

神经网络计算机系统及其现状

蔡义发 (浙江大学生物医学工程研究所)

摘要

In 1980's, studies of neural computer systems and related theories have caught world-wide attention. Scientists of computer science, cognitive science, psychology and some other disciplines are studying them from different aspects. This paper discusses the concepts of neural networks first. Then, fourteen neural network models are outlined. Two implementation strategies of neural computers and the main differences between neural computers and traditional ones have also been studied. The summarization of reported artificial neural systems and their problems end the paper.

人工神经网络(Artificial Neural Networks—ANS)的研究已超过四十年,其兴衰与人工智能的研究紧密相关。自从1982年Hopfield利用神经网络解决了TSP问题以来,神经网络再次成为一个研究热点。这方面的研究已取得了许多成果^{[1][2][3]}。由于传统的数字计算机在模式识别、符号推理、组合优化问题等方面的严重缺陷,八十年代后期人们便又重新认识到了联结机制(connectionism)的价值,并在第五代计算机、神经科学、语言学、认知科学等学科的研究结果基础上,开始了以神经网络计算机为核心的第六代计算机的研究^[4]。该研究将会在人类尖端科研方面产生重大突破,并使人类社会逐步过渡到超工业化。

1 神经网络的定义及模型

1.1 神经网络的定义

ANS是以大脑的神经网络结构为基础的。形式上说,ANS是一个以有向图为拓扑结构的动态系统。它通过对连续的和间断的输入(刺激)作出状态反馈而完成信息处理工作。网络中的节点称作处理单元(PE)。每个PE中的处理是由传递函数定义的,且依赖于输入信号的当前值和局部存储器中的值。网络中的PE或进行连续操作,或依赖一定序函数(scheduling function)来变更。典型的神经网络中,PE被划分成不同的层次。每一层上的PE具有相同的传递函数,虽然

ents and Reuse-getting more out of Your Code", IEEE Software Vol. 5, No. 2, 1988

[7] 罗晓沛, "信息系统开发的原型化方法", 研究生院, 1987.11.

[8] D. C. INCE, "Software Prototyping-Pro-

gress and Prospects", Information and Software Technology Vol. 29, No. 1, 1987.

[9] Rogers. Pressman, "Software Engineering—A Practitioner's Approach", McGraw-Hill Book Company 1982.

其局部存储器值可以不相同。

通常，我们定义三种规模的PE：

(a)小规模PE 小规模(fine-grained) PE用neuron表示。它类似于脑的海马区域的神经元。它不带有内部存储器，主要用作容错联想回忆。

(b)中规模PE 中规模(medium-grained) PE,用Neuron表示。它类似于生物传感器中作运动控制动作的神经元，带有一个内部存储器。

(c)大规模PE 大规模(large-grained) PE,用NEURON表示。它带有内部存储器，控制逻辑及通讯装置等，等同于一个个人计算机。其特征类似于祖母细胞。

这三类PE及与其有关的电路之间有四种类型的联接：兴奋性；抑制性；突发性(bursting)；分级及延迟传递。

在数学上整个网络满足：非线性；非局部性；非稳定性；和非凸性，即受限的全局优化。

1.2 神经网络模型

目前，在应用和研究中采用的至少有三十种不同类型的神经网络，其中比较有代表性的有下述十四种：

•自适应共振(ART)网——ART 1和ART 2是两种对(依靠可选参数进行粗分类的)输入数据进行分类的网络。前者用于二进制输入数据；后者用于连续数值输入^[6]。

