。1983 年,科恩(M A Cohen)和格罗斯伯格建立了按内容寻址记忆的一般原则连续 Hopfield 网络是一个特例。吸引子神经网络的一个与众不同的特征是一个学习的必要量——时间出现在网络的非线性动力学中。在这个背景下 Cohen-Grossberg 的定理非常重要。

1982 年另一个重大发展是科霍南关于使用一维或二维格形结构的自组织映射研究的著作^[377],这在某些方面与威尔肖和马尔斯伯格稍早的工作不同。在文献中科霍南工作在分析和应用方面比威尔肖和马尔斯伯格的模型得到了更多的注意,已经成为这一领域其他创新的评估标准。

1983 年克尔潘特里克(S Kirkpatrick)、格拉特(C D Gelatt)和维奇(M P Vecchi)描述了解决组合最优化的问题的称为模拟退火的新方法^[369]。模拟退火根植于统计力学,是基于梅罗波列斯(N Metropolis)等在计算机仿真中第一次使用的一个简单技术^[496]。阿克列(D H Ackley),辛顿(G E Hinton)和塞杰诺斯基(T J Sejnowski)利用模拟退火的思想发展称为玻尔兹曼(Boltzmann)机的随机机器^[321],它是多层神经网络的第一个成功实现。虽然玻尔兹曼机的学习算法没有反向传播算法的计算效率高,但它证明了明斯基和佩珀特的设计是不成立的,也为尼勒(R M Neal)随后的 sigmoid 信度网络发展做了铺垫工作^[518]。

sigmoid 信度网络完成了两件事：(1) 学习显著改善；(2) 联系了神经网络和信度网络。sigmoid 信度网络学习性能的进一步提高是沙勒(L K Saul)、雅卡拉(T Jakkolla)和约丹(M I Jordan)利用一个根植于统计力学的平均场理论作出的^[628]。

巴托(A G Barto)、苏顿(R S Sutton)和安德森关于强化学习的论文发表于 1983 年^[60]。虽然他们不是第一次使用强化学习(例如明斯基在他 1954 年的博士论文中考虑过它)，但这篇文章引起了大量关于强化学习及其在控制中应用的兴趣。特别地，他们证明了一个强化学习系统可以在没有有益教师的情况下学习倒立摆(即车上立一个杆)平衡。学习系统仅要求当杆对竖直方向倾斜超过一定角度或车到达轨道的端点时的失败信号。1996 年贝特色卡斯(D P Bertsekas)和茨茨克利斯(J N Tsitsiklis)出版了《神经动力学程序》(*Neurodynamics*)，这本书把强化学习和 Bellman 的动态规划相联系，把它放在一个恰当的数学基础上。

1984 年出版了布雷滕伯格(V Braintenberg)的书《工具：合成心理学的实验》(*Vehicles : Experiments in Synthetic Psychology*)。在这本书中布雷滕伯格提出了目标导向的自组织行为原则：利用公认的基本机制的一个综合体而非由顶向下的分析是最好了解一个复杂过程的方法。在科幻小说的形式下，布雷滕伯格通过描述各种具有简单内部结构的机器说明了这个重要原则。他直接或间接研究了 20 多年的动物大脑，这启发了他描述这些机器的特性和它们的行为。

1986 年鲁梅尔哈特(D E Rumelhart)、辛顿和威廉姆斯(R J Williams)报告了反向传播算法的发展。同一年，著名的鲁梅尔哈特和麦克莱伦德(J L McClelland)主编的《并行分布处理：认知微结构的探索》(*Parallel Distributed Processing : Explorations in the Microstructure of Cognition*)一书出版^[635]。这本书极大地促进了反向传播算法的使用造成重大影响，它已成为最通用的多层感知器的训练算法。事实上，反向传播学习在同一时间在其他两个地方被独立提出过。在 20 世纪 80 年代中公布反向传播算法后，发现早在 1974 年 8 月哈佛大学的韦勃斯(P J Werbos)已经在他的博士学位论文中描述了^[805]。韦勃斯的博士论文是有效的反向传播模式的梯度计算的第一次描述，它可以应用于包括神经网络作为其特例的一般网络模型。反向传播的基本思想可进一步追溯到勃雷森(A E Bryson)和何(Y C Ho)的书《应用最优控制》(*Applied Optimal Control*)。在该书标题为“多阶段系统”的 2.2 节中，描述了使用 Lagrange 形式的反向传播推导。但是，最终的分析得出反向传播算法的许多荣誉应给予鲁梅尔哈特、辛顿和威廉姆斯，因为他们提出了它在机器学习的应用并且演示了它怎样工作。

1988 年林斯克(R Linsker)描述了认知网络中自组织问题的新原理^[439]。原理设计成在突触连接和突触动态范围限制下最大限度地保持输入活动模式的信息。其他几个视觉研究者也做出了相似的建议。但是，是林斯克使用了根植于信息理论的抽象概念构成了极大互信息(Infomax)原理的公式。林斯克的文章重新激发了应用信息理论到神经网络中的兴趣。特别是贝尔(A J Bell)和塞杰诺斯基(T J Sejnowski)的信息理论对盲信源分离问题的应用^[68]，已经促使许多研究者探索用于求解大都熟知的盲解卷积的很大一类问题的信息理论模型。

同样在 1988 年，布鲁姆海德(D S Broomhead)和洛韦(D Lowe)描述了使用径向基函数(Radial Basis Function, RBF)设计多层前馈网络的过程。RBF 提供了多层感知器的另一选

择。径向基函数的基本想法至少追溯到佩希坎罗夫(O A Bashkirov)、布雷弗曼(E M Braverman)和姆希尼克(I B Muchnik)首先提出的势函数方法^[62],以及艾泽曼(M A Aizerman)、布雷弗曼和罗宗纳(L I Rozonnoer)发展的势函数理论。在1973年杜达(R O Duda)和哈特(P E Hart)的经典著作《模式分类和场景分析》(*Pattern Classification and Scene Analysis*)给出了势函数方法的一个描述。尽管如此,布鲁姆海德和洛韦的文章导致了联系神经网络设计和数值分析的中重要领域以及线性自适应滤波器的大量研究的努力。1990年波杰欧(T Poggio)和吉罗希(F Girosi)利用Tikhonov的正规化理论进一步丰富了RBF网络理论^[584]。

1989年米德(C A Mead)的《模拟VLSI和神经系统》(*Analog VLSI and Neural Systems*)一书出版。这本书提出了从神经生物学和VLSI技术吸取的不寻常的混合概念,包括米德和他的合作者写的关于硅视网膜和硅耳蜗的章节。

在20世纪90年代早期,瓦珀尼克(V N Vapnik)和他的合作者提出了计算上强有力并被称为支持向量机(Support Vector Machines, SVM)的一种监督学习网络,用于解决模式识别、回归和密度估计问题^[780,781],新方法是基于有限样本学习理论的结果。支持向量机的一个新颖的特征就是在它们的设计中以自然的方式包含了Vapnik-Chervonenkis(VC)维数。VC维数提供了神经网络从一个样本集中学习的容量的一种度量。

现在混沌是构成物理现象的关键方面已很好地建立了。许多人提出了一个问题:在神经网络研究中混沌起关键作用吗?在生物学意义上弗里曼(W J Freeman)相信这个问题的答案是肯定的^[241]。根据弗里曼神经模型,活动模式不是大脑外部所加的,而是从内部构建的。混沌动力学提供了对神经元群体内和它们之间自组织模式出现的必要条件进行描述的一个基础。

霍普菲尔德1982年的文章和鲁梅尔哈特和麦克莱伦德(J L McClelland)1986年的两卷书是对20世纪80年代神经网络的复兴最有影响的作品。神经网络从麦克洛奇和皮兹的早期岁月已走过了很长一段路,已确立了它作为根植于神经科学、心理学、数学、物理学和工程的交叉学科的地位,并将继续在理论、设计和应用上逐步深入。

人一出生,大脑就有精巧的构造和具有通过我们通常称为“经验”而建立它自己规则的能力。事实上经验是经时间积累的,大脑在出生后头两年内会发生非常显著的发展(即硬件线路),超越这个阶段后发展将继续进行。

一个“发展中”的大脑是与可塑的神经元同时的。可塑性允许一个发展中的神经系统适应它的周边环境。正如可塑性似乎是人类大脑中作为信息处理单元的神经元的功能的关键,它在人工神经元组成的神经网络中亦是如此。最普通的形式,神经网络是一个设计出来以对人脑完成特定任务或感兴趣功能的方法进行建模的机器;网络一般用电子器件实现或者用软件在数字计算机上模拟。在本章中,我们的兴趣大致锁定在一类重要的神经网络上,这种网络通过学习过程来实现有用的计算。为了获得好的结果,神经网络使用了一个很庞大的简单计算单元间的内部连接,这些简单计算单元称为“神经元”或者“处理单元”。据此我们给出将神经网络看作一种自适应机器的定义。

一个神经网络是一个由简单处理单元构成的规模宏大的并行分布处理器。天然具有存储经验知识和使之可用的特性。神经网络从两个方面模拟大脑。

(1) 神经网络获取的知识是从外界环境中学习得来的。

(2) 内部神经元的连接强度,即突触权值,用于储存获取的知识。

用于完成学习过程的程序称为学习算法,其功能是以有序的方式改变系统权值以获得想要的设计目标。

突触权值修改提供神经网络设计的传统方法。这种方法和线性自适应滤波器理论很接近,滤波器理论已经很好地建立起来并成功应用在很多不同领域。但是神经网络修改它的拓扑结构亦是可能的,这也和人的神经元会死亡和新的突触连接会建立等情况相适应。

神经网络在文献中也称为神经计算机、连接主义网络、并行分布处理器等。很明显神经网络的计算能力来自以下几点:①大规模并行分布式结构。②神经网络学习能力从而具有泛化能力。泛化是指神经网络对不在训练(学习)集中的数据可以产生合理的输出。这两个信息处理能力让神经网络可以解决一些复杂(大规模)的现在还不能处理的问题。但是实践中,神经网络不能单独作出解答,它们需要被整合在一个一致的系统工程方法中。具体地,一个感兴趣的复杂问题被分解成几个相对简单的子问题,神经网络分配与它的固有能力相符的子问题。但是,我们在建立一个可以模拟人脑的计算机结构(如果可能)之前还有很长路要走,认识这一点是很重要的。

神经网络的应用提供了下列有用的性质和能力。

(1) 非线性。一个人工神经元可以是线性或者是非线性的。一个由非线性神经元互联而成的神经网络自身是非线性的,并且非线性是一种分布于整个网络中的特殊性质。非线性是一个很重要的性质,特别地,如果产生输入信号(如语音信号)内部的物理机制是天生非线性的。

(2) 输入输出映射。监督学习或有教师学习是一个学习的流行的变化例,涉及使用带标号的训练样本或任务例子对神经网络的突触权值进行修改。每个样本由一个唯一的输入信号和相应期望反应组成。从一个训练集中随机选取一个样本给网络,网络就调整它的突触权值(自由参数),以最小化期望反应和由输入信号以适当的统计准则产生的实际反应之间的差别。使用训练集中的很多样本重复神经网络的训练,直到网络到达没有大的突触权值修正的稳定状态为止。先前用过的样本可能还要在训练期间以不同顺序重复使用。因此对当前问题网络通过建立输入输出映射从样本中进行学习。这样一个方法使人想起了无参数统计推断的研究,它是非模型估计的统计处理的一个分支,或者从生物学角度看,称为tabula rasa 学习。这里使用“非参数”表示的一个事实是,没有对输入数据统计模型作任何先验假设。比如,考虑一个模式分类任务,其要求是把代表具体物体或事件的输入信号分类到几个预先分好的类中去。在这个问题的非参数方法中,要求利用例子集“估计”输入信号空间中模式分类任务的任意判决边界,并且不使用概率分布模型。监督学习隐含了一个类似的观点,神经网络的输入输出之间的映射和非参数统计推断相类似。

(3) 适应性。神经网络具有调整自身连接权值以适应外界变化的能力。特别是在特定环境下可以运行的神经网络对环境的小变化可以很容易重新训练。更进一步,当它在一个时变环境(即它的统计特性随时间变化)中运行时,网络突触权值就可以设计成随时间变化。用于模式识别、信号处理、控制的神经网络耦合它的自适应能力就可以变成能进行自适应模式识别、自适应信号处理、自适应控制的有效工具。作为一个普通规则,在保证系统保持稳定时一个系统的自适应性越好,当要求在一个时变环境下运行时它就越具鲁棒性。但是,需

要强调的是,适应性不一定导致鲁棒性,实际可能相反。例如,一个暂态适应系统可能变化过快,以至于对尖峰干扰有反应,这将引起系统性能的急剧恶化。为最大限度实现适应性,系统的主要稳定时间长度应该长到可以忽略尖峰干扰,而短到可以反应环境的大体变化。这是一个稳定性和可塑性之间的二难问题^[292]。

(4) 证据响应。在模式识别的问题中,神经网络可以设计成既提供选择哪一个特定模式的信息,也提供做出判决的置信度的信息。后者可以用来拒判那些出现的过于模糊的模式。有这些信息,网络的输出结果就会有改善。

(5) 前后关系的信息。神经网络的特定结构和活动状态代表知识。网络中每一个神经元都潜在地受到网络中所有其他神经元全局活动的影响。必然地,前后关系信息自然由一个神经网络处理。

(6) 容错性。一个以硬件形式实现的神经网络有天生容错的潜质,或者鲁棒计算的能力,意即它的性能在不利运行条件下逐渐下降。比如,一个神经元或一个连接损坏了,已存的模式的回忆在质量上被削弱。但是,由于网络信息存储的分布特性,在网络的总体响应严重劣化之前这种损坏是分散的。因此,原则上,一个神经网络的性能显示了一个缓慢变差的过程而不是灾难性的失败。有一些关于鲁棒性计算的经验证据,但通常它是不可控的。为了确保网络事实上的容错性,有必要在设计训练网络的算法时采用正确的度量^[365]。

(7) VLSI 实现。神经网络的大规模并行性使它具有快速处理某些任务的潜在能力。这一特性使得神经网络很适合用超大规模集成电路(VLSI)实现。VLSI 的一个特殊优点是提供一种以高度分层的方式匹配真实复杂性行为的方法。

(8) 分析和设计的一致性。基本上,神经网络作为信息处理器具有通用性。我们这样说是在这样的意义下,即涉及神经网络的应用的所有领域都使用同样记号。这种特征体现在以下 3 个方面:①神经元不管形式如何,在所有的神经网络中都代表一个相同成分;②这种共性使得在不同应用中的神经网络共享相同的理论和算法成为可能;③模块化网络可以用模块的无缝整合来实现。

(9) 神经生物类比。神经网络的设计是由对大脑的类比引发的,而大脑是一个容错的并行处理的活生生的例子;神经网络的设计不仅在物理上是可实现的而且是快速高效的。神经生物学家将(人工)神经网络看作是一个解释神经生物现象的研究工具。另外,工程师关注神经生物学是将其作为解决复杂问题的新思路,这些问题比基于常规的硬件线路设计技术所能解决的问题更复杂。下面两个例子说明了这两种观点。

在文献[22]中比较了前庭视觉的线性系统模型。前庭视觉反射作用(Vestibulo-Ocular Reflex,VOR)是眼球运动系统的一部分,其作用是让眼球向与头转动方向相反的方向运动,以维持视觉(视网膜)图像的稳定性。VOR 由前庭核酸的前端神经元调节,前端神经元从前庭感知神经元中接受头部旋转信息并处理,将结果告知眼球肌肉的动作神经元。输入(头部旋转信息)和输出(眼球旋转)可以精确确定,因此 VOR 很适合用来建模。另外,它是比较简单的反射作用,并且其组成神经元的神经生理学的内容已经被很好阐述。在这三种神经模型中,前端神经元(反射内层神经元)在前庭核酸中是最复杂也是最引人注意的。VOR 已经用集块线性系统描述器和控制理论模型化了。这些模型对解释 VOR 的整体性质有一些作用,但是对其组成神经元特性的了解却用处不大。这种情况通过神经网络的模型已经被大大改善了。VOR 的递归网络模型能重现和解释调节 VOR 的神经元的处理信号时的静态、动态、非线性和分布性许多方面的特性,特别是前庭核酸神经元。

视网膜是我们开始将外部环境的物理图像投射到一行接受器上形成的视觉表示和第一个神经图像结合的地方。它是眼球后部的神经组织薄层,其功能是将光学图像转换成神经图像并沿光神经传输给大量的视觉中枢以便进一步处理。这是一个复杂的工作,可以从视网膜的复杂组织得到证明。在脊椎动物的视网膜中光图像转化成神经图像的过程由以下三个阶段组成^[739]。

- (1) 感应神经元层的图像传导。
- (2) 将结果信号(对光刺激的反应产生)以化学性突触的形式传输给一层二极性单元。
- (3) 同样,以化学性突触的形式把结果信号传给称为神经节的输出神经元。

在两个感应阶段(即从接受器到二极单元和从二极单元到神经节单元),有专门侧向连接的分别称为水平连接神经元和无长突细胞的神经元。这些神经元的工作是修改感应层之间的传输。另外有被叫做中间网状细胞的分配元素,它们的工作是将信号从内部感应层传到外部感应层。一些研究人员已经建立了模拟视网膜结构的电子芯片^[86],这些电子芯片称为神经形态集成电路。一个神经形态的图像传感器由成排的感光器构成,由每个元素(像素)模拟回路结合而成。它能模拟视网膜适应局部的亮度变化、检测边缘和检测运动。神经生物学模拟,例如神经形态集成电路,有另一个重要的应用:它提供了一种希望和信念,并在一定程度上提供了一种存在性证明,即对神经生物学结构的物理上的了解对电子学工艺和超大规模集成电路技术有创造性的影响。

迄今为止,按生物神经网络大规模并行分布方式构造的各种人工神经网络,虽然已经在信息处理中扮演着越来越重要的角色,但是并没有显示出人们所期望的聪明智慧来。对以仿效大脑神经系统为目的的人工神经网络的研究历程进行一些分析和反思,探讨下一步可能采取的方法步骤,对今后智能信息科学的进一步发展将是有益的。要让人工神经网络更好更快地向人脑神经网络学习,就有必要对今后人工神经网络的主攻方向、研究路线、方法步骤、关键技术和应当采取的措施等方面,作一些考虑和调整。

概括起来,为了使人工神经网络向人脑神经网络学习得更快更好,可以从以下3个方面着手。

(1) 明确人工神经网络的主要智能优势是擅长非精确性信息处理。按照输入和输出特性的不同,可将智能系统最经常处理的信息类型归纳为四类。其中,数值计算和逻辑推理所对应的信息处理任务,现行的冯·诺依曼数字计算机有着成熟而巨大的能力,人工神经网络既无必要也无优势在此等领域中与之竞争。其他三类都在输入信息或输出信息中包含有非精确信息处理的内容,因而它们是人工神经网络能发挥作用的领域。人类的绝大部分脑力劳动正是投入在这3类信息处理上。

(2) 探索新的人工神经网络体系结构。神经解剖学方面的研究表明,在人脑神经网络中除了有由神经细胞体-轴突-突触-树突构成的神经电位脉冲电路系统外,还有一个在前后突触间约 $0.02\mu s$ 的间隙中释放化学性神经递质(谷氨酸)的调节系统,这个化学递质系统的作用相当于半导体的栅极电路,起着至关重要的调控放大作用。因此,如何建立神经电脉冲系统和化学递质系统合成的耦合系统数学模型,分析其工作机理并提出可行简便算法,是一条从建模上使人工神经网络更靠近人脑神经网络的值得重视的途径。

(3) 寻求新的网络拓扑结构和相应的学习算法。迄今为止,以和积式神经元为基础的前馈型多层(特别是三层)神经网络,与误差反向传播(Back-Propagation, BP)学习算法相配合的人工神经网络结构体制,获得了最广泛的应用。但是,它在性能上仍存在若干待改进的

地方。例如,它只能调整权值,不能调整网络拓扑结构,无法实现注意力集中功能;学习新样本时,会“冲乱”原已学好保存下来的旧样本;其学习算法中,包含有较复杂的非线性激活函数的求导运算等。

3.2 神经元模型

从神经元的特性和功能可以知道,神经元是一个多输入单输出的信息处理单元,而且,它对信息的处理是非线性的。根据神经元的特性和功能,可以把神经元抽象为一个简单的数学模型。工程上用的人工神经元模型如图 3.1 所示。

