

# 第六代计算机的核心——神经网络计算机

杨熙贤 (陕西师范大学计算机科学系)

## 摘 要

神经网络计算机有可能从根本上改变计算机领域的面貌。本文阐述神经元模型和神经网络的结构与特点,指出神经网络计算机在第六代计算机中的作用。进一步讨论若干重要的神经网络模型及其技术现况和可能的应用。

现今对高速计算以及诸如语言识别、视觉与机器人等机器知觉行为的需求,是von Neumann计算机所无法满足的。近些年来,人们在前四代计算机科学与技术以及VLSI成就的基础上,试图建造第五代计算机来解决这一矛盾。可以说,第五代计算机是第四代计算机与人工智能结合的产物。另一方面,生物学家长期以来对大脑神经的研究,大大地促进了试图用大量并行的简单电子线路单元来模拟神经网络功能的设想。尤其是近两年来,由于新的网络拓朴与算法,模拟VLSI电路的实现及大规模并行性的实现,使人工神经网络的研究重新受到人们的关注。1988年日本MITI(国际贸易工业省)首先提出了称为第六代计算机计划的所谓“人类尖端科学规划(HFSP)”,后来INNS(国际神经网络学会)又将其扩充为IHFSP。这种第六代计算机可以说是一种类似于三极电子管的三体协同体系<sup>[27]</sup>,即以神经网络计算机为智能控制栅极,以光数据库机为信息源阳极,以超级计算机为数据处理阴极。这种以神经网络计算机为核心的第六代机有可能从根本上改变计算机领域的面貌。最近期间,对于神经网络模型的理论研究以及复杂的多模型系统的开发都取得了巨大的进展。本文主要讨论神经网络计算机的基本概念与若干重要的技术进展。

## 1. 神经元模型与神经网络

一个神经元或神经细胞(又称处理单元或结点)的结构组成是细胞体,其一端通过可调整的联结体突触(synapse)与多分叉神经纤维体树突(dendrite)连接(总称输入结构),而其另一端连向多分叉神经纤维体轴突(axon)(称输出结构)。人脑中共约 $10^{11}$ 个神经元,每个神经元中有 $10^4$ 个输入与输出联接体。当一个神经元的输入突触上出现刺激时,它便产生一系列脉冲传送到其轴突,在有兴奋作用的突触上脉冲速度增高;在有抑制作用的突触上,脉冲速度减低。因此输出脉冲的速度取决于输入信号的强度和突触连接的强度或加权。在简单的神经元模型中,输出脉冲的速度为输入信号及其有关连接强度的反曲函数<sup>[7,5]</sup>,即

$$u_i = f(\sum_j \gamma_j w_{ij})$$

其中 $u_i$ —输出脉冲的速度,  $\gamma_j$ —输入脉冲信号,  $w_{ij}$ —有关突触的连接强度,  $f$ —反曲函数,如 $\tanh$ 。这种模型中的神经元(处理单元)可以用电导与放大器组成的电路进行模拟<sup>[11]</sup>。

神经网络是指由许多并行运行的简单的神经元(处理单元)组成的系统,其存贮与处理能力由网络的结构,连接强度以及单个神经元所执行的处理决定。神经网络可执行如

自适应或学习这类高级任务，也可执行如由视觉任务或语言识别所感受到的数据处理这类低级任务。自适应网络可以是监督训练式，无监督训练式或自监督式的。

监督训练需要有标记的训练数据和一个外界教师。教师知道所需的正确响应，且当发生错误时便提供一个错误信号。无监督训练，有时称为自组织，使用无标记的数据且毋须外界教师。输入数据后，这种网络形成内部的类目 (categoricals) 或者簇 (clusters)，为的是压缩所需处理的输入数据量而不损失重要信息。对于自监督，当执行某种特殊任务时，由自动机从内部监督错误。例如，一种用来受控模拟眼肌以学习跟踪移动点的自动机可以通过监督模拟视网膜上像点位置与视网膜中心小凹间的距离，产生一个误差信号<sup>[18]</sup>。

神经网络中不仅有上述那种求积之和的简单结点，而且还可以含有复杂的结点。总之它们是非线性的，模拟的，速度比现代的数字电路慢，而且可能涉及暂态积分、与其它类型的时间关系以及比求和更复杂的数学运算。网络中的结点可能局部地连接到相邻结点，也可能全局地连接到所有其它结点，或者稀疏地连接到几个远程结点。此外，神经网络可以从低层向高层有外部前馈连接的层次网，或者是具有递归反馈连接的网络。