•雪崩(AVA)网——一类用于学习、识别和重演时空模式的网络^{[6][7]}。

•逆传递(BPN)网——是一种多层映射网络，能使均方误差最小，是当今使用最广的一种网络^[8]。

•双向联想存储器(BAM)——是一类单状态互联想(heteroassociative)网络。其中有些网络有学习功能^[9]。

•Boltzmann机/Cauchy机(BCM)——是使用一噪声过程来求取代价函数的全局极小值的网络^[10]。

•BSB模型(BSB)——是具有最小均方误差的单层(single stage)自联想网络^[11]。

•小脑模型(CBT)——是一种学习并根据线索再现时空命令序列的一般模式^[12]。

•反传递(CPN)网——是一种在功能上作为统计优化自组织查询表及概率密度函数分析器的网

络^[13]。

•Hopfield(HOP)网——是一类不带有学习功能的单层自联想网络^{[14][15][16]}。

•学习矩阵(LRN)——是一种单层单向非递归联想网络^[17]。

•MADALINE(MDL)——是具有最小平方误差的一组可训练的线性组合装置^[18]。

•Neocognitron(NEO)——是一个多层结构化字符识别网络^[19]。

•Perceptron(PTR)——是一组可训练的线性分类器，目前已很少使用^[20]。

•自组织图(SOM)——形成簇与簇之间的连续映射，其映射密度随着第二个簇上给定的概率密度函数而变化^[19]。

2 神经网络计算机及其与传统计算机的区别

神经网络计算机通常是按照联接机制(connectionism)及上述神经网络模型构造起来的。从实现手段上，神经网络计算机可分成两组(参见本文第三部分)：

(a)完全实现的系统 即完全借助VLSI及光学手段用物理实体装配起来的系统。目前，AT&T、TRW、Hecht-Nielsen神经计算机公司、加州理工学院等已利用VLSI及光学手段制造了试验性的神经网络计算机。

(b)虚拟系统 即采用软件及模拟的手段在传统的计算机上实现的系统。为了加快速度，往往加上专用的软件、硬件，甚至设计全新的网络模拟机，如Transputer。

由于完全实现的系统将面临过于复杂的联接等问题，从而使得虚拟实现在目前阶段更为现实。结果是当今大部分正在使用的神经网络是用虚拟处理单元结构实现的。这种结构利用分时及多路的方式，在少量的实际PE的基础上实现大量的虚拟PE，这种方法的优点是最佳地利用现有的处理机资源。

关于神经网络计算机的设计与实现的详细讨论参见[30]。

ANS系统基本上是以神经科学家，尤其是Grossberg提出的神经信息处理原理为基础的^{[5][10][21]}。它们也变得越来越复杂。它

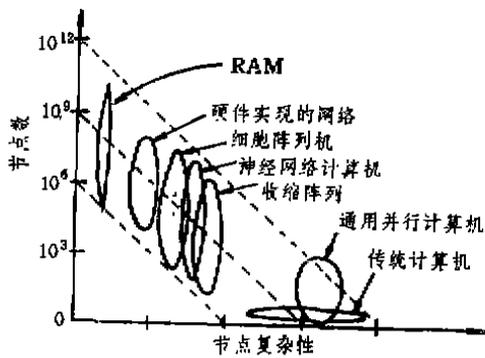


图1 神经网络计算机的结构特征
 们与数字计算机的结构关系如图 1 所示。

神经网络计算机与传统的数字计算机之

间的主要差别如表 1 所示。

3 神经网络及神经计算机系统

最近, 一些商业性的神经网络及神经计算机系统已投放市场。并且在金融、银行、国防、娱乐、汽车工业、运输、电信、安全、机器人、生产、电子学、航空以及医疗等领域开辟了新的应用。其运算速度比一些大型机快数百甚至数千倍, 而且费用只有原来的几分之一。下面我们将现今公布的神经网络及神经计算机系统分成三大类进行叙述^{[22][30]}。

表1 神经网络计算机与电子计算机的主要差别

数字化计算机	神经网络计算机
<ul style="list-style-type: none"> 以精确的数值方式处理以0和1表示的数字信息。 利用数学和逻辑函数作是/否决策。 按严格的顺序即程序处理数据, 因而整个操作总是在控制之下进行, 结果也是可预测的。 如果有足够的时间, 总希望对任何问题找到精确的答案。 搜索大型数据库以求得最佳匹配。 以特定的存储器空间存储信息以使得特定信息能被轻易地读出。 以复杂的中央处理机为核心处理数据。 以特殊的设计来实现容错。 	<ul style="list-style-type: none"> 处理连续变化的模拟信号, 比如从黑到白的各个灰度范围的信息。 在模糊、不完整甚至矛盾的信息的基础上作加权决策。 独立地通过学习和训练, 形成信息处理的方法, 结果往往是出人意料的。 对高度复杂的问题寻求快速的、好的, 但近似的答案。 搜索大型数据库以求得近似匹配。 以互联模式的形式存储信息以使得查询任何一部分信息也将回忆起所有有关的事实。 对给定的问题, 并行地集中所有简单的处理器, 即“神经元”来进行求解。 神经网络结构本身即具有容错能力。