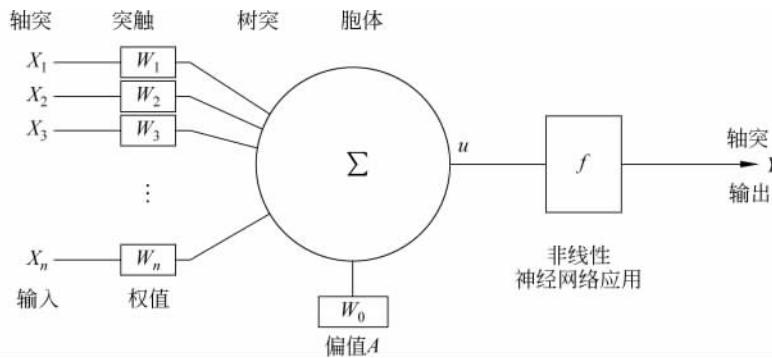


图 3.1 神经元的数学模型

在图 3.1 中, X_1, X_2, \dots, X_n 是神经元的输入,即是来自前级 n 个神经元的轴突的信息。 A 是神经元 i 的阈值; W_1, W_2, \dots, W_n 分别是神经元对 X_1, X_2, \dots, X_n 的权系数,也即突触的传递效率; Y 是神经元的输出; f 是激活函数,它决定神经元受到输入 X_1, X_2, \dots, X_n 的共同刺激达到阈值时以何种方式输出。

从图 3.1 的神经元模型,可以得到神经元的数学模型表达式:

$$f(u_i) = \begin{cases} 1 & u_i > 0 \\ 0 & u_i \leq 0 \end{cases}$$

对于激活函数 f 有多种形式,其中最常见的有阶跃型、线性型和 S 形三种形式,这三种形式如图 3.2 所示。

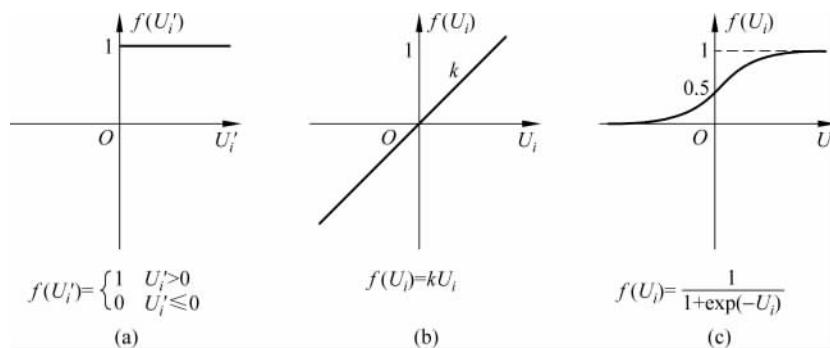


图 3.2 典型激活函数

上面所叙述的是应用最广泛而且人们最熟悉的神经元数学模型,也是历史最悠久的神经元模型。近年来,随着神经网络理论的发展,出现了不少新颖的神经元数学模型,包括逻辑神经元模型、模糊神经元模型等,也渐渐受到人们的关注和重视。

3.3 反传学习算法

近年来人工神经网络在很多领域得到了应用,其中大部分采用前馈网络(Feedforward)和反传算法(Backpropagation: BP)。BP算法是为了解决多层前馈神经网络的权系数优化而提出来的,所以,BP算法也通常暗示着神经网络的拓扑结构是一种无反馈的多层前馈网络。故而,有时也称无反馈多层前馈网络为BP模型。

3.3.1 反传算法的原理

反传算法是用于前馈多层网络的学习算法,前馈多层网络的结构一般如图3.3所示。它含有输入层、输出层以及处于输入输出层之间的隐层。在隐层中的神经元也称隐藏单元。隐层虽然和外界不连接,但是,它们的状态则影响输入输出之间的关系。这也就是说,改变隐层的权系数,可以改变整个多层神经网络的性能。有实验表示,增加隐层的层数和隐藏单元的个数不一定能够提高网络精度和表达能力。所以BP网一般都选用二级网络。

反传算法分两步进行,即正向传播和反向传播。这两个过程的工作简述如下。

(1) 正向传播

输入的样本从输入层经过隐藏单元一层一层进行处理,通过所有的隐层之后,则传向输出层;在逐层处理的过程中,每一层神经元的状态只对下一层神经元的状态产生影响。在输出层把当前输出和期望输出进行比较,如果当前输出不等于期望输出,则进入反向传播过程。

(2) 反向传播

反向传播时,把误差信号按原来正向传播的通路反向传回,并对每个隐层的各个神经元的权系数进行修改,以期望误差信号趋向最小。

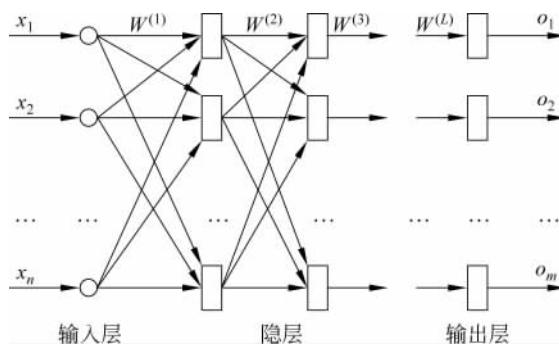


图3.3 前馈多层网络的结构

设有一个 m 层的神经网络,并在输入层加有样本 X ;设第 k 层的 i 神经元的输入总和表示为 U_i^k ,输出 X_i^k ;从第 $k-1$ 层的第 j 个神经元到第 k 层的第 i 个神经元的权系数为 W_{ij} ,各个神经元的激发函数为 f ,则各个变量的关系可用下面有关数学式表示:

$$X_i^k = f(U_i^k) \quad (3.1)$$

$$U_i^k = \sum_j W_{ij} X_j^{k-1} \quad (3.2)$$

3.3.2 反传算法的数学表达

反传算法实质是求取误差函数的最小值问题。这种算法采用非线性规划中的最速下降方法,按误差函数的负梯度方向修改权系数。为了说明反传算法,首先定义误差函数 e 。取期望输出和实际输出之差的平方和为误差函数,则有

$$e = \frac{1}{2} \sum_i (X_i^m - Y_i)^2 \quad (3.3)$$

其中: Y_i 是输出单元的期望值,它也在这里用作教师信号; X_i^m 是实际输出;因为第 m 层是输出层。

由于反传算法按误差函数 e 的负梯度方向修改权系数,故权系数 W_{ij} 的修改量 ΔW_{ij}

$$\Delta W_{ij} \propto -\frac{\partial e}{\partial W_{ij}}$$

也可写成

$$\Delta W_{ij} \propto -\eta \frac{\partial e}{\partial W_{ij}}$$

其中: η 为学习速率,即步长,一般取 $0 \sim 1$ 的数。

很明显,根据 BP 算法原则,求 $\frac{\partial e}{\partial W_{ij}}$ 最关键的。下面求 $\frac{\partial e}{\partial W_{ij}}$ 有:

$$\frac{\partial e}{\partial W_{ij}} = \frac{\partial e_k}{\partial U_i^k} \cdot \frac{\partial U_i^k}{\partial W_{ij}} \quad (3.4)$$

由于 $\frac{\partial U_i^k}{\partial W_{ij}} = \frac{\partial \left(\sum_l W_{il} X_l^{k-l} \right)}{\partial W_{ij}} = X_j^{k-l} |_{l=j}$ 故

$$\frac{\partial e}{\partial W_{ij}} = \frac{\partial e_k}{\partial U_i^k} \cdot X_j^{k-l} \quad (3.5)$$

从而有

$$\Delta W_{ij} = -\eta \frac{\partial e}{\partial W_{ij}} = -\eta \frac{\partial e_k}{\partial U_i^k} \cdot X_j^{k-l} \quad (3.6)$$

令

$$d_i^k = \frac{\partial U_i^k}{\partial W_{ij}} \quad (3.7)$$

则有学习公式:

$$\Delta W_{ij} = -\eta d_i^k \cdot X_j^{k-l} \quad (3.8)$$

其中: η 为学习速率,即步长,一般取 $0 \sim 1$ 的数。

上面实际仍未给出求 d_i^k 明显的算法公式,求 d_i^k 的计算公式如下:

$$d_i^k = \frac{\partial U_i^k}{\partial W_{ij}} = \frac{\partial e}{\partial X_i^k} \cdot \frac{\partial X_i^k}{\partial U_i^k} \quad (3.9)$$

由于

$$\frac{\partial X_i^k}{\partial U_i^k} = f'(U_i^k)$$

为了方便进行求导,取 f 为连续函数。一般取非线性连续函数,例如 Sigmoid 函数。当取 f 为非对称 Sigmoid 函数时,有

$$f(U_i^k) = \frac{1}{1 + \exp(-U_i^k)} \quad (3.10)$$

则有

$$f'(U_i^k) = f'(U_i^k)(1 - f(U_i^k)) = X_i^k(1 - X_i^k)$$

再考虑式(3.9)中的偏微分项 $\frac{\partial e}{\partial X_i^k}$,有以下两种情况:

(1) 如果 $k=m$,则是输出层,这时有 Y_i 是输出期望值,它是常数。从式(3.3)有

$$\frac{\partial e}{\partial X_i^k} = \frac{\partial e}{\partial X_i^m} = (X_i^m - Y_i) \quad (3.11)$$

从而有

$$d_i^m = X_i^m(1 - X_i^m)(X_i^m - Y_i) \quad (3.12)$$

(2) 如果 $k < m$,则该层是隐层。这时应考虑上一层对它的作用,故有

$$\frac{\partial e}{\partial X_i^k} = \sum_l \frac{\partial e}{\partial U_l^{k+1}} \cdot \frac{\partial U_l^{k+1}}{\partial X_i^k} \quad (3.13)$$

由式(3.7)可知

$$\frac{\partial e}{\partial U_l^{k+1}} = d_l^{k+1} \quad (3.14)$$

由式(3.2)可知

$$\frac{\partial U_i^k}{\partial X_i^k} = \frac{\partial \left(\sum_j W_{ij} X_j^{k-1} \right)}{\partial X_i^k} = W_{ii} \mid_{j=i} \quad (3.15)$$

故有

$$\frac{\partial e}{\partial X_i^k} = \sum_l W_{li} \cdot d_l^{k+1} \quad (3.16)$$

最后有

$$d_i^k = X_i^k(1 - X_i^k) \cdot \sum_l W_{li} \cdot d_l^{k+1} \quad (3.17)$$

从上述过程可知:多层网络的训练方法是把一个样本加到输入层,并根据向前传播的规则 $X_i^k = f(U_i^k)$,不断一层一层向输出层传递,最终在输出层可以得到输出 X_i^m 。

把 X_i^m 和期望输出 Y_i 进行比较。如果两者不等,则产生误差信号 e ,接着则按下面公式反向传播修改权系数:

$$\Delta W_{ij} = -\eta d_i^k \cdot X_j^{k-1} U_i^k = \sum_j W_{ij} X_j^{k-1} \quad (3.18)$$

其中, $d_i^m = X_i^m(1 - X_i^m)(X_i^m - Y_i)$, $d_i^k = X_i^k(1 - X_i^k) \cdot \sum_l W_{li} \cdot d_l^{k+1}$ 。

式(3.18)中,求取本层 d_i^k 时,要用到高一层的 d_l^{k+1} ;可见,误差函数的求取是从输出层开始,到输入层的反向传播过程。在这个过程中不断进行递归求误差。

通过多个样本的反复训练,同时向误差逐渐减小的方向对权系数进行修正,最终减少误

差。从式(3.18)也可以知道,当网络的层数较多时,所需的计算量很大,导致收敛速度不快。

为了加快收敛速度,一般以上一次的权系数作为本次修正的依据之一,故而有修正公式:

$$\Delta W_{ij}(t+l) = -\eta d_i^k \cdot X_j^{k-l} + \alpha \Delta W_{ij}(t) \quad (3.19)$$

其中, η 为学习速率,即步长, η 取 $0.1 \sim 0.4$, α 为权系数修正常数,取 $0.7 \sim 0.9$ 。

上面式(3.18)也称为通用的 Delta 法则。对于没有隐层的神经网络,可取

$$\Delta W_{ij} = \eta(Y_j - X_j) \cdot X_i \quad (3.20)$$

其中, Y_j 为期望输出; X_j 为输出层的实际输出; X_i 为输入层的输入。这显然是一种十分简单的情况,式(3.20)也称为简单 Delta 法则。

在实际应用中,只有通用的 Delta 法则式(3.18)或式(3.19)才有意义。简单 Delta 法则式(3.20)只在理论推导上有用。

3.3.3 反传算法的执行步骤

在把反传算法应用于前馈多层网络,采用 Sigmoid 为激活函数时,可用下列步骤对网络的权系数 W_{ij} 进行递归求取。注意对于每层有 n 个神经元的时候,即有 $i=1, 2, \dots, n$; $j=1, 2, \dots, n$ 。对于第 k 层的第 i 个神经元,则有 n 个权系数 $W_{i1}, W_{i2}, \dots, W_{in}$,另外再取一个 W_{in+1} 用于表示阈值 θ_i ; 并且在输入样本 X 时,取 $X=(X_1, X_2, \dots, X_n, 1)$ 。

算法的执行的步骤如下。

(1) 对权系数 W_{ij} 置初值。

对各层的权系数 W_{ij} 置一个较小的非零随机数,但其中 $W_{in+1}=-\theta$ 。

(2) 输入一个样本 $X=(X_1, X_2, \dots, X_n, 1)$,以及对应期望输出 $Y=(Y_1, Y_2, \dots, Y_n)$ 。

(3) 计算各层的输出。

对于第 k 层第 i 个神经元的输出 X_i^k ,有:

$$U_i^k = \sum_{j=1}^{n+1} W_{ij} X_j^{k-1}, \quad X_{n+1}^{k-1} = 1, \quad W_{in+1} = -\theta, \quad X_i^k = f(U_i^k)$$

(4) 求各层的学习误差 d_i^k 。

对于输出层 $k=m$,有

$$d_i^m = X_i^m (1 - X_i^m) (X_i^m - Y_i)$$

对于其他各层,有

$$d_i^k = X_i^k (1 - X_i^k) \cdot \sum_l W_{il} \cdot d_l^{k+1}$$

(5) 修正权系数 W_{ij} 和阈值 θ ,用式(3.18)时有:

$$\Delta W_{ij}(t+l) = \Delta W_{ij}(t) - \eta \cdot d_i^k \cdot X_j^{k-1}$$

用式(3.19)时有:

$$\Delta W_{ij}(t+l) = \Delta W_{ij}(t) - \eta \cdot d_i^k \cdot X_j^{k-1} + \alpha \Delta W_{ij}(t)$$

其中,

$$\Delta W_{ij}(t) = -\eta \cdot d_i^k \cdot X_j^{k-1} + \alpha \Delta W_{ij}(t-l) = W_{ij}(t) - W_{ij}(t-l)$$

(6) 当求出了各层各个权系数之后,可按给定品质指标判别是否满足要求。如果满足要求,则算法结束;如果未满足要求,则返回(3)执行。

这个学习过程,对于任一给定的样本 $X_p = (X_{p1}, X_{p2}, \dots, X_{pn}, 1)$ 和期望输出 $Y_p = (Y_{p1}, Y_{p2}, \dots, Y_{pn})$ 都要执行,直到满足所有输入输出要求为止。

3.3.4 对反传网络优缺点的讨论

多层前向反传网络是目前应用最广泛的一种神经网络模型,但它也不是尽善尽美的,为了更好地理解应用神经网络进行问题求解,这里对它的优缺点展开讨论。

1. 多层前向反传网络的优点

- (1) 网络实质上实现了一个从输入到输出的映射功能,而数学理论已证明它具有实现任何复杂非线性映射的功能。因此它特别适合于求解内部机制复杂的问题。
- (2) 网络能通过学习带正确答案的实例集,自动提取“合理的”求解规则,即具有自学习能力。
- (3) 网络具有一定的泛化、概括能力。

2. 多层前向反传网络的不足

- (1) BP 算法的学习速度很慢,其原因主要有:①由于 BP 算法本质上为梯度下降法,而它所要优化的目标函数又非常复杂,因此,必然会出现“锯齿形现象”,这使得 BP 算法效率下降;②存在麻痹现象,由于优化的目标函数很复杂,它必然会在神经元输出接近 0 或 1 的情况下,出现一些平坦区,在这些区域内,权值误差改变很小,使训练过程几乎停顿;③为了使网络执行 BP 算法,不能用传统的一维搜索法求每次迭代的步长,而必须把步长的更新规则预先赋予网络,这种方法将引起算法低效。

- (2) 网络训练失败的可能性较大,其原因主要有:①从数学角度看,BP 算法为一种局部搜索的优化方法,但它要解决的问题为求解复杂非线性函数的全局极值,因此,算法很有可能陷入局部极值,使训练失败;②网络的逼近、泛化能力同学习样本的典型性密切相关,而从问题中选取典型样本实例组成训练集是一个很困难的问题。

- (3) 难以解决应用问题的实例规模和网络规模间的矛盾。这涉及网络容量的可能性与可行性的关系问题,即学习复杂性问题。

- (4) 网络结构的选择尚无一种统一而完整的理论指导,一般只能由经验选定。为此,有人称神经网络的结构选择为一种艺术。而网络的结构直接影响网络的逼近能力及泛化性质。因此,应用中如何选择合适的网络结构是一个重要的问题。

- (5) 新加入的样本要影响已学习成功的网络,而且刻画每个输入样本的特征的数目也必须相同。

- (6) 网络的预测能力(也称泛化能力、推广能力)与训练能力(也称逼近能力、学习能力)间的矛盾。一般情况下,训练能力差时,预测能力也差,并且一定程度上,随着训练能力的提高,预测能力也提高。但这种趋势有一个极限,当达到此极限时,随训练能力的提高,预测能力反而下降,即出现所谓“过拟合”现象。此时,网络学习了过多的样本细节,而不能反映样本内含的规律。

3.4 Hopfield 模型

1982年,霍普菲尔德提出了可用作联想存储器的互连网络,这个网络称为 Hopfield 网络模型,也称 Hopfield 模型。Hopfield 网络模型是一种循环神经网络,从输出到输入有反馈连接。Hopfield 网络有离散型和连续型两种。

反馈神经网络由于其输出端又反馈到其输入端,Hopfield 网络在输入的激励下,会产生不断的状态变化。当有输入之后,可以求取出 Hopfield 的输出,这个输出反馈到输入从而产生新的输出,这个反馈过程一直进行下去。如果 Hopfield 网络是一个能收敛的稳定网络,则这个反馈与迭代的计算过程所产生的变化越来越小,一旦到达了稳定平衡状态,Hopfield 网络就会输出一个稳定的恒值。对于一个 Hopfield 网络来说,关键在于确定它在稳定条件下的权系数。

应该指出:反馈网络有稳定的,也有不稳定的。对于 Hopfield 网络来说,还存在如何判别它是稳定网络还是不稳定网络,依据什么进行判别的问题。

3.4.1 离散 Hopfield 网络

Hopfield 最早提出的网络是二值神经网络,神经元的输出只取 1 和 0 这两个值,所以,也称为离散 Hopfield 神经网络。在离散 Hopfield 网络中,所采用的神经元是二值神经元;故而,所输出的离散值 1 和 0 分别表示神经元处于激活和抑制状态。

首先考虑由三个神经元组成的离散 Hopfield 神经网络,其结构如图 3.4 所示。

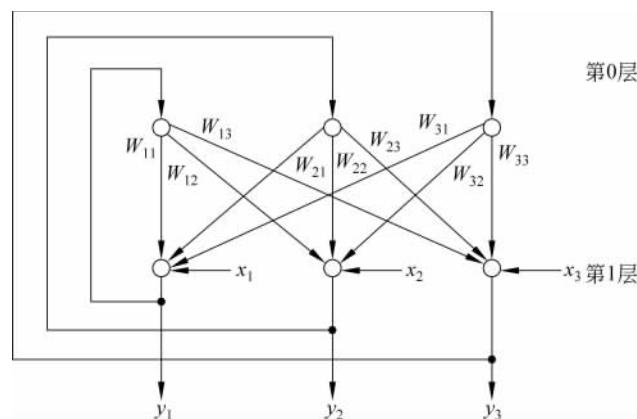


图 3.4 三神经元组成的 Hopfield 网络

在图 3.4 中,第 0 层仅仅是作为网络的输入,它不是实际神经元,所以无计算功能;而第一层是实际神经元,故而对输入信息和权系数乘积求累加和,并由非线性函数 f 处理后产生输出信息。 f 是一个简单的阈值函数,如果神经元的输出信息大于阈值 θ ,那么,神经元的输出就取值为 1;小于阈值 θ ,则神经元的输出就取值为 0。