## 2. 几个重要的神经网络计算机模型

神经网络模型可分为两个交叠的类：神经生物模型类<sup>[3]</sup>与计算模型类。前者以模拟人脑的某些功能为主，而后者则主要执行某些重要的技术性功能。下面分别讨论几种有代表性的模型。

**Hopfield网络** 是一种监督训练网络。Hopfield的早期工作是把结点完全连接成网络，使它具有两个可能的输出，且结点间有对称的连接加权该网络的输出定义成能代表所有结点状态的二元矢。

该网络用做存贮信息的可寻址存贮器时<sup>[9]</sup>，它按照与待存入参照模式有关的公式，首先将连接强度赋给加权。结点输出的初始化可将输入施加于部分的参照模式所形成的模式。然后结点个别地进行异步更新，一直到网络输出稳定且网络收敛时，便得到输出信号。网络收敛后，结点输出信号一般将表示被存贮的参照模式，其中大部分模式与输入信号相同。Hopfield证明<sup>[9]</sup>，如果加权对称且结点是异步更新的，则该网络必收敛。但最近的分析表明<sup>[2,19]</sup>，当此网络用做存贮信息的可寻址存贮器时，存贮的能力有限，而较简单的前馈网络可在较少的互连下提供更大的能力。

Hopfield首先提出明确地使用计算能量函数来理解与设计非线性神经网络。他与Tank<sup>[19,29]</sup>阐明了用此概念设计模拟网络的原理，用以解决许多组合优化问题，其中包括旅行推销员问题，线性规划问题与任务分派问题等。虽然他们的工作不失为能量函数方法应用的范例，但却未能说明这样设计的网络在解决大规模组合问题时的效率。目前的结果表明，对于含多个变量的大问题，该网络的性能欠佳，不能提供优化的解。

**单层视感控器** 首先由Rosenblatt<sup>[24]</sup>提出，是多层视感控器的基本构成单位。单层视感控器可将连续值的输入信号矢量区分为两类。这种网络示于图1左上部，其单个结点计算输入信号各元素的加权，减去一阈值，并通过一个二值非线性结构传递结果。二种可能的输出信号对应于不同区分的响应，视感控器由一个超平面分隔成两个判定区，如图1第1行所示。图中只有两个输入信号，超平面呈一直线。

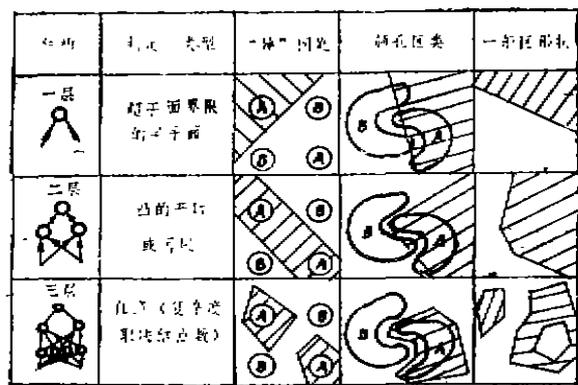


图1. 单层与多层视感控器的判定区。

视感控器中连接的加权与阈值一般用视感控器收敛过程<sup>[5,14]</sup>或LMS (最小平方平均) 算法<sup>[5,31]</sup>进

行自适应。仅在输入信号误区分时视感控器收敛过程才适用于自适应加权。可以证明,当输入类为线性可分(类成员可用单层视感控器完全区分)时这个算法收敛。LMS算法用的是半线性的非线性器,且每当实际输出与所预期的不一致时它适于自适应加权。它更适宜于类似非线性可分的情况。LMS算法是技术上最成功的神经网络算法之一,它源于Widrow与HOFF关于ADALINE网的早期工作<sup>[22]</sup>。目前广泛地用于消除电话回音与高速调制解调器<sup>[23]</sup>。

**多层视感控器**是在输入输出结点间具有一个以上结点层的前馈监督训练网络。这些附加层含有间接连接于输入输出结点的隐结点。当类不能由一个超平面分隔时就需要这种复杂结构。图1后两行示出了二层与三层视感控器的结构,其中标有A,B的平滑封闭曲线是二输入网络中输入信号对二类的分配。阴影区是单层与多层视感控器区分器造成的判定区。左边对于“异”问题所作的两类分配是互不相交的,不能由单根直线分隔。视感控器造成的A类的可能判定区由图中第1行的阴影区表示。图中第二问题的输入分配呈现筛孔,且也不能由单根直线分隔。该诸如共振峰(formant)频率一类的参数用于语言识别时就可能发生这类的饱和现象。