3.1 已投放市场的商业产品

ANZA系统 (系统参数见表2)^[33]

研制者, Hecht-Nielsen神经计算机公司(加州圣地亚哥)。

主要功能特征: 它是通用的神经计算机协处理器, 用于IBM PC/AT及其兼容机以加快神经网络的执行速度, 基本的神经网络软件包括五个已定义好互联结构及传递函数的网络软件包, 用户指定其初始状态、权值、学习速率和时间限制; 配有网络描述语言AXON; 它与ANZA协处理器的交互通过用户接口子程序库UISL进行。

应用领域: 带视觉反馈的过程控制; EKG处理; 信用评价; 抵押审查; 市场分析; 阅读手写体数字; 产品检查和质量控制; 图象分析。

Nestor开发系统

研制者: Nestor公司(罗德岛普里瓦登斯)。

主要功能特征: 采用每层内部联通性低的多层网络; 其软件模拟器能对非线性可分模式类进行自适应学习; 多模块系统; 学习系统能够模拟生物神经网络生成识别软件包, 包括用户开发的特征抽取器, Nestor学习及可训练软件。

应用领域: 模式识别; 手写字文字识别(约两

表2. ANZA系列的系统参数

神经计算机板	ANZA	ANZA Plus	ANZA Plus VME
处理机芯片	M68020 M68881	Weitek XL Chip Set	Weitek XL Chip Set
存储器容量	4MB	10MB	10MB
速度(IPS)	40000/25000	$6 \times 10^6 / 1.5 \times 10^6$	$6 \times 10^6 / 1.5 \times 10^6$
容量	30000个PE 480000个互联	2.5×10^6 个组 合PE	2.5×10^6 个组合PE
费用	\$ 6,900	\$ 14,950	\$ 25,000
主机(额外配置)	Zenith Z-386	Zenith Z-386	Sun 3

个字符/秒); 签名验证; 潜水艇信号处理; 抵御交易; 债务风险估计; 股票交易; 层次滤波以及各种训练课程。

费用: NDS 1000工作站(带有386处理机, 40MB存储器)为\$19500; NDS 1000软件为\$9500(Sun 3)和\$7500(PC/AT)。

Sigma神经计算机^[24]

·研制者: 国际科技应用公司(加州圣地亚哥)。

·主要功能特征: 80386加80387工作站; 带有ANSkit神经网络开发系统, 包括ANSim模拟器共13个大模块; 有45000个PE和权值; 速度为45000 IPS。

费用: 软件为\$495; 合计\$6500。

Delta浮点处理器

·研制者: 国际科技应用公司(同上)。

·主要功能特征: 它是PC/AT的加速板; 采用流水位浮点处理器(2 μ ECL技术, 60000门/单片); 12MB存储器, 40MB/秒带宽; 3.1M PE和权; 速度为11 MIPS, 22MFLOPS。

·费用: 124950。

ANSpec和ANSim

·研制者: 国际科技应用公司(同上)。

·主要功能特征: 它们是与Delta和Sigma-1配套的软件; ANsim是一个神经网络模型软件模拟器; ANSpec是基于ACTOR的面向对象神经网络系统描述语言, 适宜于实现复杂的神经网络。

MARK I神经计算机^[25]

·研制者: TRW公司(加州雷多海岸)。

·主要功能特征: 配有通用的神经网络环境; 带Netware开发语言; 以带VMS操作系统的Vax为主机; 有65000个PE和1130000个互联, 它们分成1~8层; 速度为0.5MIPS, 由用Pascal编程的Microvax系统快20倍; 最多可用15个68020微处理器,