对于二值神经元,它的计算公式如下:

$$U_j = \sum_i W_{ij} Y_i + X_j$$

其中, X_j 为外部输入。并且有:

$$Y_i = \begin{cases} 1, & U_i \geq \theta_i \\ 0, & U_i < \theta_i \end{cases}$$

对于一个离散的 Hopfield 网络,其网络状态是输出神经元信息的集合。对于一个输出层是 n 个神经元的网络,则其 t 时刻的状态为一个 n 维向量:

$$\mathbf{Y}(t) = [Y_1(t), Y_2(t), \dots, Y_n(t)]^T$$

故而,网络有 2^n 个状态;因为 $Y_j(t)$ ($j=1, \dots, n$) 可以取值为 1 或 0;故 n 维向量 $\mathbf{Y}(t)$ 有 2^n 种状态,即网络状态。

对于三个神经元的离散 Hopfield 网络,它的输出层就是 3 位二进制数;每一个 3 位二进制数就是一种网络状态,从而共有 8 个网络状态。这些网络状态如图 3.5 所示,图中立方体的每一个顶角表示一种网络状态。同样,对于 n 个神经元的输出层,它有 2^n 个网络状态,也和一个 n 维超立方体的顶角相对应。

如果 Hopfield 网络是一个稳定网络,那么在网络的输入端加入一个输入向量,则网络的状态会产生变化,也就是从超立方体的一个顶角转移向另一个顶角,并且最终稳定于一个特定的顶角。

对于一个由 n 个神经元组成的离散 Hopfield 网络,则有 $n \times n$ 权系数矩阵 \mathbf{W} :

$$\mathbf{W} = \{W_{ij}\} \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n$$

同时,有 n 维阈值向量 $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_n]^T$$

一般而言, \mathbf{W} 和 $\boldsymbol{\theta}$ 可以确定一个唯一的离散 Hopfield 网络。对于图 3.4 所示的三神经元组成的 Hopfield 网络,也可以改用图 3.6 所示的图形表示,这两个图形的意义是一样的。考虑离散 Hopfield 网络的一般节点状态;用 $Y_j(t)$ 表示第 j 个神经元,即节点 j 在时刻 t 的状态,则节点的下一个时刻($t+1$)的状态如下:

$$Y_i(t+1) = f[U_i(t)] = \begin{cases} 1, & U_i \geq 0 \\ 0, & U_i < 0 \end{cases} \quad (3.21)$$

$$U_j(t) = \sum_{i=1}^n W_{ij} Y_i(t) + X_j - \theta_j \quad (3.22)$$

当 W_{ij} 在 $i=j$ 时等于 0,则说明一个神经元的输出并不会反馈到它自己的输入,这时,离散的 Hopfield 网络称为无自反馈网络;当 W_{ij} 在 $i=j$ 时不等于 0,则说明一个神经元的输出会反馈到它自己的输入,这时,离散的 Hopfield 网络称为有自反馈的网络。

离散 Hopfield 网络有如下两种不同的工作方式。

1. 串行(异步)方式

在时刻 t 时,只有某一个神经元 j 的状态产生变化,而其他 $n-1$ 个神经元的状态不变,

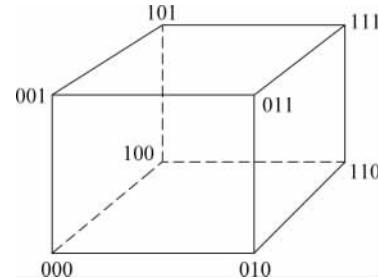


图 3.5 三神经元输出层的网络状态

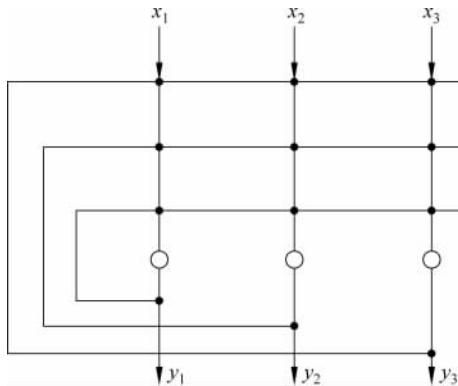


图 3.6 离散 Hopfield 网络的另外一种图示

这时称串行工作方式。并且有

$$\begin{aligned} Y_i(t+1) &= f \left[\sum_{j=1}^n W_{ij} Y_j(t) + X_j - \theta_j \right] \\ Y_i(t+1) &= Y_j(t) \quad i \neq j \end{aligned} \quad (3.23)$$

在不考虑外部输入时,则有

$$Y_i(t+1) = f \left[\sum_{j=1}^n W_{ij} Y_j(t) - \theta_j \right]$$

2. 并行(同步)方式

在任一时刻 t ,所有的神经元的状态都产生了变化,则称并行工作方式,并且有

$$Y_i(t+1) = f \left[\sum_{j=1}^n W_{ij} Y_j(t) + X_j - \theta_j \right] \quad j = 1, 2, \dots, n \quad (3.24)$$

在不考虑外部输入时,则有

$$Y_i(t+1) = f \left[\sum_{j=1}^n W_{ij} Y_j(t) - \theta_j \right] \quad j = 1, 2, \dots, n$$

对于一个网络来说,稳定性是一个重大的性能指标。对于离散 Hopfield 网络,其状态为 $Y(t)$:

$$Y(t) = [Y_1(t), Y_2(t), \dots, Y_n(t)]^\top$$

如果,对于任何 $\Delta t > 0$ 。当神经网络从 $t=0$ 开始,有初始状态 $Y(0)$; 经过有限时刻 t ,有:

$$Y(t + \Delta t) = Y(t)$$

则称网络是稳定的。

在串行方式下的稳定性称之为串行稳定性。同理,在并行方式的稳定性称之为并行稳定性。在神经网络稳定时,其状态称稳定状态。

Hopfield 网络的一个功能是可用于联想记忆,也即联想存储器。这是人类的智能特点之一。人类的所谓“触景生情”就是见到一些类同过去接触的景物,容易产生对过去情景的回味和思忆。对于 Hopfield 网络,用它作联想记忆时,首先通过一个学习训练过程确定网络中的权系数,使所记忆的信息在网络的 n 维超立方体的某一个顶角处的能量最小。当网

络的权系数确定之后,只要为网络给出输入向量,这个向量可能是局部数据,即不完全或部分不正确的数据,但是网络仍然产生所记忆的信息的完整输出。1984年 Hopfield 开发了一种用 n 维 Hopfield 网络作联想存储器的结构。在这个网络中,权系数的赋值规则为存储向量的外积存储规则(Out Product Storage Prescription),其原理如下:

设有 m 个样本存储向量 X_1, X_2, \dots, X_m , 则有

$$X_1 = \{X_{11}, X_{12}, \dots, X_{1m}\}$$

$$X_2 = \{X_{21}, X_{22}, \dots, X_{2m}\}$$

...

$$X_m = \{X_{m1}, X_{m2}, \dots, X_{mm}\}$$

把这 m 个样本向量存储入 Hopfield 网络中,则在网络中第 i, j 两个节点之间权系数的值为:

$$W_{ij} = \begin{cases} \sum X_{ik} X_{jk}, & i \neq j \\ 0, & i = j \end{cases} \quad (3.25)$$

其中: k 为样本向量 X_k 的下标, $k=1, 2, \dots, m$; i, j 分别是样本向量 X_k 的第 i, j 分量 X_i, X_j 的下标; $i, j=1, 2, \dots, n$ 。

对联想存储器的联想检索过程如下。

给定一个向量 X , 进行联想检索求取在网络中的存储内容。这时, 把向量

$$X = \{X_1, X_2, \dots, X_n\}$$

的各个分量 X_1, X_2, \dots, X_n 赋于相对应的节点 j , ($j=1, 2, \dots, n$), 则节点有相应的初始状态 $Y_j(0)$, 则有

$$Y_j(0) = X_j, \quad j = 1, 2, \dots, n$$

接下来, 在 Hopfield 网络中按动力学系统原则进行计算, 得

$$Y_j(t+1) = f[\sum W_{ij} Y_i(t) - \theta_j], \quad i, j = 1, 2, \dots, n$$

其中, f 是非线性函数, 可取阶跃函数。

状态通过不断变化, 最终会稳定下来。最终的状态是和给定向量 X 最接近的样本向量。所以, Hopfield 网络的最终输出也就是给定向量联想检索结果。这个过程说明, 即使给定向量并不完全或部分不正确, 也能找到正确的结果。在本质上, 它也有滤波功能。

3.4.2 连续 Hopfield 网络

连续 Hopfield 网络的拓扑结构和离散 Hopfield 网络相同。这种拓扑结构和生物的神经系统中大量存在的神经反馈回路是一致的。连续 Hopfield 网络和离散 Hopfield 网络一样, 其稳定条件也是 $W_{ij} = W_{ji}$ 。连续 Hopfield 网络和离散 Hopfield 网络不同点在于其函数 g 不是阶跃函数, 而是 S 形的连续函数。一般取 $g(u) = 1/(1 + e^{-u})$ 。

连续 Hopfield 网络在时间上是连续的, 所以, 网络中各神经元是以同步方式工作的。对于一个神经细胞, 即神经元 j , 其内部膜电位状态用 u_j 表示, 细胞膜输入电容为 C_j , 细胞膜的传递电阻为 R_j , 输出电压为 V_j , 外部输入电流用 I_j 表示, 则连续 Hopfield 网络可用图 3.7 所示的电路表示。

$$\begin{cases} C_j \frac{dU_i(t)}{dt} = \sum_{j=1}^n W_{ij} V_j(t) - \frac{U_j(t)}{R_j} + I_j \\ V_j(t) = g_j(U_j(t)) \quad j = 1, 2, \dots, n \end{cases} \quad (3.26)$$

其中, n 是神经网络神经元的个数, $v_j(t)$ 为输出电位, $U_j(t)$ 为输入电位。

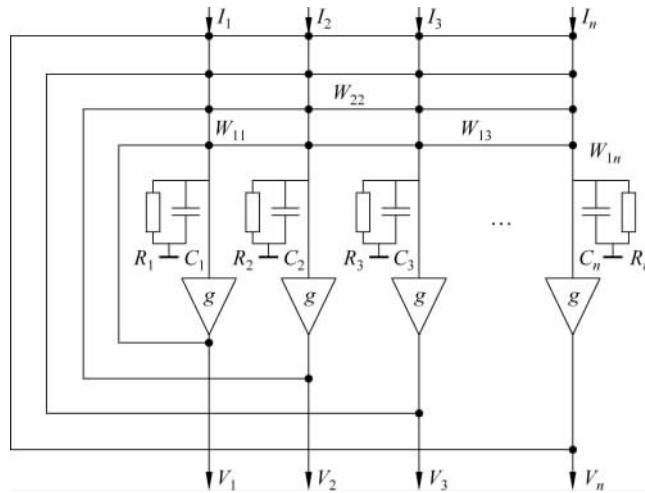


图 3.7 连续 Hopfield 网络的电路形式

对于连续 Hopfield 网络, Hopfield 给出如下稳定性定理。

给出能量函数

$$E(t) = -\frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n W_{ij} V_i(t) V_j(t) - \sum_{j=1}^n V_j(t) I_j + \sum_{j=1}^n \frac{1}{R_j} \int_0^{V_j(t)} g^{-1}(V) dV \quad (3.27)$$

其中, $g^{-1}(V)$ 是 $V_j(t) = g_j(U_j(t))$ 的反函数。

如果连续 Hopfield 网络中神经元传递函数是单调增长的连续有界函数, 并且 $W_{ij} = W_{ji}$, 则有 $\frac{dE(t)}{dt} \leqslant 0$; 当且仅当 $\frac{dV_i(t)}{dt} \leqslant 0$ 有 $\frac{dE(t)}{dt} = 0$ 。

这个定理的意义可以解释如下: 当网络神经元的传递函数是 S 函数, 并且网络权系数矩阵对称, 则随时间的变化网络的能量会下降或不变; 而且仅当输出电位随时间变化不变时, 网络的能量才会不变。换而言之, 在上述条件下的网络是能量不变或下降的。

这个定理说明 Hopfield 网络的能量函数 $E(t)$ 是单调下降的; 如果 $E(t)$ 有下界, 即有确定的极小值; 那么网络必定是稳定的。而且, 可以知道稳定点对应于能量函数的下界, 即极小值。

下一步工作, 只需证明能量函数有下界, 那么, 就可以证明网络是稳定的。

可以证明, 如果 Hopfield 网络的传递函数 g 是连续有界的, 那么, 能量函数 $E(t)$ 是有界的。

最后, 有如下结论: 当 Hopfield 网络的神经元传递函数 g 是连续且有界的, 例如 Sigmoid 函数, 并且网络的权系数矩阵对称, 则这个连续 Hopfield 网络是稳定的。在实际应用中, 任何一个系统, 如果其优化问题可以用能量函数 $E(t)$ 作为目标函数, 那么, 总可以用连续 Hopfield 网络对其进行求解。由于引入能量函数 $E(t)$, Hopfield 使神经网络和问题优化直接对应; 这种工作是具有开拓性的。利用神经网络进行优化计算, 就是在神经网络这

一动力系统给出初始的估计点,即初始条件;然后随网络的运动传递而找到相应极小点。这样,大量的优化问题都可以用连续的 Hopfield 网来求解。这也是 Hopfield 网络用于神经计算的基本原因。

3.5 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。当在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,则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,他长期致力于这方面的研究并希望用数学来刻画人类这项活动,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就是由这种理论的核心内容并经过提高发展然后得出的。

目前,ART理论已提出了三种模型结构,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。

3.5.1 ART模型的结构

ART模型源于Helmholtz无意识推理学说的协作-竞争网络交互模型^[289]。这个模型如图3.8所示。从图中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。在图3.8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。

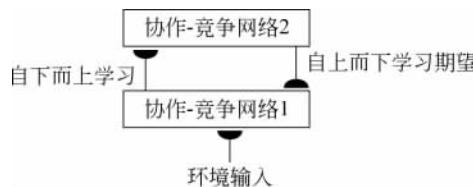


图3.8 协作-竞争网络交互模型

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,这也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。

格罗斯伯格所提出的ART理论模型有如下一些主要优点。

- (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,根据某些特征在全体中所占的比例,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映则自动提高“警觉性”,迅速识别对象。
- (5) 容量不受输入通道数的限制,存储对象也不是正交的。

ART的基本结构如图3.9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模型的框图如图3.10所示。

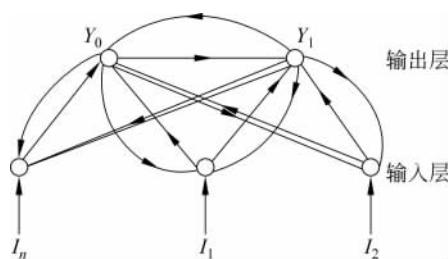


图3.9 ART的基本结构

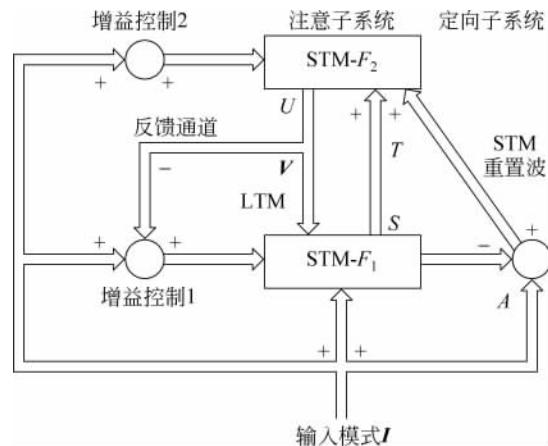


图3.10 ART模型的框图

它由两个子系统组成,一个称为注意子系统(Attentional Subsystem),一个称为取向子系统(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的熟悉事件的编码。

调整子系统的作用是对不熟悉事件产生响应。在有不熟悉事件输入时,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。

简而言之,注意子系统的功能是完成由下向上的向量的竞争选择,以及完成由下向上向量和由上向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模式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。

3.5.2 ART的基本工作原理

在ART模型中,其工作过程采用2/3规则的。所谓2/3规则,就是在ART网络中,三个输入信号中要有两个信号起作用才能使神经元产生输出信号。ART网络的整个工作过程中,2/3规则都在起作用。在说明ART模型的工作原理之前,先介绍2/3规则。

1. 2/3规则

考虑如图3.10所示的ART模型,很明显在 F_1 层中,有三个输入信号源:输入信号I,增益控制输入,自上而下的模式输入。所谓2/3规则,就是指 F_1 中这三个输入信号对 F_1 的激发作用和关系。

2/3规则可以用如图3.11所示的图形来说明。

(1) 自上而下的单输入情况

这时的情况如图3.11(a)所示。 F_1 从三个输入信号源中,只接收来自 F_2 的自上而下的引发模式。故而, F_1 中的三个输入信号源中,并没有两个输入信号源起作用,而只有一个输入信号源即来自 F_2 的自上而下的引发模式。所以 F_1 中的神经元不会被激活, F_1 的神经元不会产生信号输出。

(2) 自下而上的双输入情况

这时的情况如图3.11(b)所示。这时,在 F_1 的三个输入信号源中,有输入信号 I 进行输入,并且有 I 通过增益控制后所产生的对 F_1 的输入;由于这两个输入信号起作用,故而 F_1 中的神经元被激活, F_1 能产生信号输出。

(3) 自下而上输入及自上而下引发模式输入的情况

这时的情况如图3.11(c)所示。它说明了自下而上输入模式和自上而下的引发模式共同作用于 F_1 的过程。这个过程也就是输入模式 I 和来自 F_2 的自上而下的引发模式匹配过程。在这时, F_2 的输出信号会加到增益控制中对其中的输入信号 I 产生抑制作用,所以,增益控制不会产生信号送去 F_1 。在 F_1 中,同时接收到自下而上输入信号以及自上而下的 F_2 输出信号的神经元才会被激活;而只接收到其中一个信号的神经元则不会被激活。

(4) 模态竞争情况

当注意子系统从一个模态向另一个模态转移时,在这个转移的瞬间会禁止 F_1 被激活,因为,这是一个过渡过程,它不反映模式的实质内容。故 F_1 不能被激活。模态竞争的情况如图3.11(d)所示。

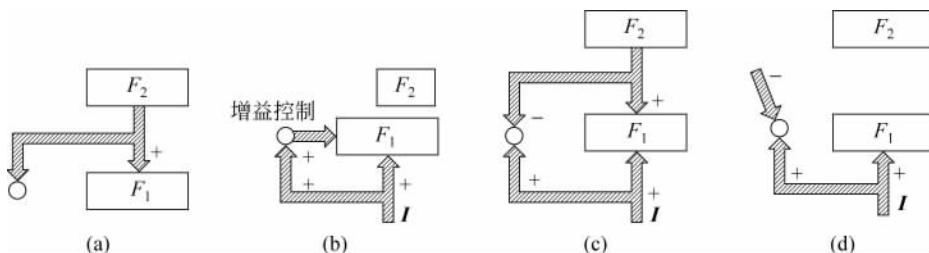
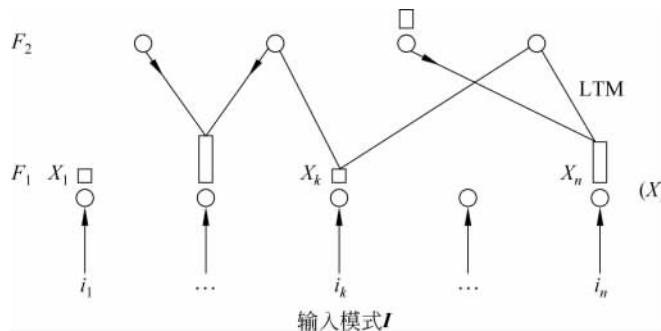