多层视感控器克服了单层的许多限制,但由于无可行的算法,这种视感控器过去未得到使用。最近发现所谓反传播的新训练算法<sup>[25]</sup>后情况有所改观。该算法是LMS算法的推广,因使用一种最速下降搜索技术将错误信号从输出传回较低层而得名。虽然不能证明该算法收敛,但对于很多感兴趣的问题,该算法是成功的<sup>[25]</sup>,如按音素规则形成字母<sup>[26]</sup>,区分元音与数字读音<sup>[12,20]</sup>,在语言识别器中识别音素<sup>[30]</sup>,非线性信号处理<sup>[16]</sup>等。

反向传播算法的功效在于具有二隐层的视感控器的惊人计算能力。这些网络可形成所需的任意判定区<sup>[18]</sup>,如图1底部所示,故可与传统的确定性区分器相匹敌。Kolmogorov证明了一个定理<sup>[19]</sup>,表明三层网络可形成输入信号的任何非线性连续函数。因此从理论上说,三层视感控器可以创造区分器中所需的任何连续似然函数。经验表明,利用反向传播与其它构造技术可以实现许多有用的非线性映射。

反向传播训练的困难是在判定区或所需映射变得复杂时,算法收敛常要求训练数据有多种表示形式。这在训练时期是重要的,在使用时期并不影响

网络的响应。目前已找到了可训练网络的大小的一个实际限制,但对于受限连接的单层与多层网络这不是严重问题。

**CMAC模型**即小脑模型连接控制器,它起初是作为小脑神经生物学模型研制的<sup>[4]</sup>。这种脑结构是为协调精细的运动神经动作所必需的。CMAC模型用一种类似于多层视感控器的前馈结构适当地形成连续值输入与输出间的非线性映射。它可以迅速地自适应,因为仅有一层连接体需要训练。在细心选择自适应区的情况下,自适应层输入中的连续值输入可用二值结点表示。这里的自适应区指的是使二元结点有“高”输出值的所有模拟输入变量区,可以用具有固定加权与阈值的多层视感控器形成。在有多输入变量的应用中需要用过量的二值结点来表示所有可能的输入状态。在这种情况下可以使用计算机科学中常用的杂凑技术来降低二值结点的数量<sup>[1]</sup>。二值结点的输出信号与在单层视感控器中一样,可以进行加权与求和,唯一不同的是求和结点是线性的。加权重监督训练方式并用一种类似于LMS算法的算法进行自适应。在机器人应用中,当它执行预定的活动时加权就进行自适应。

最初已证明CMAC模型能控制模拟机器人的手臂活动<sup>[4]</sup>。最近借助于自适应模拟手臂的反动力学<sup>[23]</sup>已成功地改进了商用机器人手臂的动态行为,在较慢的串行计算机上使用查表技术,用不多的加权就能实时地实现。目前正在进一步研究用于机器人的这类算法及由之派生的更高级算法<sup>[13,22]</sup>。

**特征图模型**非监督的簇化技术可用于诸如语言与图像识别这类应用,以执行数据压缩与减少区分的计算量。Kohonen<sup>[14]</sup>提出了一种顺序簇化算法,能产生一种所谓的自组织特征图。该图类似于新皮质(neocortex)中常见的那种感觉图。图(maps)是二值结点网络,其中每个结点都代表一个簇中心。Kohonen的算法保证网格中空间邻近的结点对于Euclidean意义上邻近的输入矢量有最佳响应。这个算法是k-平均簇化算法<sup>[5]</sup>的一种顺序化变体,它不需要存贮所有的训练符号并使用一种简单的加权修正规则。Kohonen算法能调整从公共输入结点到按二维网格排列的输出结点的加权,因而创造了一个矢量量化器(vector quantizer)。经训练后,一个输入信号将引起最靠近输入结点且对应于簇中心的输出结点有“高”的输出信号。