并配有M68881浮点协处理器。

·应用领域: 信号处理; 时空模式学习; 时变光谱图分类; 图象分析; 传感器数据处理; 机器人和汽车控制; 模糊知识处理。

·价格: 约\$80000。

MARK V神经计算机^{[26][27]}

·研制者: TRW公司(同上)。

·主要功能特征: 用M68020作主处理器, 带有1~16个从处理器, 用M68020、M68881和Weitek乘法器/ALU; 带4MB RAM; 速度分别为16 MIPS(10个处理器)和11.5MIPS(8个处理器); 用VME与Microvax接口; 软件系统包括: TRW神经网络系统环境、MARK V实时监控程序、MARK V主处理器及从处理器软件; 该计划由DARPA资助。

·费用: 未公布。

NX模拟器(作者拟名)

·研制者: Human Devices公司(纽约)。

·主要功能特征: 以Macintosh I或SE为主机; 配置MacBrain; 可配1~16个RISC处理器, 最多为四块板; 每个处理器的指标是: 10MIPS/2MFLOPS、1MB RAM; NX-16最多有1600000个神经元和权, 速度为8MIPS(无学习功能)或3M IPS(有学习功能)。

·价格: NX-1/SE为\$4150; NX-16/I为\$49200。

plexus神经网络工作站

·研制者: Symbolics公司(马萨诸塞州坎布利奇)。

·主要功能特征: 以APROC 110向量处理器为基础, 用软件实现; 速度为65000 IPS, 包括3个自定义的模块。

神经网络软件模拟器

- 研制者, Neuralware公司(宾州斯威克里)。
- 主要功能特征, 以80386 PC为基础, 含有4000个PE, 425000个联接, 运算速度为100000 IPS, 以Sun为基础的含有65000个PE, 1500000个互联, 速度为100000 IPS, 有13种网络类型和14种学习规则。

- 价格: \$995 (PC), \$2995 (Sun)。

基于PC的ANS模拟软件

- 研制者, Olmstead & Watkins公司(加州埃斯康迪多)。

- 主要功能特征: 以PC为基础, 采用链接软件模块, 系统中有十个大模块。

- 价格: \$849。

N-Net 模拟器

- 研制者, AI-ware公司(俄亥俄州克利夫兰)。

- 主要功能特征: N-Net 210 PC软件模拟器, 用C语言编写, 有窗口功能, N-Net 810用于DEC Vax机。

- 价格: \$695。

3.2 已公开交流的ANS模拟软件 ASSOC和AUTOCRAY

- 研制者: James Anderson (布朗大学)。

- 主要功能特征: 用VMS Pascal编写, 在Vax上开发, ASSOC生成能使向量对联想起来的记忆矩阵; AUTOCRAY生成大型的状态向量, 即刺激, 并实现简单的非线性层次网络模型。

罗彻斯特大学模拟器

- 研制者: 罗彻斯特大学。

- 主要功能特征: 在BBN Butterfly/Vax或Sun系统上UNIX支撑下开发的, 每个单元/链接被当作一种数据结构来加以实现, 模拟则通过独立的函数调用来完成。

PC ANS模拟器

- 研制者, D. E. Rumelhart和J. L. McClelland。

- 主要功能特征: 在PC上运行软件模拟器, 该软件同他们的新作《Parallel Distributed Processing, Vol. 3, MIT Press (1987)》配套。

3.3 正在开发的以及尚未提供市场的ANS产品

(1) Bell实验室的硬件装置^[28]

Bell实验室的H. Graf, L. Jackel和J. Denker等人利用硬件手段来构造神经网络。他们开发的硬件芯片列于表3。

表3. Bell实验室的硬件芯片

芯片	字长	主要功能特征
• HPG-1	• 256×2	• 突触数为256×256×2, 固定连接
• HPG-2	• 612	• 突触数为512×512, 光学实现
• HPG-3	• 54×2	• 突触数为54×54×2, 用SRAM实现
R1	• 192	• 突触数为192×64×2, 用SRAM实现
R2	• 256	• 以Bayes最佳匹配方式工作, 用DRAM实现
R3	• 1048	• 以逆传播方式工作, 用DRAM实现