图 3.11 2/3 规则

2. ART模型的基本工作原理

在ART模型中,显然分为 F_1 、 F_2 两层神经网络。对于注意子系统, F_1 和 F_2 这两层的作用可以用图3.12表示。

图 3.12 F_1 和 F_2 层的信息处理

F_1 层接收输入模式 I ,则在 F_1 中被转换成激活模式 X , X 由 F_1 中的激活神经元表示,如图3.12中的长方形所示。这个模式 X 被短期存储在 F_1 中。只有激活值足够高的神经元才能产生输出信号并通过连接传送到 F_2 的神经元去。

在 F_1 中,由 \mathbf{X} 所产生的 F_1 输出模式为 \mathbf{S} , \mathbf{S} 模式通过连接送到 F_2 的神经元输入端,并在 F_2 的神经元的输入端产生一个和 \mathbf{S} 不同的模式 \mathbf{T} 。从 \mathbf{S} 到 \mathbf{T} 的转换称为自适应滤波。无论 F_1 还是 F_2 ,其神经元是一般形式的神经元结构。一般而言,这些神经元的状态,输入和输出并不相同。

在 F_2 层中,模式 \mathbf{T} 经过 F_2 神经元的相互作用会迅速地被转换。这个相互作用是对输入模式 \mathbf{T} 的比较及除弱增强的过程。其结果产生一个短期存储在 F_2 中的模式 \mathbf{Y} ,这也是 F_2 的状态。

在一般情况时,从 \mathbf{T} 到 \mathbf{Y} 这个比较,除弱增强的转换会使多个神经元处于激活状态。这时,这种转换结果变为由 F_2 中的多个神经元群来表达。这个转换过程自动地把 F_1 的输入模式 \mathbf{I} 划分到不相交的各个识别聚类中去,每个类对应于 F_2 中的某个特征神经元。在特殊情况下,从 \mathbf{T} 到 \mathbf{Y} 的比较、除弱增强过程就是在 F_2 中选择一个与当前输入 \mathbf{I} 相对应的而输出值最大的神经元的过程,所选择的神经元就是用于表示激活模式的唯一神经。为了说明 ART 模型有关工作的基本原理,下面分 5 点进行介绍。

(1) 自下而上的自适应滤波和 STM 中的对比度增强过程。

输入信号 \mathbf{I} 加到注意子系统的 F_1 的输入端,经过 F_1 的节点变换成激活模式 \mathbf{X} ,这一过程起到特征检出作用。在 F_1 中,激活值较高的神经元就会有输出到 F_2 的信号,并成为输出模式 \mathbf{S} 。 \mathbf{S} 经过 F_1 到 F_2 的连接通道时受到加权组合(LTM),变换成模式 \mathbf{T} 后作用于 F_2 的输入端。 \mathbf{S} 到 \mathbf{T} 的转换称为自适应滤波。 F_2 接收到 \mathbf{T} 后通过神经元间的相互作用迅速产生对比度增强了的激活模式 \mathbf{Y} ,并且存储于 F_2 中。如图 3.13(a)所示。

这个阶段的学习是一个变换系列: $\mathbf{I} \rightarrow \mathbf{X} \rightarrow \mathbf{S} \rightarrow \mathbf{T} \rightarrow \mathbf{Y}$ 。

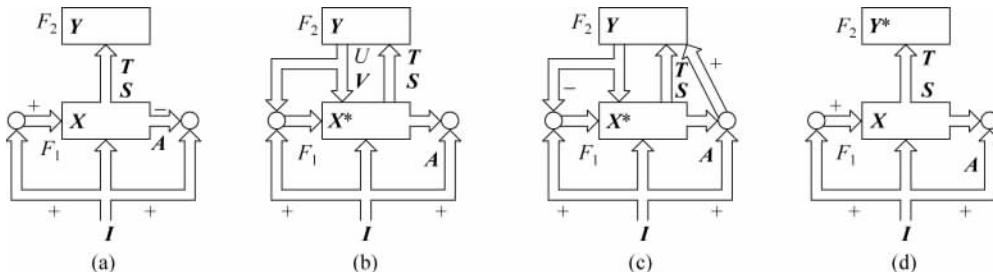


图 3.13 ART 模型的工作过程

(2) 自上而下的学习期望匹配和对已学习编码的稳定。

一旦当自下而上的变换 $\mathbf{X} \rightarrow \mathbf{Y}$ 完成之后, \mathbf{Y} 就会产生自上而下的输出信号的模式 \mathbf{U} ,并送向 F_1 ,只有激活值足够大的才会向反馈通道送出信号 \mathbf{U} 。 \mathbf{U} 经加权组合变换成模式 \mathbf{V} 。 \mathbf{V} 称为自上而下的模板,或学习期望。

由于 \mathbf{X} 和 \mathbf{V} 这两个模式对 F_1 输入,则它们的共同作用在 F_1 中产生激活模式 \mathbf{X}^* 。一般而言, \mathbf{X}^* 和只由输入信号产生的 \mathbf{X} 模式不同。这时, F_1 的作用就是试图使 \mathbf{V} 和 \mathbf{I} 匹配,其匹配结果确定了以后的作用过程。这时的情况如图 3.13(b)所示。

(3) 注意子系统和取向子系统相互作用过程。

这个过程和输入 \mathbf{I} 有关。在图 3.13(a)中,在输入模式 \mathbf{I} 产生 \mathbf{X} 的同时,也会激活取向子系统 \mathbf{A} ;只是在 \mathbf{A} 产生输出之前, F_1 中的 \mathbf{X} 对 \mathbf{A} 所产生的输出端起禁止作用。当 F_2 的

反馈模式 \mathbf{V} 与 F_1 的输入模式失配时, 就会大大减弱这一禁止作用, 当减弱到一定的程度时, \mathbf{A} 就被激活。如图 3.13(c)所示。

\mathbf{A} 被激活之后就向 F_2 送出重置信号, 并作用于 F_2 的全部神经元, 从而改变 F_2 的状态, 取消了原来的自上而下的学习期望 \mathbf{V} ; 终止了 \mathbf{V} 和 \mathbf{I} 的失配; 于是输入 \mathbf{I} 再次作用直到 F_2 产生新的状态 \mathbf{Y}^* 。如图 3.13(d)所示。

\mathbf{Y}^* 会产生新的自上而下的学习期望 \mathbf{V}^* , 如果 \mathbf{V}^* 仍然和 \mathbf{I} 失配, 那么, 取向子系统 \mathbf{A} 继续起作用; 这样, 产生一个快速的一系列匹配与重置过程。这个过程控制 LTM 的搜索从而调整了 LW 对外界环境的编码。这个过程一直执行下去, 直到 F_2 送出的模式 \mathbf{V} 和输入 \mathbf{I} 相互匹配为止。

(4) 需考虑的一些特点。

在注意子系统的增益控制及起动这一自上而下的学习期望匹配过程中, 还应考虑一些有关特点。

例如在 F_1 输出自下而上的作用之前, F_2 已被激活。这时 F_2 就会产生自上而下的学习期望并作用于 F_1 ; 这时 F_1 也会被激活, 并产生自下而上的作用过程。显然, 需要对来自外部输入的激活以及来自 F_2 的反馈激活进行区分。所以, 设置一个辅助机构进行区分激活来源的工作。这个辅助机构称为注意增益控制。

为 F_2 被激活时, 注意起动机机构会向 F_1 发送学习期望信号, 注意增益控制就会给出禁止作用, 从而影响 F_1 对输入响应灵敏度, 使得 F_1 可以区分激活信号的来源。

(5) 匹配。

采用 2/3 规则, 以确定 F_1 的输出。这实际上是把存储模式和输入模式进行匹配的规则。

3. ART 模型的工作过程

在图 3.10 所示的 ART 结构中, F_1 可称为比较层, F_2 可称为识别层。

比较层 F_1 接收输入模式 \mathbf{I} , 初始时不作任何变动作作为输出向量 \mathbf{S} 送去识别层 F_2 ; 此后, F_1 同时接受识别层输出的向量 \mathbf{V} 和输入模式 \mathbf{I} , 还有增益控制的输出, 并按 2/3 规则产生输出。在初始时, 增益控制的输出为 \mathbf{I} , 而 \mathbf{V} 设置为 0, 故有 \mathbf{S} 等于输入 \mathbf{I} 。

识别层 F_2 是用作输入向量分类器的。在识别层中, 只有一个神经元和输入的向量 \mathbf{S} 最优匹配, 这个神经元就会被激活, 而其他神经元则被抑制。根据神经元的结构原理, 最优匹配规则如下:

$$\sum_i W_{ik} \mathbf{S}_i = \max \sum_i W_{ij} \mathbf{S}_i \quad (3.28)$$

其中, \mathbf{S} 是输入 F_2 的向量; $\mathbf{S} = (S_1, S_2, \dots, S_n)$; \mathbf{W}_j 是识别层中第 j 个神经元和比较层中神经元从 F_1 — F_2 的权系数向量 $\mathbf{W}_j = (W_{1j}, W_{2j}, \dots)$; \mathbf{W}_c 是识别层中最优匹配神经元 C 从 F_1 — F_2 的权系数向量 $\mathbf{W}_c = (W_{1c}, W_{2c}, \dots)$ 。

注意: 最优匹配神经元 C 到比较层神经元有从 F_2 — F_1 的权系数向量 \mathbf{W}_c' , $\mathbf{W}_c' = (W_{c1}, W_{c2}, \dots)$ 。很明显, \mathbf{W}_c 和 \mathbf{W}_c' 就组成了输入向量的类别样本, 也是权系数的形态表示一类模式。

在识别层中, 为了使一个神经元有最大输出值取得竞争的优胜, 并抑制其他神经元。故而识别层有横向连接, 每个神经元的输出和正的权系数相乘后作为本神经元的一个输入, 而其他神经元的输出和负权系数相乘后再作为本神经元的输入。这种作用等于加强自身, 抑制其他神经元。

制其他,从而保证了只有一个神经元被激活。这种情况如图 3.14 所示。

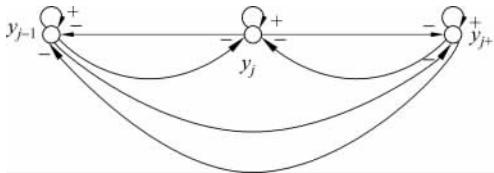


图 3.14 F_2 层的横向连接

增益控制有两部分,它们的作用功能不同。识别层 F_2 的增益控制输出原则为:只要输入向量 I 有一个元素为 1,则输出 1。比较层 F_1 的增益控制原则为:只要在 I 有一个元素为 1,同时 F_2 的输出向量 U 全部元素为 0 时,才输出 1。

重置作用是在输入信号 I 和 F_1 的输出 S 之间的匹配存在问题,差别大于某警戒值时,则发清零信号到 F_2 。以便重新进行识别。

ART 网络的学习分类过程分为三步,即识别、比较和搜索。下面作简要介绍。

(1) 识别

初始化时,网络无输入信号,故 I 全部元素为 0;识别层 F_2 增益控制输出为 0,识别层 F_2 输出全部为 0。在有模式 I 输入后, I 必有元素为 1,故 F_1 增益控制、 F_2 增益控制均输出 1;比较层 F_1 按 2/3 规则全部复制 I 作为输出; $S=(S_1, S_2, \dots, S_n)$ 。接着识别层 F_2 的每个神经元 j 执行下面操作;从而求出最优匹配神经元 C :

$$\sum_i W_{ik} S_i = \max \sum_i W_{ij} S_i$$

则神经元 C 输出 1,其余输出 U 。这些输出送回比较层 F_1 。 F_2 输出的值为 $U=(U_1, U_2, \dots)$ 。找寻最优匹配神经元 C 的过程就是识别。

(2) 比较

从识别层 F_2 反馈到 F_1 的向量 U 不再全部为 0, F_1 增益控制输出 0。按 2/3 规则,只有输入及反馈向量 U 的元素同时为 1,所激励的神经元才会被激活。从另一个角度讲,就是来自 F_2 的反馈强迫输入向量 I 中那些不匹配存储模式 U 的 S 元素为 0。

如果 I 与 U 不匹配,则产生的 S 只有少数元素为 1,这也说明模式 U 不是所要寻找的 I 模式。取向子系统对 I 和 S 的匹配程度进行判别,如果低于给定的警戒值,则发出重置信号,使识别层 F_2 激活的神经元清零;这也说明该神经元失去竞争的资格。到此这个分类阶段比较过程结束。如果 I 与 U 匹配,则输入模式 I 所属的类别已找到,分类结束。

(3) 搜索

在 I 与 U 不匹配时,为了找到较好的匹配必须对其他的模式进行搜索。重置信号把识别层 F_2 的神经元全部清 0,则 F_1 增益控制又输出 1,网络返回到初始状态。输入模式 I 再进行输入,识别层的另一个神经元会取得优胜,则反馈一个新的存储模式 U 送回比较层 F_1 。接着又进行匹配比较,如不匹配,则又重置识别层……不断执行下去。搜索过程直到产生下列情况之一才会停止。

① 找到一个存储模式,在警戒值范围内和输入模式 I 匹配,则 ART 网络进入学习阶段。修正和匹配神经元 C 相关的权系数 W_{ik} 和 W_{ci} 。

② 搜索了全部模式后,没有一个模式能够和 I 相似匹配,则网络也进入学习阶段。把原

来分配模式的神经元 j 赋于输入式 \mathbf{I} , 构造相应的权系数 \mathbf{W}_{ij} 和 \mathbf{W}_{ji} , 并作为样本模式存储。

特别应指出的是: 搜索过程包含了识别和比较二个阶段。搜索是识别-比较-识别-比较的多次重复。

严格地说, ART 应分成搜索和学习这两种最主要的过程和功能。

3.5.3 ART 模型的数学描述

在 ART 模型中, F_1 或 F_2 中的神经元用 N_k 表示, 神经元被激活后产生的激活值用 X_k 表示, 从神经生理学的研究结果, 可以知道神经元的激活值, 即神经元未经 S 函数处理的输出 X_k 满足下面的微分方程:

$$\epsilon \frac{dX_k}{dt} = -X_k + (1 - AX_k)J_k^+ + (B + CX_k)J_k^- \quad (3.29)$$

其中, ϵ 是远小于 1 的正实数, J_k^+ 是送到神经元 N_k 的所有激励输入之和, J_k^- 是送到神经元 N_k 的所有抑制输入之和, A, B, C 是非负常数, X_k 的取值范围为 $[-BC-1, A-1]$ 。

1. F_1 层的数学描述

用 N_i 表示 F_1 的神经元, 并且 $i=1, 2, \dots, n$, 则有

$$\epsilon \frac{dX_i}{dt} = -X_i + (1 - A_1 X_i)J_i^+ + (B_1 + C_1 X_i)J_i^- \quad (3.30)$$

很明显, 有 F_1 的激活模式:

$$X = \{X_1, X_2, \dots, X_n\}$$

(1) J_i^+ 的形式

由于 F_1 神经元 N_i 的激励输入 J_i^+ 是自下而上的输入 \mathbf{I}_i 以及自上而下的输入 \mathbf{V}_i 之和, 故有 $J_i^+ = \mathbf{I}_i + \mathbf{V}_i$, 其中 \mathbf{I}_i 是一个 n 维输入向量, $\mathbf{I} = \{I_1, I_2, \dots, I_n\}$;

$$\mathbf{V}_i = D_1 \sum_i f(X_i) \mathbf{W}_{ji} \quad (3.31)$$

这里, $f(X_j)$ 是 F_2 中神经元 N_j 的输出, \mathbf{W}_{ji} 是 N_j 到 N_i 的连接权系数; D_1 是系数。 $\mathbf{V} = \{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n\}$, 也是 n 维向量。

(2) J_i^- 的形式

对 F_1 层, 抑制输入 J_i^- 是由注意子系统的增益控制信号来控制, 即

$$J_i^- = \sum_i f(X_i) \quad j = n+1, n+2, \dots, n+m \quad (3.32)$$

当且仅当 F_2 的激活值很高时, $J_i^- = 0$, 否则 $J_i^- > 0$ 。

2. F_2 层的数学描述

用 N_j 表示 F_2 的神经元, 并且 $j=n+1, n+2, \dots, n+m$, 则有

$$\epsilon \frac{dX_j}{dt} = -X_j + (1 - A_2 X_j)J_j^+ + (B_2 + C_2 X_j)J_j^- \quad (3.33)$$

则有 F_2 的激活模式 $Y = \{X_{n+1}, X_{n+2}, \dots, X_{n+m}\}$ 。

选择 F_2 中的激活模式的输入和参数, 使到 F_2 中具有来自 F_1 的最大输入的神经元取得竞争的胜利。故而对 J_j^+ 和 J_j^- 考虑应有如下形式。

(1) J_j^+ 的形式

$$J_j^+ = g(X_j) + T_j \quad (3.34)$$

其中, $g(X_j)$ 为 N_j 的自反馈信号, T_j 是从 F_1 来的到 F_2 的输入模式:

$$T_j = D_2 \sum_i h(X_i) W_{ji} \quad (3.35)$$

这里的 $h(X_i)$ 是 F_1 中神经元 N_i 的输出, D_2 是系数, W_{ji} 是 F_1 到 F_2 的神经元的连接权系数。

(2) J_i^- 的形式

$$J_i^- = \sum_{k \neq j} g(X_k) \quad (3.36)$$

迄今为止, 按生物神经网络(Biological Neural Networks, BNN)巨量并行分布方式构造的各种人工神经网络, 虽然已经在信息处理中扮演着越来越重要的角色, 但是并没有显示出人们所期望的聪明智慧来。对以仿效大脑神经系统为目的的人工神经网络的研究历程进行一些分析和反思, 探讨下一步可能采取的方法步骤, 对今后智能信息科学的进一步发展将是有益的。要让 ANN 更好更快地向 BNN 学习, 就有必要对今后 ANN 的主攻方向、研究路线、方法步骤、关键技术和应当采取的措施等方面, 作一些考虑和调整。

概括起来, 为了使 ANN 向 BNN 学习得更好更快, 可以从以下几个方面着手。

(1) 明确 ANN 的主要智能优势是擅长非精确性信息处理。按照输入和输出特性的不同, 可将智能系统最经常处理的信息类型归纳为四类。其中, 数值计算和逻辑推理所对应的信息处理任务, 现行的冯·诺依曼数字计算机有着成熟而巨大的能力, ANN 既无必要也无优势在此等领域中与之竞争。其他三类都在输入信息或输出信息中包含有非精确信息处理的内容, 因而它们是 ANN 能发挥作用的领域。人类的绝大部分脑力劳动正是投入在这三类信息处理上。

(2) 探索新的 ANN 体系结构。神经解剖学方面的研究表明, 在 BNN 中除了有由神经细胞体-轴突-突触-树突构成的神经电位脉冲电路系统外, 还有一个在前后突触间约 $0.02\mu s$ 的间隙中释放化学性神经递质(谷氨酸)的调节系统, 这个化学递质系统的作用相当于半导体的栅极电路, 起着至关紧要的调控放大作用。因此, 如何建立神经电脉冲系统和化学递质系统合成的耦合系统数学模型, 分析其工作机理并提出可行简便算法, 是从建模上使 ANN 更靠近 BNN 的一条值得重视的途径。

(3) 寻求新的网络拓扑结构和相应的学习算法。迄今为止, 以和-积式神经元为基础的前馈型多层(特别是三层)神经网络, 与误差反向传播学习算法相配合的 ANN 结构体制, 获得了最广泛的应用。但是, 它在性能上仍存在若干待改进的地方。例如, 它只能调整权值, 不能调整网络拓扑结构, 无法实现注意力集中功能; 学习新样本时, 会“冲乱”原已学好保存下来的旧样本; 其学习算法中, 包含有较复杂的非线性激活函数的求导运算过程等。

3.6 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^[305]。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个方面,即对结论生成方法以及对网络个体生成方法。