Kohonen<sup>[14]</sup>描述了许多有关此算法的例子,他也说明了此算法做矢量量化器用于语言识别器的方

法<sup>[15]</sup>。在使用前可规定所希望的簇数量而训练数据的数量又比所希望的簇数量大时,该算法就是一个可行的顺序矢量量化器。

**Darwin II 模型** 是一种自监督的复杂网络<sup>[6,7]</sup>,它模拟由许多网络组成的自动机。这个自动机的输入来自一个模拟眼睛,它在一对对抗肌控制下所看到的是一个二维输入阵列。该眼睛在其视网膜上有一个颇大的低分辨外部视野和一个小的分辨内部视野。输入信号也可以由一个多重连接的手臂提供,该手臂借助于眼肌的协调运动可以达到并“感知”呈现在输入视野中的物体。**Darwin II**借助眼肌的协调动作,经过训练去跟踪呈现在输入阵列中的物体。学习该任务的误差信号就是物体在视网膜上的位置距视网膜中心小凹间的距离。然后,**Darwin II**再用自监督训练技术,经过训练使其能够达到、触及与感知物体的边缘。最终,它可以被训练得能够用监督及视觉与触觉输入信号来区分物体。

**Darwin II**是制造成功的用以探索大脑理论的一系列模型中最复杂的一个。该理论称为“神经达尔文主义”(Neural Darwinism)<sup>[6]</sup>,它与其它神经网络设计过程不同之处在于它强调达尔文主义者所主张的选择性的重要意义,即只选择那些对特殊刺激因素有最佳响应的神经元群(groups)或子网。该理论假定称为全部技能(repertoire)的神经元群集合是在胚胎形成时形成的。由于神经元增长的随机性,全部技能中的神经元群对于交叉且类似的输入模式有最佳响应。全部技能中一个以上的群将响应每个输入模式,以便保证响应未预期的重要输入信号。训练涉及群间的竞争,它增强了特定群对特定因素的响应,并将那些群彼此结合起来,也与适当的特殊响应相互结合起来。

神经达尔文主义与普通的网络拓扑设计方法以及监督训练提供所需响应的方法有着根本的不同。它认为可以有任意多的子网络,但只有训练时所需响应的那些才能选中。此法涉及的重要问题包括响应未预期的刺激因素的必要性,区分与归纳的必要性以及与环境相互作用的重要性等问题。这种建造一个具有良定义任务、辨识与操纵外界物体的自动机的方法对于了解智能功能可能具有重大意义<sup>[10]</sup>。此外,模拟如**Darwin II**这种包含约8,000个结点与200,000个相互连接的网络的规模多模型系统之经验,对于建造更有魅力的模型可能也是重要的。

**视网膜片模型** Carver Mead<sup>[20,21]</sup>一直使用一种独特的网络系统的设计方法。他的哲学思想是用

新的模拟VLSI设计技术尽可能准确地在硅片中模拟生物系统。迄今他已制造了一个人工视网膜片<sup>[21]</sup>与一个人工耳蜗片<sup>[20]</sup>。视网膜片预定用作一个视觉系统的前端,模拟人类的视网膜。它包含一个具有对数输出的光电检测器,一个在空间上能局部平均输入信号的宽动态范围的电路和一个增强对比度暂态变化的电路。最早的视网膜片尺寸很小(48×48个象点),需要局部的连接线,包含约十万个晶体管,并经过活的输入信号验证。更大的视网膜片正在研制中。

耳蜗片预定作为语言识别系统的前端,用于模拟人类的耳蜗。它包含许多非线性带通滤波器组和自动增益控制机构并能够对射击时的同步信息在听觉神经纤维上产生脉冲。同步信息对于诸如两耳听觉定位这一类任务是重要的,对于语言识别也可能是重要的。

Carver Mead的工作可能有着重要深远意义,因为它可以制造出非常小而有力的模拟神经网络硬件。如果体积小价格低的话,它在许多甚至是商用的图像与语言识别系统中都会得到广泛应用。目前对于高级神经系统功能的了解还比较少,并且还考虑到硅所引起的限制与生物学的限制是非常不同的。

### 3. 结束语

本文所述的几种神经网络模型涉及许多新思想、新技术以及广泛的应用领域。这些模型或多或少具有基本网络的特点,而更复杂的系统是由不同神经网络组合而成的。目前,获得对模型理论的了解,设计更好的训练算法,构造模拟时间顺序行为的网络仍是许多研究者关心的论题。

电子神经网络的制造技术基本上沿着两种途径进行。一种办法是事先决定网络中的电阻,可以用一台外部计算机进行这种计算。但有时事先决定神经元之间的互连强度与位置是不可能的,所以得采用另一种可编程的突触(相当于电阻的电导)来克服这种不足。例如可使用CMOS与标准静态RAM阵列技术相结合。这种方法比较灵活,已得到广泛采用。但由于片中元件的数量很大,所以对可靠性的要求十分高。