(2) 卡内基—梅隆大学的Warp收缩阵列

CMU的Warp收缩阵列包含十个处理器, 其结构如图2所示。其逆传播数据划分过程如图3所示。

在Warp收缩阵列上的模拟过程是逐层进行的, 在同一时间每个细胞机上有10个模式。其运算速度为17 MIPS。在16K的CM-1机上为2.6 MIPS。

(3) 并行Transputer结构

S. Y. Kung提出的用于构造多层ANS的大规模收缩神经网络如图4所示。图6是用于逆传播学习的大规模收缩网。

其每一层上的操作被转换为矩阵向量乘积。估

计的运算速度为2.5MIPS。

其余的一些研究工作列于表4。

以上我们从多方面讨论了ANS及神经计算机的基本概念、原理, 并概述了已公布的主要研究成果。从中可以看出它强大的生命力。然而, 在进入实用化前尚需解决下述问题:

- 对人脑神经网络的机理认识不足。
- 缺乏一致的ANS系统的设计与实现技术。
- 网络上的知识表示问题。
- 高效的学习及训练算法问题。逆传播算法在用VLSI实现时有一定的限制。因为局部值的调整

方
卷
之
五

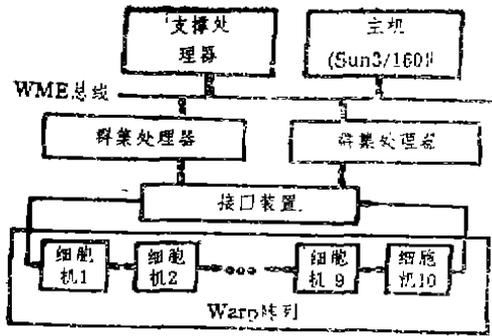


图2. CMV的Warp阵列

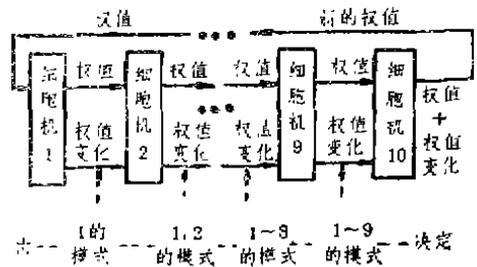


图3. Warp阵列的数据划分

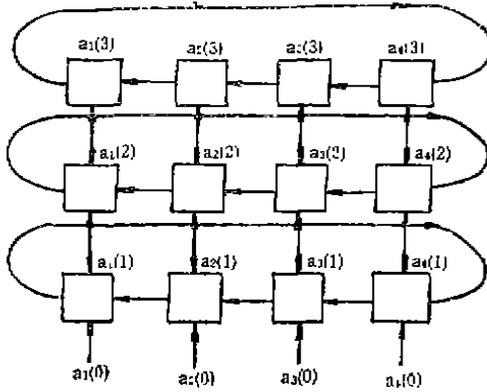


图4

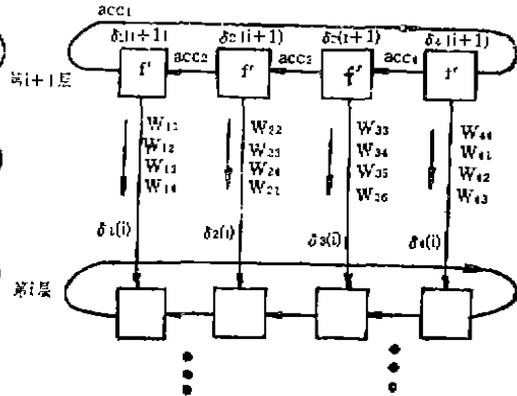


图5

表4

产品	研制者	主要功能特征
NEP	IBM 公司	<ul style="list-style-type: none"> NEP是以PC/XT作为主计算机的网络模拟处理器。 系统中可配置256个NEP，从而达到10^6个PE和4×10^6互联。
NETtalk	T. Sejnowski 和 C. Rosenberg	<ul style="list-style-type: none"> 它是一个网络模拟程序，能朗读英文文章。 309个神经元组成一个三层的网络，有18629个联接，它利用逆传播方法修改网络的权值。 它与DECtalk语言合成器及 Ridge 32微机连接构成完整的系统。 以1000字的字典训练12小时后正确率高达95%。
语言识别系统	J. Lazzaro 和 P. Mueller	<ul style="list-style-type: none"> 利用400个模拟神经元构造了语言模式分析和识别系统。 利用电阻和电容构成的互联高达100,000个。
光学非线性 联想存储器	G. Dunning, E. Maron, Y. Owechko 和B. Soffer(休斯研 究实验室)	<ul style="list-style-type: none"> 在相位共轭镜形成的光学空腔中，利用全息图构造光学非线性联想存储器。 存储多重重叠的图像，并能根据存储图像的一部分来重建整个图像。 正在开发的液晶光闸作为相位共轭镜之外的另一种选择。
CAP计划及 ABH结构	D. Hammers- trom等(俄勒冈 研究生中心)	<ul style="list-style-type: none"> CAP为认知结构计划。 采用晶片级集成(water scale integration)技术，用0.8μCMOS电路在0.8平方米面积上实现10^9个连接。 增强型广播层次(ABH)结构，它在局部通讯时采用重迭广播区域结构，而在长距离信息传递时采用点对点通讯。

需要非局部信息，因而需要全局控制。而且网络必须事先指定。

- 具有竞争性的工业应用。
- 大量神经元显现的统计及混沌问题。
- 与其它系统，如AI系统的协调问题。

所有这些都待多学科的国际合作。我们期望在不久的将来，人类在更进一步认识自身的基础上创造出更接近于智能生物的机器。

参考文献

- [1] Cai yifa, et al, "A New Neuron Model—C Model", in *Proc. of IEEE 1988 System Man & Cybernetics Conf.*
- [2] D. E. Rumelhart, J. L. McClelland (eds.), *Parallel Distributed Processing*, Vol.1, Vol.2, (1986), Vol.3(1987), MIT Press.
- [3] R. Eckmiller, C. v. d. Malsberg, *Neural Computers*, Springer-Verlag, Berlin, 1988
- [4] 蔡义发, 关于第六代计算机的研究, 《计算机科学》, No.6, 1989.
- [5] S. Grossberg, *The Adaptive Brain*, Vol. I, Vol. I, North-Holland, 1982.
- [6] S. Grossberg, *Studies of Mind and Brain*, D. Reidel, 1982.