3.6.1 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^[305]。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

$$p_{\text{err}} = \sum_{k>N/2}^N \binom{N}{k} p^k (1-p)^{N-k} \quad (3.37)$$

在 $p < 1/2$ 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(3.37)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^[389]。假设学习任务是利用 N 个神经网络组成的集成对 $f: \Re^n \rightarrow \Re$ 进行近似,集成采用加权平均,各网络分别被赋以权值 w_a ,并满足式(3.38)和式(3.39)。

$$w_a > 0 \quad (3.38)$$

$$\sum_a w_a = 1 \quad (3.39)$$

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

$$\bar{V}(X) = \sum_a w_a V^a(X) \quad (3.40)$$

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

$$E^a = \int dx p(x) (f(x) - V^a(x))^2 \quad (3.41)$$

$$E = \int dx p(x) (f(x) - \bar{V}(x))^2 \quad (3.42)$$

各网络泛化误差的加权平均为

$$\bar{E} = \sum_a w_a E^a \quad (3.43)$$

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

$$A^a = \int dx p(x) (V(x) - V^a(x))^2 \quad (3.44)$$

$$\bar{A} = \sum_a w_a A^a \quad (3.45)$$

则神经网络集成的泛化误差为

$$E = \bar{E} - \bar{A} \quad (3.46)$$

式(3.46)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。

3.6.2 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^[244],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(3.47),其中 ε_t 为预测函数 h_t 的训练误差, $\gamma_t = 1/2 - \varepsilon_t$ 。

$$\begin{aligned} H &= \prod_t [2 \sqrt{\varepsilon_t(1-\varepsilon_t)}] \\ &= \prod_t \sqrt{1-4\gamma_t^2} \leqslant \exp\left(-2t \sum_t \gamma_t^2\right) \end{aligned} \quad (3.47)$$

从式(3.47)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^[101]。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(3.48)和式(3.49)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)。

$$r^* = \int \max_j P(j | x) P_X(x) \quad (3.48)$$

$$r_A = \int_{x \in C} \max_j P(j | x) P_x(dx) + \int_{x \in C'} \left[\sum_j I(\phi_A(x) = j) P(j | x) \right] P_X(x) \quad (3.49)$$

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。

3.7 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——脉冲耦合神经网络模型(Pulse-Coupled Neural Network, PCNN)。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,霍奇金与哈斯利(A F Huxley)开始研究神经元电化学特性^[324]。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^[325,324]。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^[201,283]。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^[202]。对猴的大脑皮层进行的试验中,也得到了相类似的试验结果。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^[346]。通过对艾克霍恩提出的模型进行改进,就形成脉冲耦合神经网络(PCNN)模型。于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,在解决图像处理具体应用时取得令人满意的性能。

3.7.1 Eckhorn 模型

1990年,根据猫的视皮层的同步振荡现象,艾克霍恩提出一个脉冲神经网络模型^[202],如图3.15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,和起调制作用的连接(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。

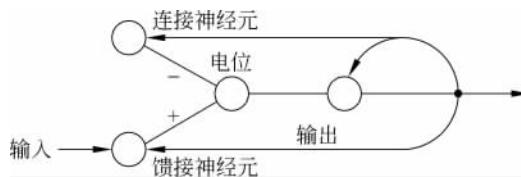


图3.15 Eckhorn神经元模型示意图

Eckhorn模型可用如下方程描述:

$$U_{m,k} = F_k(t)[1 + L_k(t)] \quad (3.50)$$

$$F_k(t) = \sum_{i=1}^N [w_{ki}^f Y_i(t) + S_k(t) + N_k(t)] \otimes I(V^a, \tau^a, t) \quad (3.51)$$

$$L_k(t) = \sum_{i=1}^N [w_{ki}^l Y_i(t) + N_k(t)] \otimes I(V^l, \tau^l, t) \quad (3.52)$$

$$Y_k(t) = \begin{cases} 1, & U_{m,k}(t) \geq \theta_k(t) \\ 0, & \text{其他} \end{cases} \quad (3.53)$$

这里,一般表示为

$$X(t) = Z(t) \otimes I(v, \tau, t) \quad (3.54)$$

即

$$X[n] = X[n-1]e^{-t/\tau} + VZ[n], \quad n = 1, 2, \dots, N \quad (3.55)$$

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

3.7.2 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),即连接域(Linking Field)。如果去掉连接输入部分,Eckhorn模型中的神经元模型与常规的神经元模型没什么不同,而正是连接输入的引入,使我们对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^[341]。它很快被应用到图像处理领域,而它和它的许多变种模型被一起称为脉冲耦合神经网络(PCNN)。

图3.16给出了脉冲耦合神经元示意图。神经元主要有两个功能单元构成:馈接输入域和连接输入域,分别通过突触连接权值 M 和 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。馈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:

$$F_{ij}[n] = e^{aF\delta_n} F_{ij}[n-1] + S_{ij} + V_F \sum_{kl} M_{ijkl} Y_{kl}[n-1] \quad (3.56)$$

$$L_{ij}[n] = e^{aL\delta_n} L_{ij}[n-1] + V_L \sum_{kl} K_{ijkl} Y_{kl}[n-1] \quad (3.57)$$

式中, F_{ij} 是第 (i,j) 个神经元的馈接, L_{ij} 是耦合连接, Y_{kl} 是 $(n-1)$ 次迭代时神经元的输出。两个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 M 和 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, M 和 K 有多种取值选择方式,但选择要合适,一般不宜过大。

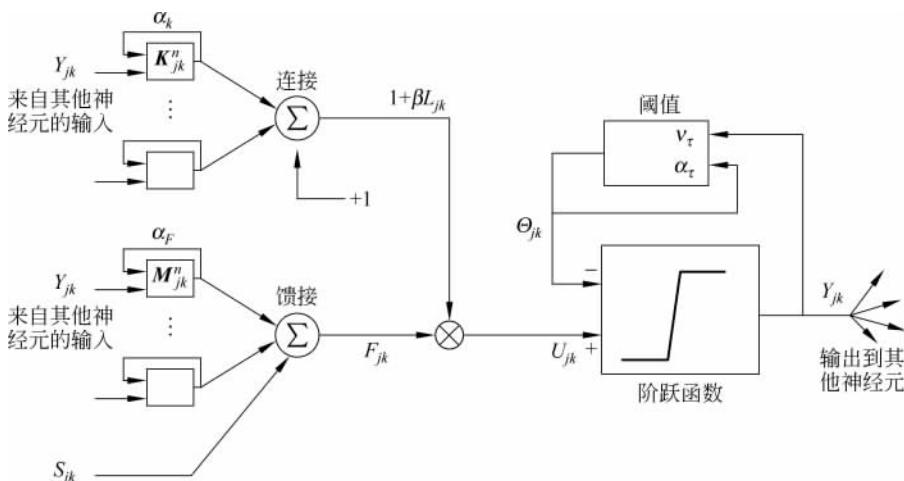


图3.16 脉冲耦合神经元示意图

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:

$$U_{ij}[n] = F_{ij}[n]\{1 + \beta L_{ij}[n]\} \quad (3.58)$$

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:

$$Y_{ij}[n] = \begin{cases} 1, & U_{ij}[n] > \Theta_{ij}[n] \\ 0, & \text{其他} \end{cases} \quad (3.59)$$

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

$$\Theta_{ij}[n] = e^{-\theta\delta_n} \Theta_{ij}[n-1] + V_\Theta Y_{ij}[n] \quad (3.60)$$

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(3.56)~式(3.60),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(3.60)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,每一时刻的发放图就是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,每一时刻的发放图就是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输入的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 神经元间存在连接但不一定存在影响,存在影响但不一定存在连接,这一现象更加突出 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起的有序状态,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,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连接。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。从而这一连接会引起邻近神经元比原来更快地点火,这样就导致了在图像的一个大的区域上产生同步振荡: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。

3.7.3 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(Feeding)输入,另一类是连接(Linking)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 3.17 给出了我们模型中的一个神经元输入耦合方式的示意。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。

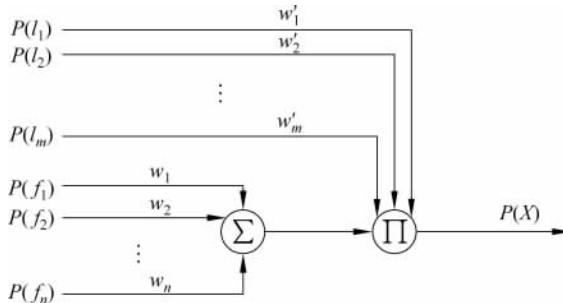


图 3.17 BLFN 模型中神经元输入耦合示意

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 Feeding 输入和 Linking 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。

3.8 神经场模型

人工神经网络被认为是具有广泛非线性关系逼近机理的学习机和非线性数据关系分类器，极其重要的是它提供了一种对于非线性函数关系的逼近理论。1985 年甘利俊一提出了信息几何并用于神经计算理论的研究^[14]，其基本的思想是把所有神经网络变换所组成的空间看作为流形空间，流形空间是一般欧氏空间和 Hilbert 空间的推广，它使我们可以在一个更一般化的非线性空间和非欧氏空间上建立逼近理论和拓扑校正理论，通过几何结构和拓扑结构分析来更理解神经网络模型的变换机理和学习问题。史忠植和张建从微分流形和拓扑变换的角度来分析神经网络结构的表示、系统的识别、变换机理，以此来理解神经网络在更一般信息处理系统空间中的组织结构、定位机理。我们考虑神经计算的整体结构，提出了场组织变换与场反应变换的整体结构耦合，作为人工神经网络模型信息处理、变换机理新的表示方法；引入了模块化、层次化，还有模型的可扩展性，利用流形上的对偶几何结构对它作解释，提出相应的学习算法^[974]。

3.8.1 神经场表示

神经网络包括输入单元、隐单元和输出单元。由于有隐单元存在，隐单元被看作为不可观测量，在信息处理中是需要被估计的值。我们把神经网络的信息处理理解为非完整数据下的信息处理。我们对神经网络的变换机理给出一种新的形式化描述。神经网络变换机理包括两个过程：神经场组织过程，通过隐单元实现从输入模式的表示；神经场反应过程，实

现从表示到输出的变换(如图3.18所示)。这里, X 是输入空间, $X \times Z$ 是表示空间, 或称作为神经场表示空间, Y 是输出空间。

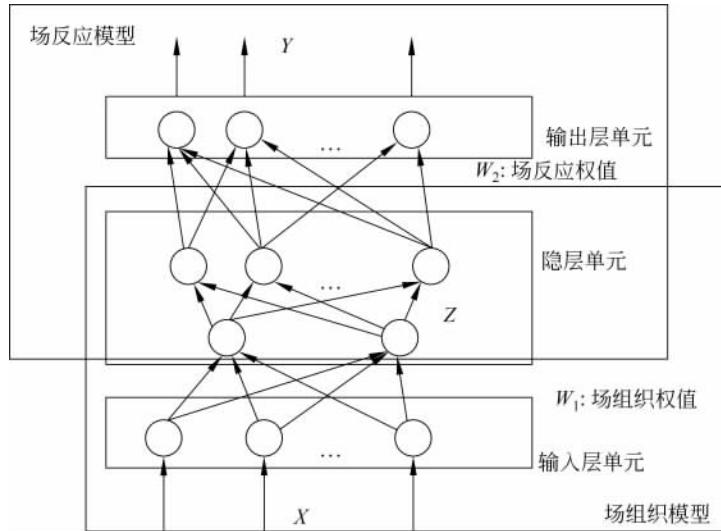


图 3.18 神经场模型

场组织变换模型表示为一个变换 $T: X \rightarrow X \times Z$, 其中, Z 是不可观测量空间, 或是辅助信息量空间, 场组织模型实现输入模式的特征表示, 它可能由算法或是映射而实现; 它的目的是抽取输入模式的特征结构, 使期望的输出模式更容易; 在神经网络中, 由场组织模型所包含的连接权值, 称为场组织权值或认知权值。在场组织模型权值为 W_1 的情形下, 场组织模型表示为

$$q_z(x) = z(x; W_1) \quad (3.61)$$

场反应模型表示一个变换 $g: X \times Z \rightarrow Y$, 表示从表示空间到输出空间的变换, 即通过对输入模式特征的估计后, 由特征模式表示下的输出。由场反应模型所包含的神经网络连接权值, 我们称之为场反应权值或产生权值。在场反应权值为 W_2 的情形下, 场反应模型表示为

$$p(y) = \sum_z p(z) p(y | z; W_2) \quad (3.62)$$

完整的人工神经网络系统形式化为含隐变量的概率分布 $P(y, z(x, W_1) | x; \theta(W_2))$, 其中, W_1, W_2 分别表示神经网络中的场组织权值和场反应权值, θ 是用来识别统计模型的参数。下面, 用统计流形的理论来研究人工神经网络模型 $P(y, z(x, W_1) | x; \theta(W_2))$ 的结构, 作为信息处理的假设, 我们把人工神经网络系统嵌入到一个相关的统计流形结构中, 称所有概率分布 $P(y, z | x; \theta)$ 组成的流形 S 为人工神经网络模型 $P(y, z(x, W_1) | x; \theta(W_2))$ 的相关信息系统

$$S = P(y, z | x; \theta) \quad (3.63)$$

一般假设 S 是一个可编码成指数簇结构的平坦流形结构, 而神经网络可编码成 S 的子流形结构:

$$\begin{aligned} &P(y, z(W_1) | x; \theta(W_2)) \\ &= \exp\{\sum_i r_i(y, z(W_1)) \theta_i(W_2) + k(r) - \psi(\theta)\} \end{aligned} \quad (3.64)$$

S 中的概率分布可由 e -坐标 θ 和 m -坐标 η 进行识别, 即

$$S = \{\theta \mid P(r(y, z) \mid x; \theta)\} = \{\eta \mid P(r(y, z) \mid x; \eta)\} \quad (3.65)$$

在对偶流形中, 神经网络的场组织模型结构表示为

$$R = \{\eta(W_1) \mid P(r(y, z(x, W_1)) \mid x; \eta(W_1))\} \quad (3.66)$$

表示在隐变量已知情况下, 即是完整数据下, 模型的估计。

神经网络的反应模型表示为

$$M = \{\theta(W_2) \mid P(r(y, z(x, W_1)) \mid x; \theta(W_2))\} \quad (3.67)$$

在对偶流形中, 神经网络的场组织模型和场反应模型变换分别表示为

$$\text{场组织变换: } T: x \rightarrow \eta(y, z(x, W_1)) \quad (3.68)$$

$$\text{场反应变换: } g: \eta(y, z(x, W_1)) \rightarrow \theta(W_2) \quad (3.69)$$

复杂结构神经网络模型可以处理简单输入/输出关系的逼近问题, 但是由于复杂模型可变量(自由度)的增加, 所得出的结果不一定会反映问题本身的结构, 复杂模型处理简单问题泛化能力差。因此根据 Occam's razor 原理, 在有冗余参数下的模型结构中, 应该寻找最简模式。在神经流形结构中, 数据子流形 D 和产生模型子流形 M 的拓扑结构由单复形进行编码, 可以通过单复形的边缘链分解, 实现模型的分解和组织。通过边缘链分解的分维组织有非常明确的几何直观解释, 对于冗余参数结构的模型, 通过边缘链分解(边缘投影), 消去与问题无关的冗余参数, 边缘投影算子可以通过同伦变换而实现。

有了神经网络模型整体结构的流形和拓扑结构表示, 神经网络对于复杂系统的信息处理和模型化可以描述为在非完整性数据估计与冗余参数结构下参数选择耦合的过程, 对复杂系统模型化神经网络的组织过程描述为如图 3.19 所示的链结构。

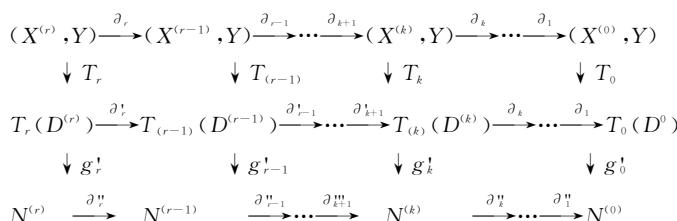


图 3.19 神经场自适应神经网络模型结构

图 3.19 中, $T_i(r^i) = T_i(r^i(X, Z))$, ∂ 是边缘链分解投影变换, T 是拓扑校正逼近变换, g 是对偶投影变换。通过图 3.19, 我们提供了一个基于结构分析的框架来讨论神经网络的分维组织和信息处理机理, 它对复杂系统的模型化是通过多集模型的神经网络通过层次化的组织而实现, 复杂系统的信息处理理解为在非完整性数据估计和冗余参数结构下的逼近理论。以上的结构具有表达模型的可扩展性, 信息处理的分维、简单模型嵌入复杂模型的机理。层次化表达为如下两个过程。

(1) 定性分析(拓扑校正逼近变换 T): 对输入模式的结构进行定性分析, 把它嵌入相关的模型结构中, 可以引入适当的隐变量和辅助量, 对输入模式进行估计, 这是基于结构逼近的基础。

(2) 定量分析(信息几何框架下的投影变换 g): 在拓扑变换的指导下, 对输入模式的特征表示进行定量分析, 选择优化的输出。

3.8.2 神经场学习理论

人工神经网络最重要的特征是学习能力,通过神经网络内部权值的更替,学习外部环境结构的特征和表示。神经网络的学习主要涉及神经网络拓扑结构的建立和设计,在固定拓扑结构的意义下,网络权值(参数)的优化和选择。一般地,神经网络的学习主要是优化某种行为的度量和测度,而这个度量和测度代表了学习问题的拓扑结构特性和模型特征。神经网络可以被看成一种学习机,如图 3.20 所示。

1. EM 学习算法

EM(Expectation-Maximization)学习算法是由登朴斯特(A P Dempster)等在 1977 年提出的统计模型参数估计的一种算法框架,通过交替搜索方法,使样本集的邻域函数最大化^[180]。当可观测的数据被表示为 x 时,我们认为它是非完整的数据,拥有一个完整的数据向量 y ,这个数据包括损失的变量 z ;我们识别一个概率统计模型,应该确定在完整数据下的概率分布 $P(x, z; \theta)$ 。完整数据的邻域函数表示为 $l_c(\theta, y) = \log P(x, z; \theta)$,而非完整数据的邻域函数表示为 $l(\theta, x) = \log P(x; \theta)$ 。

EM 算法的每一次交替包括两个步骤:估计 E 步和最大化 M 步。 M 步包括对于在一次交替过程中每一个 E 步所定义的邻域函数的最大化。具体 EM 算法的步骤分为以下的两步。

(1) E 步: 给予可观测数据 x 和由目前模型参数量 $\hat{\theta}^k$ 表示的模型,计算完整数据邻域的数学期望,得到

$$\begin{aligned} Q(\theta, \hat{\theta}^k) &= E_z[l_c(\theta | x)] \\ &= \int P(z | x, \hat{\theta}^k) \log P(x, z | \theta) dz \end{aligned} \quad (3.70)$$

其中, E 是数学期望, θ 是参数向量,是指对于非完整数据的一个估计,也就是说它是对隐变量的数学期望或完整数据的一个估计。

(2) M 步: 确定更替值 $\hat{\theta}^{k+1}$, 它计算

$$\hat{\theta}^{k+1} = \arg_{\theta} \text{Max}(Q(\theta, \hat{\theta}^k)) \quad (3.71)$$

在一些情况下,有时很难计算 $Q(\theta, \hat{\theta}^k)$,这是由 M 步所要求的过程。我们仅仅能计算 $\hat{\theta}^k$ 的一个新的值使 Q 的值增加,这样的情况下,我们称之为广义的 EM 学习算法,参数值的变换满足

$$\hat{\theta}^{k+1} = M(\hat{\theta}^k) \quad (3.72)$$

这里

$$Q(M(\hat{\theta}^k), \hat{\theta}^k) \geq Q(\hat{\theta}^k, \hat{\theta}^k)$$

以上的方法,提供了一种特殊的求解问题的方法。EM 算法的一个交替的步骤是选择一个参数值来增加 Q 的值,完整邻域的期望。

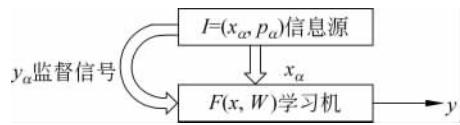


图 3.20 神经网络的简单学习过程

2. 对偶校正学习算法

在信息几何中,Amari提出了em算法^[15],把神经网络模型表示成为下述的指数簇结构

$$P(y, z | x; \theta) = P(y, z | x; \theta(W_2)) \quad (3.73)$$

把这个模型嵌入到下面的指数簇结构中

$$P(y, z | x; \theta(W_2)) = \exp\{\sum r_i(y, z) \theta_i(W_2) + k(r) - \psi(\theta(W_2))\} \quad (3.74)$$

