

通用神经网络硬件中神经元基本数学模型的讨论

王守觉,李兆洲,陈向东,王柏南

(中国科学院半导体研究所,北京 100083)

摘要: 在介绍了作者实现通用神经网络硬件中应用的通用计算公式的基础上,提出了一种能同时模拟包括 RBF 与传统 BP 网络神经元在内的各种神经元通用的新的数学计算模型,并把基于这种通用数学计算模型的神经网络 CASSANDRA- 型神经计算机结构设计中并予以硬件实现.文中还讨论了它所模拟神经网络的灵活性.

关键词: 神经网络;神经计算机;神经元数学模型

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2001) 05-0577-04

Discussion on the Basic Mathematical Models of Neurons in General Purpose Neurocomputer

WANG Shou-jue, LI Zhao-zhou, CHEN Xiang-dong, WANG Bai-nan

(Institute of Semiconductors, CAS, Beijing 100083, China)

Abstract: Based on the introduction of the traditional mathematical model of neurons in general-purpose neural network hardware realization, a novel all-purpose mathematical model has been proposed, which can simulate all kinds of neuron architecture, including RBF and BP models, etc. At the same time this model has been realized using hardware and implemented in the new CASSANDRA- neurocomputer that can be used to form many neural networks with multiple mathematical models of neurons. In this paper the flexibility of the new neurons has also been discussed in constructing neural networks.

Key words: neural network; neurocomputer; mathematical model of neurons

1 引言

人工神经网络的优越特性之一在于它适合于大量并行计算.因而神经网络硬件可以用较小的硬件开销而获得很高的网络联接计算速度与规模.作为通用目的的神经网络硬件可以是芯片集成或多芯片系统甚至大圆片集成(wafer scale integration WSI)以至多圆片系统^[1,2].但不论集成度的大小如何,也不论神经网络是用硬件实现还是用软件在通用计算机上模拟,其最基本的单元是神经元.因而,神经元的基本计算方法和功能是决定神经网络性能的最基本的基础.

从模拟脑细胞的激活和抑制的机制出发最早就提出了经典的神经元数学模型^[3]:

$$Y = f \left(\sum_{i=0}^n W_i X_i - \right) \quad (1)$$

式中: Y 为神经元输出; f 为神经元激励函数(非线性函数); X 为神经元的输入; W 为神经元的权值; \quad 为神经元的激活阈值.

由式(1)可知神经元的输出是由两个因素确定.一个是激励函数 f , 另一个是函数的基 $\left(\sum_{i=0}^n W_i X_i - \right)$. 不难看出, 这一函数的基就是输入空间中输入点离一个超平面的距离(一侧为

正,另一侧为负),该超平面的方程是:

$$\sum_{i=0}^n W_i X_i - = 0 \quad (2)$$

如果激励函数是阶跃函数则神经元的功能就是在多维输入空间中作一个超平面,输入位于此超平面的一侧时其输出为 1,而在另一侧时输出为 0. 这种功能常用于模式分类器^[4]中.

很早人们就提出了用多维空间中的封闭超曲面取代式(2)超平面的设想^[5]. 径向基函数(RBF)神经网络,它的神经元激励函数的基不是离一个超平面的距离,而是离一个核心点的距离^[6]. 实验结果表明径向基函数(RBF)神经网络在模式识别与函数拟合的应用中具有比式(1)所描述的神元更优越的性能.

径向基函数神经元的数学模型可表示为:

$$Y = f \left(\sum_{i=0}^n (W_i - X_i)^2 - \right) \quad (3)$$

可以看到,如果式(3)RBF神经元的激励函数 f 为阶跃函数,则该神经元相当于在输入空间中以核心 W_i 为球心,以 \quad 为半径作一个超球面.当输入点在此超球面内时输出为“0”,

在超球面以外时输出为“1”。因而,可以认为 RBF 神经网络是一种最简单的封闭超曲面高阶神经网络。它的性能既然优于式(1)超平面神经网络,这说明超曲面高阶神经网络在性能上的优越性。

本文的目的就在探讨一种适用于通用神经网络硬件化的通用性强、功能齐全的神经元基本算法模型,使在此基础上实现的通用神经网络硬件,能有更强的适应能力和更好的网络性能。本文以下所讨论的新的超曲面高阶神经元基本算法已被应用于 CASSANDRA- I 型神经计算机的设计与实践中。