- [7] R. Hecht-Nielsen, "Nearest Matched Filter Classification of Spatiotemporal Patterns", *Appl. Opt.*, 20, No.10, 1987
- [8] B. Kosko, "Bidirectional Associative Memory", *BYTE*, 1987
- [9] B. Kosko, "Fuzzy Associative Memory" in A. Kandel (ed.), *Fuzzy Expert Systems*, Addison-Wesley, 1987
- [10] G. E. Hinton, and T.J. Sejnowski, "Optimal Perceptual Inference," in *Proc. of IEEE Comp. Soc. Conf. on Comp. Vision and Patt. Recog.*, Washington, D.C., 1983. pp. 448-453.
- [11] J. A. Anderson, and E. Rosenfeld (eds.), *Neurocomputing: A Collection of Classic Papers*, MIT Press, 1987
- [12] P. F. C. Gilbert, "A Theory of Memory that Explains the Function and Structure of the Cerebellum," *Brain Research* 70, pp.1-18, 1973
- [13] R. Hecht-Nielsen, "Counterpropagation Networks", *Applied Optics*, Dec. 1987.
- [14] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Properties", in *Proc. Natl. Acad. Sci., USA*, 79, 2554-2558, 1982.
- [15] J. J. Hopfield, and D. W. Tank, "Neural Computation of Decisions in Optimization Problems", in *Biological Cybernetics*, 52, pp.141-152, July 1985.
- [16] J. J. Hopfield, "Neurons with Graded Response have Collective Computational Properties like those of Two State Neurons", in *Proc. of Natl. Acad. Sci., USA* 81, pp. 3088-3092, May 1984.
- [17] G. Palm, "On Associative Memory", *Biological Cybernetics*, 36, pp. 19-31, 1981.
- [18] K. Fukushima, S. Miyake, "Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shifts in Position", *Pattern Recognition* 15,, No.6, 455-469, 1984.
- [19] T. Kohonen, *Self-Organization and Associative Memory*, Second-Edition, Springer-Verlag, 1987.
- [20] R. Hecht-Nielsen, *Performance Limits of Optical, Electro-Optical and Electronic Neurocomputers*, TRW Rancho AI Center, Camel, CA, 1985.
- [21] S. Grossberg, et al, *Neural Dynamics of Adaptive Sensor-Motor Control*, North-Holland, 1986
- [22] B. J. Wah, *Neural Networks and Applications*, Coordinated Science Lab., Univ. of Illinois, 1988
- [23] ANZA *User's Guide*, Hecht-Nielsen Corp., Release 1.00, 1987.
- [24] DELTA/SIGMA/ANSim, Editorial, *Neurocomputers*, 2(1), 1988.
- [25] *The TRW Mark II-1 Artificial Neural*