我们把神经网络的权值分离成场组织权值和场反应权值,神经网络的模型被描述为

$$P(y, z | x; \theta) = P(y, z(W_1) | x; \theta(W_2)) \quad (3.75)$$

把这个模型嵌入到下面的指数簇结构中

$$\begin{aligned} & P(y, z(W_1) | x; \theta(W_2)) \\ &= \exp\{\sum r_i(y, z(x; W_1)) \theta_i(W_2) + k(r(y, z(W_1))) - \psi(\theta(W_2))\} \end{aligned} \quad (3.76)$$

在训练样本集 $Y=\{x^t, y^t\}, t=1, 2, \dots, N\}$ 下,神经网络模型的信息几何学习理论,称为对偶校正学习算法(DCL)。神经网络包括两类连接权值,场组织模型连接权值 W_1 ;由场组织模型作用产生表示 $z(W_1)$,称为非完整数据结构的估计;同样的道理,神经网络的场反应变换连接权值 W_2 ,当表示为 z ,由产生权值作用所产生的输出模式为 y ,神经网络模型表示为概率分布 $P(y, z(W_1) | x; \theta(W_2))$ 。

对偶校正学习算法是基于非完整数据假设的微分流形投影理论的一种方法,其基本的结构依赖于神经网络模型的指数簇结构编码。当神经网络模型 M 作为子流形嵌入到带隐变量的指数簇流形 S 中,由 θ 坐标识别;训练集 $Y=\{x^t, y^t\}, t=1, 2, \dots, N\}$ 被看作是非完整数据, z 是不可观测量,即 z 可任意取值,这样对于训练集 Y ,形成 S 中的一个子流形 D ,称为数据子流形,由 η 坐标识别。对于输入样本 (x_i, y_i) ,可以得到数据子流形 D_i

$$D_i = \{W_1 | P(r(y_i, z(W_1))) | x_i; \eta(W_1)\} \quad (3.77)$$

同样可以引入产生模型子流形 M ,

$$M_i = \{W_2 | P(r(y_i, z) | x_i; \theta(W_2))\} \quad (3.78)$$

从信息几何的观点来看,神经网络的学习算法的优化目标是减小 D_i 和 M_i 的 Kullback-Leibler 差异度量

$$K_i(D_i, M_i) = \text{Min}_{W_1 \in D_i, W_2 \in M_i} K_i(W_1, W_2) \quad (3.79)$$

在训练样本集 $\{(x_i, y_i), i=1, 2, \dots, T\}$ 的情况下,学习被描述为优化下式的过程:

$$K(D, M) = \sum_{i=1}^T k_i(D_i, M_i) \quad (3.80)$$

D 和 M 的 Kullback-Leibler 熵减小要通过对偶校正学习算法实现。

对偶校正学习算法 DCL 有如下步骤。

- (1) 初始化: 取神经网络初值,对 W_2 进行随机的初始化赋值为 W_2^0 ,从而确定 $P^0 \in M$ 。
- (2) 从第 i 步起, $i=0, 1, \dots$,重复以下的过程。
 - ① E 步: 沿 e -测地线投影 P_i 到 D ,得到投影点 $Q_i \in D$,减小 Kullback-Leibler 差异度量 $K(Q, P_i), Q \in D$ 。
 - ② M 步: 沿 m -测地线投影 Q_i 到 M ,得到投影点 $P_{i+1} \in M$,减小 Kullback-Leibler 差异度量 $K(Q_i, P), P \in M$ 。

3. 基于拓扑结构逼近的学习理论

在神经场理论的框架下基于拓扑结构逼近的学习算法理论,基本的思想可表示为人的感知系统模型,用隐变量来估计和发现输入模式的主要结构,形成输入模式相关的问题结构表示,通过这个结构表示来指出相关的结构估计,表示空间的结构是一个中间结构,它有助于得到期望的结果。这一部分,我们发展了基于代数拓扑概念的结构表示,用拓扑结构表示作为人工神经网络结构抽取表示的中间过程,对于未知系统的逼近密切关系着表示空间中的中间过程。这个表示空间与PCA、LVQ、SOM和MDL的表示机理不同,也不同于Helmhotz机的描述,它突出由代数拓扑中单、复形结构的组织空间作为表示空间,对于神经网络数值结构的表示和编码机理。神经网络变换机理包括两个变过程,一是抽取输入的拓扑表示,形成神经场表示;另一个是在表示下期望值的输出。与之相适应的学习过程也包括两个方面,抽取输入模式特征和最佳结构的估计以满足期望的输出。

我们把人的感知系统看作为统计学习机,用流形上的拓扑概念来发展学习理论,一个输入为 x 输出为 y 的未知目标系统,我们用 $q(y|x)$ 来表示,一个基本的问题就是怎样来描述目标系统的结构问题,把它嵌入到一个什么样的信息处理结构中,与问题相关的系统的结构我们用统计流形学习机

$$S = \{P(y, z | x, \Theta)\} \quad (3.81)$$

来进行表示,这里 x, y 是可观测数据, z 是辅助数据量或是不可观测数据,也就是在非完整信息处理的框架下,对偶统计流形作为系统的结构描述方法。作为对偶流形,可以在 S 上引入对偶坐标 θ 和 η ,即用两种坐标来编码系统,来对系统进行识别; S 可表示为

$$S = \{\theta | P(y, z | x; \theta)\} = \{\eta | P(y, z | x; \eta)\} \quad (3.82)$$

$P(y, z | x, \Theta)$ 可由对偶坐标 θ 和 η 来进行表示和识别。

在神经场理论中,采用代数拓扑中的单复形结构来表示空间。单复形结构是模型结构嵌入最直观的拓扑概念,为研究简单模型向复杂模型的嵌入给出了非常直观的形象。在非完整数据信息处理的框架下,假设训练样本集 $(x_i, y_i), i = \{1, 2, \dots, T\}$ 所载信息量表示为 $t(x, y, z)$, t 是 (x, y, z) 的函数,由于 z 为辅助量或不可观测量,当一个辅助量 z 确定,就可看作为一个完整信息,从而决定一个系统,用 η 坐标来识别系统,当 z 变化时,形成一个数据子流形,为

$$D = \{\eta | P(t(x, y, z) | x; \eta)\} \quad (3.83)$$

D 是 S 的一个子流形,下面为了研究对于输入信息量结构的描述,我们引进下述的 D 的拓扑结构编码表示,即 D 的表示空间为 V 。

$$V = \left\{ \eta \in S \mid \eta = \sum_{i=0}^r c_i \eta_{q_i}, \sum_{i=0}^r c_i = 1, c_i \geq 0 \right\} \quad (3.84)$$

这里, q_i 是 S 上 r 个点。实际上, V 是由流形 S 上 r 个数据点张成的平坦流形, D 是 V 的弯曲子流形, V 是 η 平坦子流形结构;从拓扑的结构看, D 和 V 有相同的拓扑结构。引入了表示空间 V ,则数据子流形的拓扑结构可由 V 的结构进行表示。数据子流形 D 作为 V 的弯曲子流形嵌入 V 中。同样也可以进一步地用复形结构表示输入模式的结构。表示空间 V 称为复形结构编码的表示空间。

可分子流形结构编码:假设在 η 坐标编码中,可观测变量与隐变量是可分离的,则数据

子流形 D 的拓扑结构被编码成为 V

$$V = \{\eta = (\eta_v, \eta_p) \mid P(t(x, y, z) \mid x; \eta)\} \quad (3.85)$$

这是最为简单的结构编码。

和数据子流形的结构编码一样,神经网络模型产生结构可以由 θ 坐标进行表示,形成神经网络子流形 M , M 的拓扑结构也可以编码成 S 上的 θ -复形结构, M 的拓扑结构编码为

$$N = \left\{ \theta \in S \mid \theta = \sum_{i=0}^r c_i \theta_{pi}, \sum_{i=0}^r c_i = 1, c_i \geq 0 \right\} \quad (3.86)$$

N 是 S 上的 θ -平坦子流形结构,而 M 是 S 的弯曲子流形。 N 可以看作为 M 的主要结构, N 和 M 的相同的拓扑不变结构。从拓扑不变量的角度看, M 和 N 是同伦不变的,即 M 可以通过同伦变换收缩到 N 。 N 实际上也是对于 M 结构定性描述。

数据子流形是对于输入数据模型结构的描述,是对认知模型结构的认识,由神经网络的认知权值结构作用而产生;而神经网络模型是对产生式结构的描述,由神经网络产生模型而形成,由产生权值作用而产生。以上两个方面可以看作为同一个过程认识的多集模式表示,即一个问题由两种方式表示,从形式上说,认知模型表示为相关系统 S 子流形 D 和对应的拓扑结构编码 V :

$$D = \{W_1 \mid P(y, z(W_1) \mid x; \eta(W_1))\} \quad (3.87)$$

$$V = \left\{ \eta \in S \mid \eta = \sum_{i=0}^r c_i \eta_{qi}, \sum_{i=0}^r c_i = 1, c_i \geq 0 \right\} \quad (3.88)$$

产生模型表示为 S 的子流形 M 和对应的拓扑结构编码 N 。

$$D = \{W_2 \mid P(y, z \mid x; \theta(W_2))\} \quad (3.89)$$

$$V = \left\{ \eta \in S \mid \eta = \sum_{i=0}^s c_i \eta_{pi}, \sum_{i=0}^s c_i = 1, c_i \geq 0 \right\} \quad (3.90)$$

作为同一个问题的不同表示方法,学习的问题就是两种表示的耦合问题,两种表示拓扑结构的定性逼近和在此基础指导下的定量分析逼近。

我们看到对于数据子流形 D ,可以形成 D 的拓扑结构表示 V , V 是 D 的拓扑结构编码;同样对于产生式神经网络子流形 M ,也有一个单复形结构拓扑编码 N 作为其结构的描述。学习的策略包括两个方面,一个是使数据子流形的拓扑结构 V ,与 N 的结构尽可能一致,即通过拓扑特征映射 T ,减小 V 与 N 间的差异

$$T: V \rightarrow N$$

这个拓扑变换,将实现拓扑特征抽取,把神经场的表示变换到神经网络产生式模型的决策结构中,实现结构的逼近;另一个过程是对偶校正学习算法,在拓扑校正后,通过两个表示间交替地减小 Kullback-Leiber 差异度量,即通过对偶学习算法,减小 $T(D)$ 与 M 间的 Kullback-Leibler 差异度量。更具体地说,基于拓扑结构逼近的学习算法 TAC 包括如下两个过程。

(1) TAC 拓扑变换:在前面看到,拓扑特征变换可由变换算子实现或由算法实现。在结构编码的情况下,我们用变换算子来实现拓扑变换。假设 V 和 N 的顶点满足关系 $K(q_1, p_1) \leq K(q_2, p_2) \leq \dots \leq K(q_r, p_r)$ 且 $r \leq s$,则

$$T: V \rightarrow N, \quad \eta = \sum_{i=0}^r c_i \eta_{qi} \rightarrow \sum_{i=0}^r c_i \theta_{pi} \quad (3.91)$$

这里, K 是 Kullback-Leibler 差异度量

$$E_{\text{topo}}(V, N) = \sum_{i=0}^r K(q_i, p_i) \quad (3.92)$$

称为拓扑信息熵, 经过 T 变换我们看到拓扑信息熵为 0。

(2) 对偶校正学习算法(DCL): 经过拓扑校正映射 T 的作用后, 形成了 D 与 M 间拓扑结构的校正, 即 $T(D)$ 与 M 间有相同的拓扑结构; 下一步, 即是在相同的拓扑指导下, 运用信息几何的学习理论(对偶校正学习算法或 em 学习算法)交替减小 $T(D)$ 与 M 间的 Kullback-Leibler 差异度量, 学习的优化目标为

$$K(W_1, W_2) = \text{Min}K(T(D), M) = \text{Min}_{Q \in T(D(W_1)), P \in M(W_2)} K(Q, P) \quad (3.93)$$

上述过程可用信息几何的 em 学习算法实现, 描述为

① e -投影步: 固定 W_2 , 得到 P , e -投影 P 到 $T(D)$, 从而得到 $T(Q(W_2)) \in T(D)$, 找出 W_1 , 减小 $K(W_1, W_2)$ 。

② m -投影步: 固定 W_1 , 得到 Q , m -投影 $T(Q)$ 到 M , 从而得到 $P \in M(W_2)$, 找出 W_2 减小 $K(W_1, W_2)$ 。

3.9 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, 是因为它有很多优点: 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。但另一方面缺乏快速学习方法, 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)^[335], 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 3.21 所示。

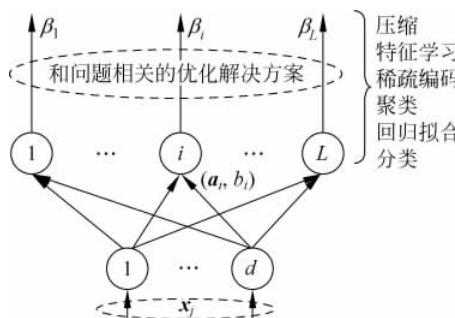


图 3.21 超限学习机的结构

图 3.21 中, 超限学习机的网路输出:

$$f_L(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, \mathbf{x}) \quad (3.94)$$

超限学习机的特征映射：

$$\mathbf{h}(x) = [G(\mathbf{a}_1, b_1, \mathbf{x}), \dots, G(\mathbf{a}_L, b_L, \mathbf{x})] \quad (3.95)$$

隐节点的输出函数

$$\begin{aligned} \text{Sigmoid: } \quad G(\mathbf{a}_i, b_i, \mathbf{x}) &= g(\mathbf{a}_i \cdot \mathbf{x} + b_i) \\ \text{RBF: } \quad G(\mathbf{a}_i, b_i, \mathbf{x}) &= g(b_i \| \mathbf{x} - \mathbf{a}_i \|) \\ \text{Fourier Series: } \quad G(\mathbf{a}_i, b_i, \mathbf{x}) &= \cos(\mathbf{a}_i \cdot \mathbf{x} + b_i) \end{aligned} \quad (3.96)$$

3.10 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,以及其他联合皮层当中都存在功能柱结构。这些结果表明,功能柱是皮层中一个普遍的结构,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^[517]。

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)等人模拟视皮层中发现的同步振荡现象^[664];詹森(B H Jansen)等人提出了耦合功能柱模型产生了类 EEG 波形和诱发电位^[342];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^[250]。另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)等人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)等人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。

李速等采用模型神经元作为基本单元,按皮层功能结构组织起来的功能柱模型,探索这些发放模式与外界输入和网络结构的关系,研究多个功能柱联结起来组成的网络模型,在活动模式上有什么新的特点^[879]。

3.10.1 模型与方法

1. 神经元模型

希望从发放的单神经元出发来构建功能柱模型,因此选用了 Rose-Hindmarsh 方程来描述单神经元:

$$\begin{cases} \dot{x} = y + ax^3 - bx^2 - z + I_{\text{syn}} + I_{\text{stim}} \\ \dot{y} = c - dx^2 - y \\ \dot{z} = r[s(x - x_0) - z] \end{cases} \quad (3.97)$$

其中, x 代表膜电位, y 表示快速回复电流, z 描述慢变化的调整电流, I_{syn} 表示突触电流, I_{stim} 表示外界输入, a, b, c, d, r, s, x_0 均为常数,在这里取值为 $a=1, b=3, c=1, d=5, s=2,$

$x_0 = -1.6$ 。在 Rose-Hindmarsh 模型中的时间尺度为 5 单位相当于 1ms。

根据生理试验的结果,皮层功能柱中的神经元按生理特性分主要有两类:规则发放型细胞(Regular-Spiking,RS)和快速发放型细胞(Fast-Spiking,FS)。RS 细胞是兴奋型的,在形态上均为锥体细胞,它的特性是明显而快速的发放频率适应性,对于持续电流刺激下随着时间延长发放频率快速降低。FS 细胞是抑制型的,在形态上通常为非锥体细胞,它的生理特性是对持续电流输入的频率适应性较低。我们采用不同的参数 r 来表现这两种细胞的特征: $r_{\text{RS}} = 0.015, r_{\text{FS}} = 0.001$ 。在图 3.22(a)当中画出了两种神经元的发放图样和频率-时间曲线。

2. 突触模型

模型采用基于电流的突触模型,在突触前细胞的每个动作电位都将触发突触后细胞的 I_{syn} 输入。突触电流 I_{syn} 表示为:

$$I_{\text{syn}} = g_{\text{syn}} V_{\text{syn}} (e^{-t/\tau_1} - e^{-t/\tau_2}) \quad (3.98)$$

其中 g_{syn} 为膜电导, τ_1 和 τ_2 是时间常数, V_{syn} 表示突触后电位。用 V_{syn} 来调节突触耦合的强度。RS 细胞之间兴奋型连接的 V_{syn} 用 V_{RR} 表示, 同理, 从 RS 细胞投射到 FS 细胞、从 FS 细胞投射到 RS 细胞的 V_{syn} 分别表示为 V_{RF} 和 V_{FR} 。参数设置为 $g_{\text{RR}} = 4, \tau_{1(\text{RR})} = 3, \tau_{2(\text{RR})} = 2, g_{\text{RF}} = 8, \tau_{1(\text{RF})} = 1, \tau_{2(\text{RF})} = 0.7, g_{\text{FR}} = 4, \tau_{1(\text{FR})} = 3, \tau_{2(\text{FR})} = 2$ 。 V_{FR} 始终设为 -1。 V_{RR} 和 V_{RF} 在模拟过程中从 0.1 到 1 之间变化。几种突触中的电流和突触后电位变化见图 3.22(b)。

3. 网络结构

网络结构不同于以往模型的是,功能柱由发放的单神经元构成,再以功能柱为基本结构模块,构成更大型的网络。作为模块的功能柱,其内部结构是相对固定的,功能柱内部神经元连接也较为丰富,而功能柱与功能柱之间的连接则稀疏得多。

对功能柱中的神经元数量和类型做了简化,按生理资料,兴奋型神经元约占 80%,抑制型神经元约占 20%。我们的模型中,一个功能柱由 15 个神经元组成,12 个 RS 神经元,3 个 FS 神经元。

在功能柱中,锥体细胞的轴突上发出丰富的回返侧枝,投射到同一功能柱内其他锥体细胞,形成兴奋型回路。这种连接把一个功能柱内的锥体细胞耦合成为一个相互兴奋的系统。在本文的模型中,每个 RS 细胞随机连接到其他 6 个 RS 细胞。这种 RS 到 RS 的连接(以下称为 RR)的延时是 1.2 单位时间,标准差为 2.5%。

锥体细胞的轴突侧枝也终止于抑制性中间神经元,抑制性中间神经元则在锥体细胞上形成抑制性突触,这样便形成了功能柱内的抑制性反馈回路。在我们的模型中,每个 FS 神经元接受从 5 个 RS 神经元来的突触,也就是说平均每个 RS 细胞投射到 1.25 个 FS 细胞。对每个 RS 细胞而言,它发出的纤维终止于 RS 细胞和 FS 细胞的比率符合解剖学统计结果。RF 突触连接的延迟是 0.8 ± 2.5% 时间单位。

在形态学观察中,锥体细胞具有较长的轴突,而 FS 细胞往往轴突很短。我们假定抑制性的 FS 神经元只能作功能柱内的局域投射,而只有 RS 神经元能够将轴突分支投射到其他功能柱内。功能柱之间相互连接。我们假定功能柱内部的连接较强,功能柱间则连接稀疏