神经网络计算机是人类所面临的最富挑战性的尖端科学领域, 学科间的密切合作将会带来更有成果的前景。

#### 参考文献

1. J. Albus, *Brain, Behavior, and Robotics*, BYTE Books, 1981
2. E. B. Baum, J. Moody, and F. Wilczek, Univ. of California (Santa Barbara) Preprint NSF-ITP-88-138, Institute for Theoretical Physics, 1987.
3. J. D. Cowan and D. H. Sharp, *Daedalus*, vol. 117, pp. 85-121, 1988
4. H. L. Dryfus and S. E. Dryfus, *Daedalus*, vol. 117, pp. 15-43, 1988
5. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. John-Wiley & Sons, New York, 1973
6. G. M. Edelman, *Neural Darwinism*. Basic Books, N.Y. 1987
7. J. J. Hopfield, *Proceedings of the National Academy of Sciences, USA*, vol. 79, pp. 2554-2558, 1982
8. J. J. Hopfield, *ibid.*, vol. 81, pp. 3088-3092, 1984
9. J. J. Hopfield and D. W. Tank, *Science*, vol. 233, pp. 625-633, 1986
10. W. Y. Huang and D. W. Lippmann, *Confer. on Neural Information Procession System-Natural and Synthesis*, IEEE, Nov. 1987.
11. L. D. Jackel, R. E. Howard, and H. P. Graf, *Proc. of 3rd International Conference on Supercomputing*, pp. 41-46, 1988
12. M. Kawato, K. Furukawa, and R. Suzuki, *Biological Cybernetics*, vol. 57 pp. 169-185, 1987
13. T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, 1984
14. T. Kohonen, K. Makasara, and T. Saramaki, *Proc. of the 7th International Conf. on Pattern Recognition*, IEEE, Aug. 1984.
15. A. Lapedes and F. Farber, *Conf. on Neural Information Procession Systems-Natural and Synthetic*, IEEE, Nov. 1987
16. R. P. Lippmann, B. Gold, and M. L. Malpass, *Technical Report TR-769*, MIT Lincoln lab, 1987
17. R. P. Lippmann, *IEEE ASSP Magazine*, vol. 4, pp. 4-22, Apr. 1987
18. R. P. Lippmann and E. A. Martin, *ICASSP 87*, pp. 705-708, Apr. 1987
19. R. P. Lippmann, *Proc. of 3rd International Conference on Supercomputing*, pp. 35-40, 1988
20. C. A. Mead, *Conf. on Neural Information Procession Systems-Natural and Synthetic*, IEEE, Nov. 1987
21. C. A. Mead, *1st International Conf. on Neural Networks*, IEEE, Jun. 1987
22. W. T. Miller, F. H. Glanz, and L. G. Kraft, *International Journal of Robotics*, vol. 6, pp. 84-98, Feb. 1987
23. G. N. Reeke and G. M. Edelman, *International Journal of Supercomputer Applications*, vol. 1, pp. 44-68, 1987
24. F. Rosenblatt, *Perceptrons and the theory of Brain Mechnes*, Spartan Books, 1962
25. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Parallel Distributed Processing*. (D. E. Rumelhart and J. L. McClelland, eds.) ch. 8, MIT Press, 1986
26. T. Segnowski and C. R. Rosenberg, *Technical Report JHU/EECS-86/01*, Johns Hopkins University, 1986
27. O. G. Selfridge, R. S. Sutton, and C. W. Anderson, *Univ. of California (Santa barbara) Preprint TR87-509.4*, GTE Laboratories Inc., Waltham, M. A., Dec. 1987
28. H. Szu, *Proc. of 3rd International Conference on Supercomputing*, pp. 47-53, 1988
29. D. W. Tank and J. J. Hopfield, *Scientific American*, vol. 257, pp. 104-114, Dec. 1987
30. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, *Technical Report TR-1-006*, ART Interpreting Telephony Research Laboratones, Japan, 1986
31. B. Widrow and M. E. Hoff, 1960 *Wescon Convention Record Part IV*, pp. 96-104, Aug. 1960
32. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice-Hall, NJ, 1985
33. G. V. Wilson and G. S. Pawley, *Biological Cybernetics*, vol. 58, pp. 63-70, 1988