人工神经网络的学习算法(1)

张鸿宾 (北京计算机学院)

摘要

人工神经网络可以看作是一种具有学习和自我组织能力的智能机器。最近几年,由于出现了一些强有力的学习算法,大大推动了有关神经网络的理论和应用研究。本文介绍并分析其中几个有代表性的学习算法。

关于智能机器的制造方法,很早以来一直有两种不同的策略在互相竞争着。第一种是以逻辑为基础的,认为只有符号运算才称得上是所谓智能的本质。这种策略是在通常的串行计算机上实现的,在专家系统和计算机博弈等领域中取得了很大的成功。但在处理和外部世界有较多联系的对象,诸如图像处理 and 声音、物体识别,以及常识推理等时,却显得能力不强,暴露出其局限性。

第二种策略是试图模仿人脑的信息处理机制,通过把大量的结构非常简单的计算单元(神经元)相互连接起来,构成人工神经网络,实现高度并行和分散的信息处理,用来进行物体和声音识别、联想存储和常识推理等。这种神经网络型计算机不需要编写程序,它可以从示例中进行学习,实现网络的自我组织。比起第一种策略来,这种策略在

过去一些年进展不大,成果也不怎么显著。究其原因,主要有以下两点。一是过去的神经网络学习算法不够完善和强有力,信息表现的方式也不适当。二是在通常的串行机上模拟神经网络的效率非常低。然而,随着VLSI和CAD技术的进步,神经网络的硬件实现已经变得较为容易了。特别是由于近几年提出了一些强有力的学习算法,使得一度沉寂的神经网络型计算机的研究又出现了一个新的高潮。人们期待这个一度落后了的“小兄弟”能够成长起来,和“老大”——串行数字机携手,在未来的智能机器中发挥更大的作用。

本文在回顾以往的神经网络学习算法的基础上,介绍和分析最近几年新提出的两个主要学习算法——误差反向传播算法和Boltzmann Machine的学习算法,展望它们的

System Processor, Product Description,
TRW MEAD AI Center, San Diego,
1987

[26] R. W. Kuczewski, et al, "Neurocomputer Workstations and Processors: Approaches and Applications", *IEEE ICNN Conf.*, 1987.

[27] J. Prichett, "The New Cybernetics: ADAPting to AI", *TRW Message s*,4(1).

1986.

[28] L. D. Jackel, et al, "Electronic Neural Networks", in *Proc. of Supercomputing'88*, pp.41-46.

[29] L. D. Jackel, et al, "Electronic Neural Networks Chips", *Applied Optics*, 26, 1987.

[30] 蔡义发,丁德恒,吕维雪,电子神经计算机的设计与实践,《计算机学报》(待发表)。