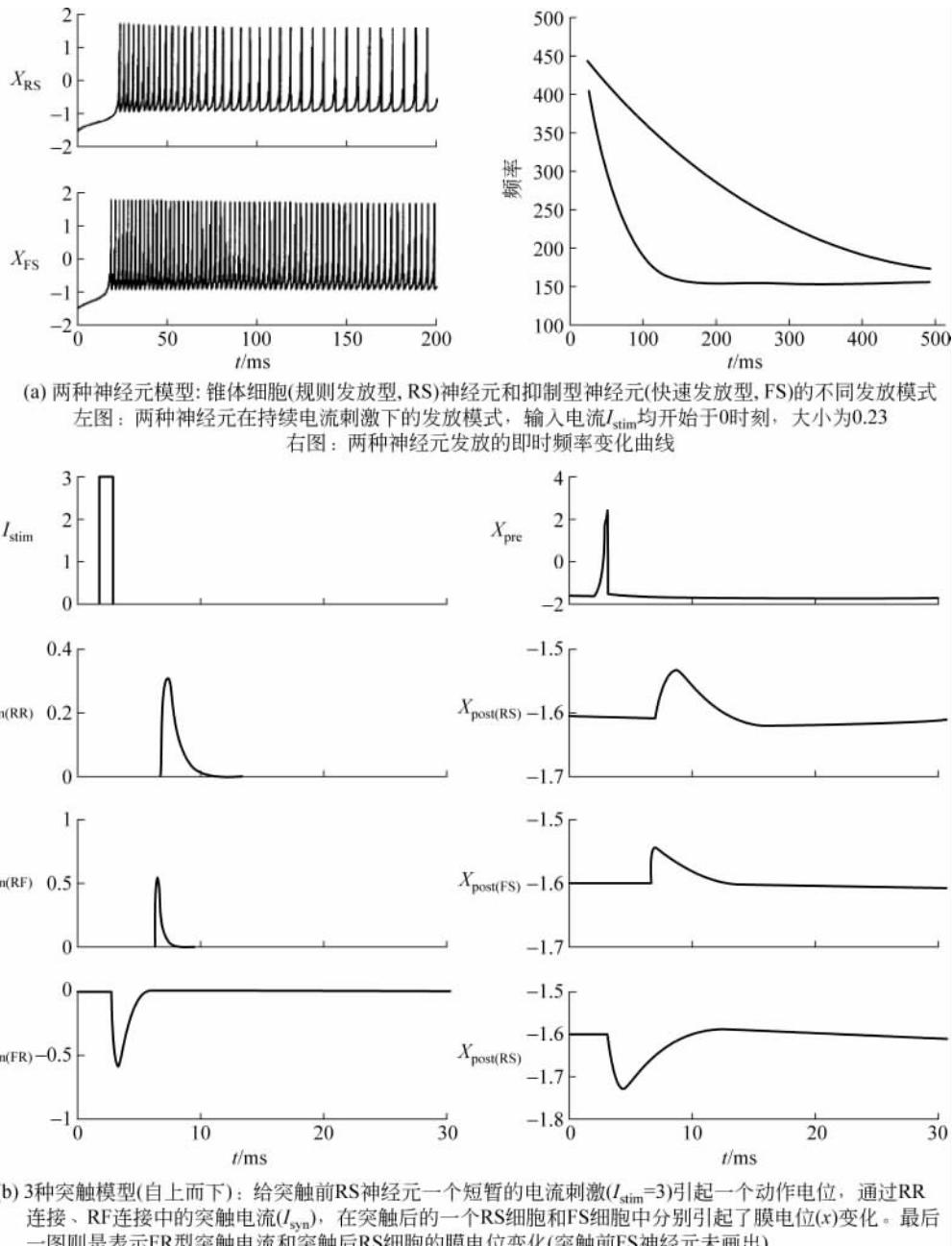


图 3.22 神经元和突触模型

而且耦合度弱。在目前的网络模型中, 我们只设定了一种功能柱之间的连接, 即从一个功能柱内的 RS 神经元连接到目标柱内的 RS 神经元, 形成功能柱之间的兴奋性连接(以下称为 iRR, 符号前标注 i 表示功能柱之间, 下同)。这种连接的数量是, 2 个突触前细胞投射到 6 个突触后细胞。可以看出, 相对于柱内连接, 柱间的连接是非常稀疏的。连接参数为 $g_{iRR}=4$, $\tau_{1(iRR)}=3$, $\tau_{2(iRR)}=2$, 延迟是 $2 \pm 2.5\%$ 时间单位。

4. 网络的输入

网络的输入是随机的脉冲输入,模拟功能柱接受的来自丘脑和皮层其他区域来的脉冲刺激。外界输入只刺激12个RS细胞,每个细胞的脉冲序列是随机的,并在各细胞间相互独立。每个脉冲都在受刺激的细胞中触发 I_{stim} 电流输入,其方程如下:

$$I_{\text{stim}} = g_{\text{stim}} V_{\text{stim}} (e^{-t/\tau_1} - e^{-t/\tau_2}) \quad (3.99)$$

其中, $g_{\text{stim}}=4$, $V_{\text{stim}}=0.08$, $\tau_1(\text{stim})=3$, $\tau_2(\text{stim})=2$ 。可以看出外界刺激的幅度比功能柱内部的突触耦合强度要小得多。

外界输入的强度定义为每个RS细胞在每个时间间隔中收到一个脉冲刺激的概率 P_{stim} 。

注意: 时间间隔和 P_{stim} 都很小,可以认为一个细胞在一个时间间隔内至多只能接受一个脉冲刺激。

这里采用了两种刺激方式,一种是恒定刺激,其中 P_{stim} 保持不变,另一种是周期性刺激, P_{stim} 定义为时间的函数:

$$P_{\text{stim}}(t) = E_{\text{stim}} + A_{\text{stim}} * \sin(\omega t + 2\pi/T) \quad (3.100)$$

E_{stim} 和 A_{stim} 分别是 P_{stim} 的均值和振幅。 T 等于5000(在模型中等于1s),因此 ω 就是脉冲概率振荡的频率。

5. 度量指标和计算方法

功能柱的局部场电位 U 定义为12个RS细胞膜电位的平均值。功能柱输出的功率普通过场电位自相关函数的傅里叶变换得到。

为衡量功能柱内神经元发放的同步程度,引入同步性 κ 这一度量,它的计算基于功能柱神经元对的相关系数。把一段长的时间 T 按分解为小片断 λ ,每一小段内的膜电位加以平均,就得到了一个新的膜电位序列 $X_i(l), l=1, 2, 3, \dots, n (n=T/\lambda)$ 。计算神经元 i 和 $j (1 \leq i, j \leq 12)$ 之间的相关系数 κ_{ij} :

$$\kappa_{ij} = \frac{\sum_l (x_i(l) - \bar{x}_i)(x_j(l) - \bar{x}_j)}{\sqrt{\sum_l (x_i(l) - \bar{x}_i)^2 \sum_l (x_j(l) - \bar{x}_j)^2}} \quad (3.101)$$

群体的相关性 κ 定义为每两个RS细胞间 κ_{ij} 的平均值。本文中时间片断 λ 设为1,即相当于实际的0.2ms。

为避免网络设定和输入刺激的初始条件的影响,网络输出的前200ms舍弃,所有结果都从模拟输出200ms以后开始计算。由于网络连接和刺激有随机性,以下结果均是重复3~5次的平均值。数值计算的时间步长为0.01ms。

3.10.2 单功能柱模型的模拟结果

我们首先对功能柱模型输入恒定刺激,研究功能柱内部的锥体细胞的发放的同步性和整个网络输出的节律现象。

图3.23显示了几种不同的节律输出。其中,图3.23(a)是典型的alpha节律,其场电位波形有明显的纺锤形振荡,其振幅增大减小的周期约为1.5s。场电位的能谱表明振荡频率

处于 alpha 频段 ($14 \sim 30\text{Hz}$)。图 3.23(b) 中显示的是处于 beta 频段 ($8 \sim 13\text{Hz}$) 的振荡。图 3.23(c) 中的振荡频率为 5Hz 左右的方波, 发放期较长而且振幅很大。图 3.23(d) 显示的是 40Hz 振荡。在图 3.23(e) 中, 神经元进入持续发放状态, 能谱也没有明显的峰, 这种状态下文中称为持续兴奋。以上各种节律均是由模型在不同参数条件下产生, 下文中将详细讨论不同的条件对皮层节律和同步性的影响。

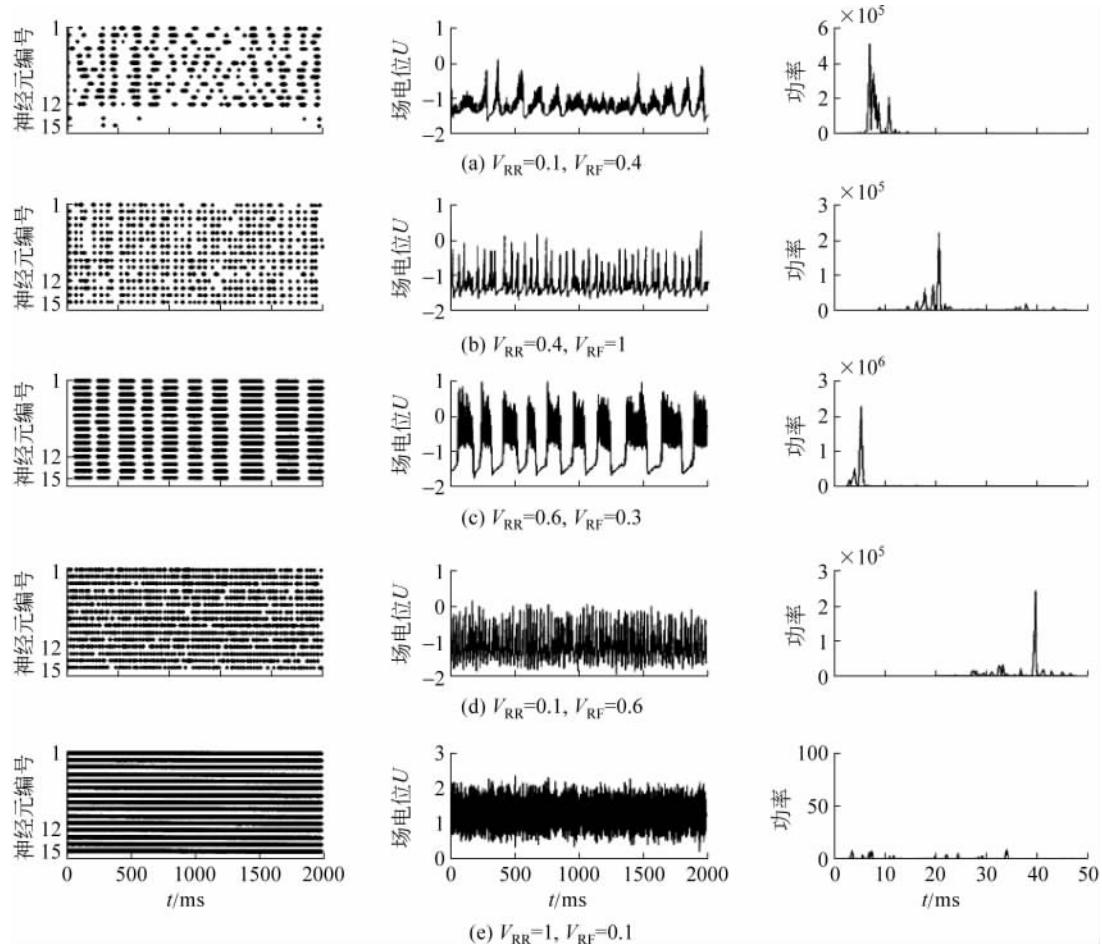


图 3.23 单功能柱模型的几种活动模式

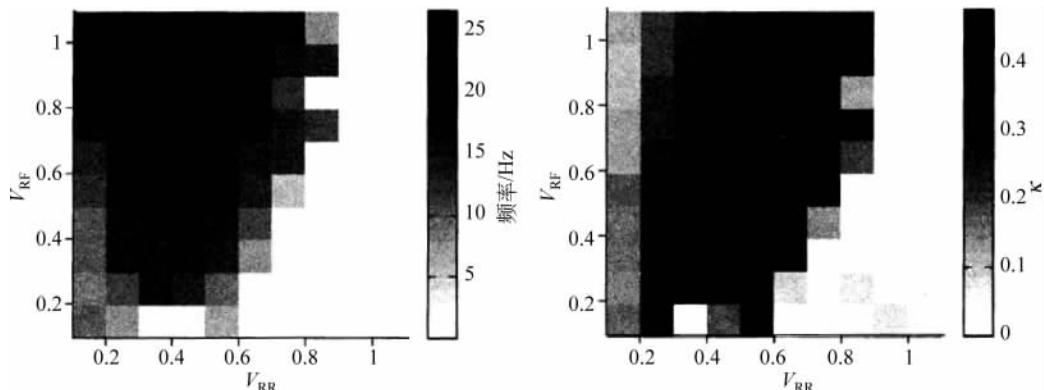
左列: 神经元发放图形。纵轴 15 个神经元中, 1~12 为 RS 细胞, 13~15 为 FS 细胞, 每个点表示一个动作电位。中列: 功能柱的场电位。右列: 场电位的功率谱, 表示功能柱的输出频率。(a) $V_{RR} = 0.1$, $V_{RF} = 0.4$; (b) $V_{RR} = 0.4$, $V_{RF} = 1$; (c) $V_{RR} = 0.6$, $V_{RF} = 0.3$; (d) $V_{RR} = 0.1$, $V_{RF} = 0.6$; (e) $V_{RR} = 1$, $V_{RF} = 0.1$ 。除(d)图中 $P_{stim} = 0.04$ 外, 其余各图 $P_{stim} = 0.025$ 。

1. 同步性与突触耦合强度的关系

为进一步研究同步振荡和突触耦合强度的关系, 在模拟过程中改变 V_{RR} 和 V_{RF} 的值。 P_{stim} 设定为 0.025 且不变。图 3.24(a) 中, 网络输出的频率表示在 V_{RR} 和 V_{RF} 构成的二维参数空间中。随着兴奋性连接强度的增大, 输出频率先增大, 再减小, 直到在图的右下角, 即兴奋型连接强度大而抑制型连接强度小的区域, 输出节律为 0, 对应于图 3.23(e) 中的持续兴奋。

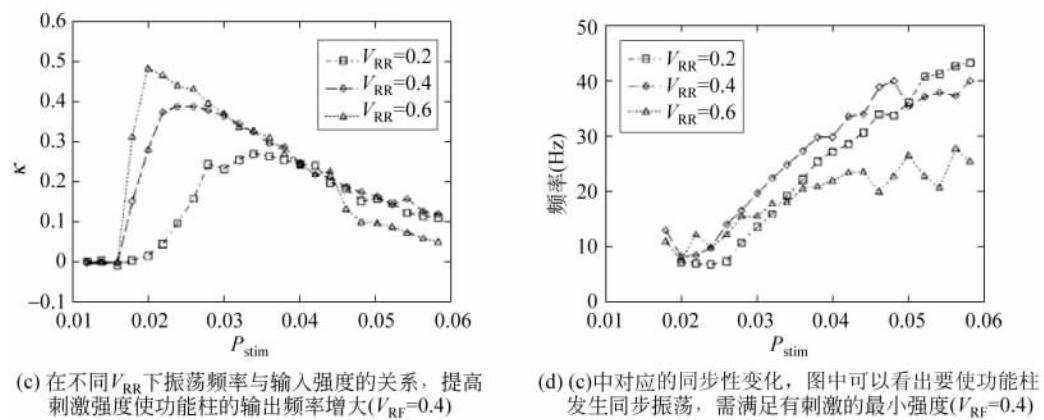
奋状态,此时网络过兴奋。而抑制性回路中耦合强度 V_{RF} 的增强则单调地增加网络的频率,并且使网络不容易陷入过度兴奋状态。在当前刺激强度下,最高频率(25Hz, beta 节律)出现在 $V_{RR}=0.4, V_{RF}=1$ 处。

图 3.24(b)显示了同步性和连接强度之间的关系。可以看到在图 3.24(a)白色区域的对应处也是白色,说明持续兴奋下的功能柱同步性几乎为 0,除此以外,同步性系数 κ 随着兴奋型连接强度的增加而增加,随着抑制性连接强度的增加而减少。同步性最强的区域总是出现在持续兴奋的边缘,此处对应图 3.23(c)中大振幅的振荡。



(a) 在不同连接强度下, 功能柱活动的振荡频率可在3~20Hz之间变化($P_{stim}=0.025$)

(b) (a)中相应功能柱活动的同步性($P_{stim}=0.025$)



(c) 在不同 V_{RR} 下振荡频率与输入强度的关系, 提高刺激强度使功能柱的输出频率增大($V_{RF}=0.4$)

(d) (c)中对应的同步性变化, 图中可以看出要使功能柱发生同步振荡, 需满足有刺激的最小强度($V_{RF}=0.4$)

图 3.24 兴奋性和抑制性回路的耦合强度及输入强度对功能柱同步振荡频率和同步性的影响

2. 同步性与刺激强度的关系

我们接着探讨网络行为与刺激强度的关系。刺激强度 P_{stim} 从 0.01 到 0.05 逐渐变化, 观察网络的周期性和同步性(见图 3.24(c)(d)), 同时在不同 V_{RR} 下刺激强度的作用。在图 3.24(c)和(d)当中都可以明显地看到, P_{stim} 必须大于一个临界值约 0.016, 才有同步振荡的出现, 而且不同的 V_{RR} 值对这个临界值几乎没有影响, 说明使功能柱进入同步状态需要一个刺激的最小强度。当 P_{stim} 超过临界值, 则同步性迅速上升, 很快达到最大值, V_{RR} 越大, 这个爬升的速度就越快, 同步性 κ 的峰值出现的就越早, 当超过峰值以后, P_{stim} 的增大均使 κ 下

降,而且不同的 V_{RR} 下曲线非常接近。说明要使网络较好地同步,需要最佳的 P_{stim} 。 P_{stim} 过大,将使模型进入持续发放状态,同步性降为 0。网络振荡的频率随着 P_{stim} 增大而单调递增,从 6Hz 变化到 43Hz,覆盖了 α 频率到 γ 频率。

3. 网络对周期性刺激的反应

给网络施加正弦刺激,以考察其行为对周期性输入的反应。 P_{stim} 由式(3.100)描述。首先我们选取了以下参数 $V_{RR} = 0.3$, $V_{RF} = 0.8$, $E_{stim} = 0.025$, $A_{stim} = 0.01$ 为不变量,当 $\omega = 0\text{Hz}$ 时,即如前面所讨论的恒定刺激,在恒定刺激下的频率称为此参数条件下的功能柱的固有频率。刺激的频率 ω 从 0 到 100Hz 间变化,功能柱的输出表现出 3 种反应模式(图 3.25(a))。在图 3.25(b)中输出的功率谱按 ω 排列,表示输出频率和输入频率之间的关系。可以看出当 $\omega < 10\text{Hz}$ 时,场电位的固有频率叠加在缓慢振荡上。在频谱中,在对角线上的峰表示输出频率中对应输入频率的成分,在 30Hz 处另有频率响应,也就是模型在最大输入 $P_{stim} = E_{stim} + A_{stim} = 0.04$ 时的固有频率。当 $\omega > 10\text{Hz}$ 且 $< 30\text{Hz}$ 时,频谱中只有输入频率可见,网络的固有频率被外界输入同步化。当 ω 大于 30Hz,场电位中又可见到两个波叠加的情况。此时频谱中除输入频率外,另有 25Hz 左右的响应,并且随着 ω 增加,该峰逐渐向低频偏移并逼近 20Hz,也就是功能柱在平均输入 P_{stim} 下的固有频率,同时频谱中输入频率的峰值也逐渐减小。

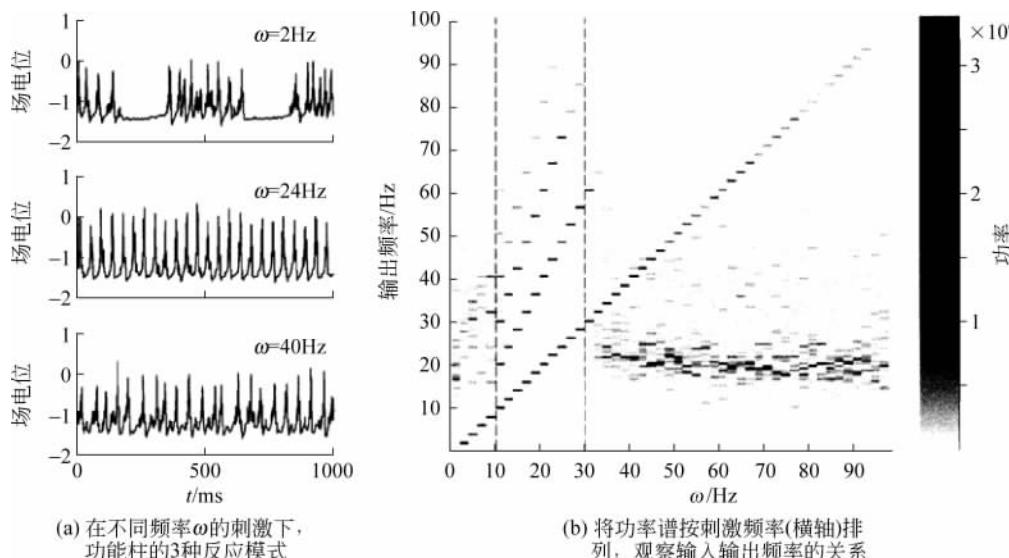


图 3.25 功能柱对周期性输入的反应

3.11 神经元集群的编码和解码

3.11.1 概述

脑的信息编码研究由来已久,在上世纪 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年赫布提出的经典细胞群假设^[311],到1972年巴洛(Barlow)的单个神经元的编码假设^[58],以及1996年藤井(Fujii)等人提出的动态神经元的集群时空编码假设^[249],不同观点间的争论仍在继续。其中重要的一个问题是: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,主要是由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,就必须将信息分配于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^[625]。集群编码的一个关键特性在于它的鲁棒性和可塑性,由于信息的编码在许多神经元共同活动的基础上得以完成,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储存的形成等;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是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无疑的。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,基于这种广泛的共识,产生了大量分析集群编码与解码的研究。有研究如何量化集群编码的信息,罗尔斯(E T Rolls)等人运用信息理论方法量化猕猴下颞叶可视皮层中神经元集群的编码信息,研究发现编码信息大致随着神经元数目线性增长^[629];佛朗哥(L Franco)等人研究发现编码信息不仅存在于单个神经元的放电次数中,也受神经元活动的互相关性的影响,指出由独立于刺激的神经元互相关性影响产生的基于放电率的编码信息的冗余性是很小的,同时依赖于刺激的互相关性影响对下颞叶可视皮层中的编码信息的贡献也极小^[231]。有从熵和互信息来探讨集群编码与解码的特性^[548],康(K Kang)利用统计动力学方法计算神经元集群的互信息,结果发现在离散刺激下互信息随着集群规模指数饱和,而饱和指数恰恰是不同刺激诱导的神经元响应概率中的Chernoff距离^[360];潘宁斯基

(L Paninski)探讨了集群编码中熵与互信息的非参数估计问题^[548]。从计算角度来看, Fisher 信息量是衡量编码与解码精度的常用手段和有效指标^[594]。例如阿博特(L F Abbott)和戴安(P Dayan)研究了互相关神经元的放电率差异性对集群编码精度的影响,通过对编码过程的 Fisher 信息量的计算,指出一般情况下只要集群的规模足够大的话,相关性并不限制编码精度的增长,并且讨论了某些情形下(例如加性噪声和乘性噪声)相关性怎样提高编码精度^[5]; 萨默波林斯基(Haim Sompolinsky)等通过 Fisher 信息量的计算发现,相对于不相关的神经元集群,正相关性降低神经网络的估计能力,信息容量随集群中神经元数目增长而趋于饱和,而负相关性则大大提高集群的信息容量^[733]; 托约竹米(T Toyoizumi)研究了对基于峰电位的集群解码(不考虑峰电位发生时刻)中损失的 Fisher 信息量的估计问题^[760]。

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^[831],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^[595]; 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 Knill)等人称之为贝叶斯编码假说^[374]。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^[182,183]。葛杨(Yang Ge)和蒋文新(Wenxin Jiang)探讨了采用逻辑回归混合模型的贝叶斯推断的一致性^[255]。

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,为脑的工作机理作出客观合理的数学解释。

3.11.2 熵编码理论

信息熵是随机事件的不确定性及信息量的量度,又可以反映由偏度、峰度代表的不规律性,因此适合表达非线性的神经元发放编码。熵编码方法包括:二值串方法、直接方法、渐近线无偏估计方法、神经元发放间隔方法、神经元发放对数间隔和互信息方法等。

1. 二值串方法

1952 年麦凯(Mackay)和麦克罗奇(McCulloch)首次把熵的概念用于神经编码的表达,提出了二值串熵编码,该方法的原理是将在 Δt 时间内包含 r 个发放的发放序列离散成多个带宽为 ΔT 的窄带,在每个窄带中根据有无发放分别用“1”和“0”表示编码。这种“1”和“0”序列的熵表达式为:

$$S[c] = \log_2 (^n C_r) \quad (3.102)$$

式中

$$^n C_r = \frac{n!}{r!(n-r)!} \quad (3.103)$$

n 为二值串的长度,即窄带 ΔT 的个数; r 为二值串中“1”的个数。

一般二值串都很长,难于计算其长度,所以常用近似公式。当 $n, r, (n-r)$ 都很大,且 $r \ll n$ 时式(3.102)可以转化为:

$$S[c \mid M, \Delta\tau] \approx \log_2 \frac{e}{M\Delta\tau} \quad (3.104)$$

式(3.104)中, M 为平均发放频率; $\Delta\tau$ 为窄带的带宽; e 为自然对数。

二值串方法可以近似计算神经元发放序列的熵,但该方法存在“带宽问题”的局限性,即对于不同的发放序列,离散的带宽 $\Delta\tau$ 并没有明确的标准来确定,因此导致计算的熵值可能与实际不符^[621]。

2. 直接熵编码

1997 年普林斯顿大学的史帝文柯(Steveninck)等人提出了直接熵编码方法^[779]。该编码方法的思路是基于在一个刺激前后的特定时间段内发放的时间之间的联系。编码表达的形式为:将发放序列离散成多个带宽为 Δt 的窄带,在每个窄带中根据有无发放分别用“1”和“0”表示编码,在某一个设定的常数 L 下,由 L 个相邻窄带编码的不同组合表达不同的“词”,再根据不同的“词”在整个发放序列中的概率分布计算出发放序列的熵

$$S = -\frac{1}{L\Delta\tau} \sum_{i=1}^n p(w_i) \log_2 p(w_i) \quad (3.105)$$

式(3.105)中, L 为设定的常数; n 为发放序列中包含的“词”的个数; w 表示长度为 L 的第 i 个“词”; $p(w_i)$ 为第 i 个“词”在发放序列中出现的概率。

直接方法的优点是熵与窄带分布和刺激特性无关,但是也同样有如在二值串方法中存在的“带宽问题”;同时还存在“数据采样问题”,即为了准确估计“词”出现的概率,需要在实验中采集大量的数据,这常常难以实现。

史帝文柯等人于 1997 年对苍蝇视觉中枢的运动神经元施加恒定刺激并记录神经元的响应,再应用直接方法用熵定量计算了神经元发放序列的可重复性和变异性,从而研究了在某类特定刺激下神经信息编码的方式。

3. 渐近线无偏估计熵编码

2002 年维克特(J Victor)提出了熵编码的渐近线无偏估计方法。方法的基本原理是将神经元发放序列“嵌入”一个矢量空间集,再从欧几里得空间中的有限数据集中估计出此发放序列连续分布的熵^[780]。

具体表达编码的方法是:将神经元发放序列携带的信息分成两个分布:一个是关于发放计数的分布,即在 Δt 时间内发放个数的分布;另一个是关于发放发生时间的分布。因此可以用离散为窄带的方法来表达发放计数携带的信息编码,用非窄带方法估计发放时间携带的信息编码。该方法使非线性发放序列线性、连续地嵌入矢量空间集,避免了“带宽问题”,但是目前该方法的编码结果和实际生理机制相差甚远,与实际应用尚有距离。

4. 神经元发放间隔的熵编码

神经元发放间隔熵编码的主要思路是应用神经元发放脉冲间隔(Inter-Spike Interval,

ISI)来表达编码。首先用发放间隔图(Inter-Spike Interval Histogram, ISIH)来表示 ISI 的分布情况,再找出合适的间隔概率密度函数 $f_w(w|I)$ 来拟合 ISIH,再由 $f_w(w|I)$ 计算出 ISI 的熵:

$$S[w] = - \int_{S_w} f_w(w|I) \log_e f_w(w|I) dw \quad (3.106)$$

式(3.106)中, w 表示 ISI 序列; $f_w(w|I)$ 为一个合适的条件概率密度函数,例如可以选择 gamma 函数、正态函数或高斯概率密度函数等^[468]; S_w 是 $f_w(w|I)$ 的定义域。

ISI 熵编码方法可以识别出被“直接方法”所忽略的信息。但是,因为 ISI 的均值就是发放频率的倒数,所以 ISI 熵编码方法在本质上还是一种基于频率的编码方法。目前,应用 ISI 的熵来表达编码已应用于很多神经元和细胞发放序列的分析中,例如 1996 年马修(P. Matthews)用 ISI 熵编码方法分析了人类多种不同肌肉的运动神经元的发放编码,用以解释突触噪声对刺激响应的影响; 2001 年凌(G Leng)等将它用于解释神经元对持续增强的突触输入的抑制效应^[413]; 2004 年魏特摩(D Wetmore)和贝克(S Baker)对猴运动皮质细胞发放编码应用了 ISI 的熵编码方法,得出了猴运动皮质的单个神经元在发放动作电位 30ms 后再次发放的概率逐渐增加的结论^[808]等等。该方法存在的问题是,选择 ISIH 的拟合函数比较困难,拟合精度有待提高。

5. 神经元发放对数间隔的熵编码

神经元发放对数间隔的熵编码在 ISI 熵编码的基础上,用 ISI 的对数值代替 ISI 值作为 ISIH 的横坐标,构造出对数发放间隔图,再用合适的间隔概率密度函数 $f_w(w|I)$ 来拟合对数发放间隔图。布姆勃拉(G Bhumbra)等人提出用双峰高斯函数很好地拟合了多种神经元发放序列的对数间隔^[75]。

如果 x 的单位是 $\log_e(\text{ms})$,则双峰函数 $f(x)$ 可以用加权系数 c ,两个均值 μ_1 和 μ_2 ,两个标准差 σ_1 和 σ_2 ,包含以上 5 个参数的拟合函数如下:

$$f(x) = \frac{c}{\sigma_1 \sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} + \frac{1-c}{\sigma_2 \sqrt{2\pi}} e^{-(x-\mu_2)^2/2\sigma_2^2} \quad (3.107)$$

应用 Levenberg-Marquardt 迭代算法来优化以上 5 个参数,再用 K-S 检验参数 D 来衡量数据和双峰函数的符合程度。式中 D 代表的是期望累积密度和实际累积概率之间的最大差距,期望累积密度函数 $f(x_i \leq x)$ 的表达式为:

$$f(x_i \leq x) = \int_{-\infty}^{x_i} f(x) dx \quad (3.108)$$

再将 $f(x)$ 代入式(3.106)计算出对数间隔熵。

对数间隔图的局限性是,它不能表达间隔发生次序的信息,因为这些信息可能是很重要的,所以进一步用相邻间隔的互信息来衡量一个发放前后两个间隔之间的信息。计算互信息的过程如下。

- (1) 用对数间隔 Y 及其前驱 x 画出对数间隔散点图。
- (2) 用一个标准差为对数间隔标准差六分之一的二维高斯核与散点图进行卷积来平滑数据。
- (3) 构造联合对数间隔图。
- (4) 得出相邻间隔的联合概率质量分布 $P(x_i, y_i)$ 。

(5) 由 $P(x_i, y_i)$ 计算出前驱间隔 $P(x_i)$ 及后驱间隔 $P(y_i)$ 的边缘概率分布:

$$P(x_i) = \sum_{j=1}^{N_y} P(x_i, y_j) \quad (3.109)$$

$$P(y_i) = \sum_{i=1}^{N_x} P(x_i, y_i) \quad (3.110)$$

式中 N_y 和 N_x 分别是 $P(y_i)$ 和 $P(x_i)$ 的窄带数。 $P(x_i)$ 的熵 $S(X)$ 可表示为:

$$S(X) = -\sum_{i=1}^{N_y} P(x_i) \log_2 P(x_i) \quad (3.111)$$

(6) 联合概率质量分布 $P(x_i, y_i)$ 的联合熵表示为:

$$S(X, Y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} P(x_i, y_j) \log_2 P(x_i, y_j) \quad (3.112)$$

相对熵 $D(X, Y \| XY)$ 表示为:

$$D(X, Y \| XY) = S(X) + S(Y) - S(X, Y) \quad (3.113)$$

相对熵 $D(X, Y \| XY)$ 是相邻间隔互信息的近似值。但是此近似值很可能偏高, 所以需要引入随机化方法, 通过随机和重列进行校正。并可以用 Monte-Carlo 法验证互信息是不是显著的大于零。

2001 年菲尔霍(D Fairhall)将发放对数间隔分析方法应用于苍蝇视觉系统对快速变化的刺激的响应分析中, 揭示了短发放间隔在神经信息表达中的重要性^[216]。2004 年布姆勃拉应用发放对数间隔分析方法对鼠视上核神经元核周细胞、抗利尿激素细胞和催产素细胞的发放模式进行了定量比较^[75]。该方法也存在对数发放间隔图很难精确拟合的局限性。

上述熵编码方法, 二值串熵编码存在对于不同的发放序列, 离散的带宽 Δf 没有明确标准, 因此导致计算的熵值可能与实际不符的局限性; 直接熵编码存在为了准确估计“词”出现的概率, 需要在实验中大量采集数据, 又常常难以实现的局限性; 渐近线无偏估计熵编码存在与实际应用距离较远的局限性。与以上四种方法相比, 应用神经元放电序列对数间隔熵和互信息熵编码, 通过熵值与互信息的结合, 可以更好地表达神经元放电的编码。熵编码研究的发展动态之一是结合对数间隔熵和互信息熵, 更好地揭示神经元放电编码的机制。

3.11.3 贝叶斯集群编码

在给定的随机模型下, 编码与解码通过贝叶斯法则相互联系。这里 r 表征单个神经元或集群神经元在刺激条件下的响应, 刺激用相位参量 θ 表示。 n 个神经元的响应记为 $r = (r_1, r_2, \dots, r_n)$, 这里 $r_i, i=1, 2, \dots, N$ 表示用峰电位计数的第 i 个神经元的放电率, 当然除了放电率, 还有其他描述神经元响应的参量。下面我们引入描述 N 神经元活动的概率函数。

$P(\theta)$: 用参量 θ 表示刺激的概率, 它经常被称为先验概率或先验知识。

$P(r)$: 实验中记录到的响应 r 的概率。

$P(r, \theta)$: 刺激 θ 与记录到的响应 r 的联合分布概率。

$P(r|\theta)$: 由刺激 θ 激发响应 r 的条件概率。

$P(\theta|r)$: 在响应 r 被记录的条件下刺激为 θ 的条件概率。

这里需要注意的是, $P(r|\theta)$ 是在刺激取值 θ 时观察到放电率为 r 的概率, 而 $P(r)$ 表示放电率为 r 时的概率, 它不依赖于刺激取的某个特定值, 由此 $P(r)$ 可以用 $P(r|\theta)$ 在所有刺激值的概率的权重之和表示, 即

$$P(r) = \sum_{\theta} P(r | \theta) P(\theta) \quad (3.114)$$

同理有

$$P(\theta) = \sum_r P(\theta | r) P(r) \quad (3.115)$$

由条件概率的定义我们知道关于刺激 θ 与响应 r 的联合概率表示为

$$P(r, \theta) = P(r | \theta) P(\theta) = P(\theta | r) P(r) \quad (3.116)$$

由此有

$$P(\theta | r) = \frac{P(r | \theta) P(\theta)}{P(r)} \quad (3.117)$$

这里假定 $P(r) \neq 0$, 此即为从 $P(\theta|r)$ 到 $P(r|\theta)$ 的贝叶斯推理理论。编码可以通过一组关于所有刺激与响应的概率 $P(r|\theta)$ 描述, 另一方面, 解码某个响应, 相当于获取概率 $P(\theta|r)$ 。由贝叶斯理论, $P(\theta|r)$ 可由 $P(r|\theta)$ 获取, 但却需要刺激概率 $P(\theta)$ 。因此解码需要用到实验中或自然发生的刺激的统计性质的知识。

3.11.4 贝叶斯集群解码

贝叶斯方法在解码问题上的好处在于促使我们作出明确的假设, 在处理似然度时我们可通过先引入一些简单的假设, 然后研究怎样使假设越来越实际。这里先给出刺激 θ 下 n 个神经元放电率分别为 r_i 的联合似然度为 $P(r_1, r_2, \dots, r_n | \theta)$ 。如果我们知道刺激 θ , 并且单个神经元的放电率依赖于 θ 且独立于其他神经元的放电率, 则假定不同神经元放电率在刺激 θ 下是条件独立的, 则有

$$P(r_1, r_2, \dots, r_n | \theta) = \prod_{i=1}^n P(r_i | \theta) \quad (3.118)$$

这个假定意味着所有神经元放电率的联合似然度就等于各自似然度的乘积。

从数学意义上讲, 编码过程可用在给定刺激 θ 下的条件概率 $P(r|\theta)$ 描述, 而解码就是从观察到的集群活动 r 中推断刺激 θ 的值。文献[835]中指出大多数的解码方法可以用极大似然推断 ML (Maximum Likelihood Inference) 或最大化后验估计 MAP (Maximize a Posteriori) 来系统的表述, 具体通过选择适当的似然函数和关于刺激的先验分布来实现。它们可以归纳为如下的一致方法。

由贝叶斯法则, 在给定 r 的条件下关于刺激 θ 的后验分布 $P(\theta|r)$ 为

$$P(\theta | r) = \frac{P(r | \theta) P(\theta)}{P(r)} \quad (3.119)$$

这里 $P(r|\theta)$ 为似然函数, $P(\theta)$ 为关于 θ 的分布, 表征先验知识, $P(r)$ 为标准化因子。MAP 就是通过对后验分布的对数 $\ln P(r|\theta)$ 的最大化来实现对刺激的估计, 即

$$\hat{\theta} = \operatorname{argmax}_{\theta} \ln P(\theta | r) = \operatorname{argmax}_{\theta} [\ln P(r | \theta) + \ln P(\theta)] \quad (3.120)$$

当先验知识 $P(\theta)$ 未知或平直的, MAP 还原为 ML

$$\hat{\theta} = \operatorname{argmax}_{\theta} \ln P(r | \theta) \quad (3.121)$$

注意在解码阶段,由估计者假定的 $P(r|\theta)$ 要是等于实际的模型 $Q(r|\theta)$,我们就称该估计者运用了一个可置信模型;当 $P(r|\theta)$ 不等于 $Q(r|\theta)$,则称为非置信模型。运用非置信模型主要基于两个原因,一是解码系统经常不知道编码系统的精确信息,因此不得不使用非置信模型,特别是对实验中的数据分析;二是通过合适的简化的非置信模型,可以在不牺牲太多的解码精度的要求下,实现对计算成本的大幅降低。

极大似然推理类型估计量 $\hat{\theta}$ 的获得可通过最大化似然对数 $\ln P(r|\theta)$ 得到,即解方程 $\nabla \ln P(r|\hat{\theta})=0$,这里 $\nabla P(s)$ 表示 $dP(s)/ds$, $P(r|\theta)$ 称为解码模型。这里考虑三种解码模型,它们都采用极大似然推理,定义如下。

(1) 置信模型 $P_F(r|\theta)$ (Faithful Model),它用到所有的编码信息,解码模型就是真实的编码模型,即 $P_F(r|\theta)=Q(r|\theta)$,这种方法简记为 FMLI。

(2) 非置信模型 $P_U(r|\theta)$ (Unfaithful Model),它运用描述神经元响应活动的调置函数的信息,但是却忽略神经元之间的相关作用,简记为 UMLI。

(3) 矩法(也称为矢量法)模型 $P_C(r|\theta)$,它并不涉及运用编码过程的任何信息,但却对调置函数作出了粗糙的简单假定,同时忽略神经元之间的相关作用,简记为 COM。

这里需要指出的是利用非置信模型(例如 $P_U(r|\theta)$ 或 $P_C(r|\theta)$)有重要的意义。当实验研究人员从记录到的数据中重建刺激时,它们实际上运用的是非置信模型,因为真实的编码过程是无从知道的。进一步讲,真实的神经元之间的相关性经常是复杂的并且随着时间变化,导致大脑很难储存并利用所有这些信息,因此基于非置信模型的极大似然推理(忽略信息的某些方面)是解决这样的信息“灾难”的关键。