2 通用神经网络硬件中的通用计算式

作为通用神经网络硬件的基本要求,是要能适应各种各样神经网络模型联接、各种各样神经元激励函数以及尽可能灵活的神经元算法模型。一种比较实用的通用神经网络计算机的结构方式就是把所有需要灵活可变的因素都以可变参数的形式放在一个通用计算式中,神经计算机就并行地不断重复进行该通用计算式的计算,而这些可变参数却是按需要随时随处在变动着的,从而实现复杂灵活的神经网络结构^[7]。例如我国在 1995 年实现的 CASSANDRA-I 神经计算机就是建立在下述通用计算式基础上的:

$$O_{mi}(t+1) = F_{k_i} \left[C_i \left(\sum_{j=0}^{n-1} S_{ji} W_{ji} I_{mj} + \sum_{g=0}^{n-1} S_{gi} W_{gi} O_{mg}(t) \right) - \theta_i \right] \quad (4)$$

式中: $O_{mi}(t+1)$ 为输入空间输入第 m 个对象时第 i 个神经元在 $(t+1)$ 时间的输出状态值,其中 m 最大不超过 128, i 不超过 256; n 为输入节点数量(即输入空间维数)和神经元的最大数量(不超过 256); F_{k_i} 为第 i 个神经元的输出非线性函数,其下标 k_i 为第 i 个神经元所用非线性函数在函数库中的序号(1~64); I_{mj} 为第 m 个输入对象中的第 j 个(即第 j 维)输入值(输入至网络第 j 个输入节点); $O_{mg}(t)$ 为在计算输入空间第 m 个对象时第 g 个神经元在时间 (t) 的输出状态值; W_{ji} 为由第 j 个输入节点至第 i 个神经元的联接权值; W_{gi} 为由第 g 个神经元至第 i 个神经元的联接权值; S_{ji} 和 S_{gi} 为决定网络拓扑结构模型的参数, S_{ji} 为 0 时表示第 j 个输入节点与第 i 个神经元无联接, S_{gi} 为 0 时表示第 g 个神经元输出与第 i 个神经元无联接,反之则为 1 时表示有联接; θ_i 为第 i 个神经元的阈值; C_i 为扩大运算动态范围用的比例因子(1,2,4,8)。

由通用计算式(4)可知 CASSANDRA-I 可以计算任意拓扑结构的前馈网络或反馈网络;各个神经元可以在非线性函数库中任意选择不同的激励函数;可以有不同阈值和不同比例因子。因此,可以说在网络拓扑结构和神经元参数方面它有很大的灵活性与适应能力。

然而,不难看出通用计算式(4)的神经元基础是式(1)的数学模型,因而它只能作“超平面”神经网络的运算,而较难作超曲面神经网络的运算。利用式(1)的基本运算来计算矢量间的夹角,以夹角作为神经元非线性函数的基,实现一种方向基函数(DBF)神经网络可以实现高阶神经元的功能^[8];但它所实现的封闭超曲面功能是在降低一维的超空间中实现

的,输入各矢量的模值在计算归一化过程中将被抛弃。

本文以下将主要探讨一种通用的功能强灵活性高的高阶神经元基本运算模型和对它的功能进行讨论。

3 超曲面神经元基本数学模型的通用性探讨

为通用神经计算机选择的超曲面神经元基本数学模型应具有下列基本要求原则:

- (1)可以同时覆盖传统的超平面神经元和径向基函数 RBF 神经网络的功能;
- (2)有实现尽可能多种不同超曲面特性的可能性;
- (3)特性的改变可以用更动少数参数来实现;
- (4)能较为方便地用硬件来进行快速运算。

由要求(1)可知通用神经元基本数学运算公式必须兼有式(1)和式(3)的基本运算内容。所以对于每一个输入节点来的信号必须有两个权值。一个确定方向的权值和一个确定核心的权值。选用的基本计算式如下:

$$Y = f \left[\sum_{j=1}^n \left(\frac{W_{ji}}{|W_{ji}|} \right)^S |W_{ji}(X_j - W_{ji})|^p - \right] \quad (5)$$

式中: Y 为神经元输出; f 为神经元激励函数; θ 为神经元阈值; W_{ji} 和 w_{ji} 为由第 j 个输入端接至神经元的两个权值; X_j 为第 j 个输入端输入(正值); n 为输入空间维数; S 为决定单项正负号方法的参数 $S=0$ 时单项符号永为正, $S=1$ 时单项的符号与 W_{ji} 的符号相同; p 为幂参数。

不难看出,如果所有的 w 均为“0”, $S=1, p=1$, 则式(5)就与式(1)相当;如果所有的 w 均为“1”, $S=0, p=2$, 则式(5)就与式(3)相当。说明该基本算式符合要求(1)。

再看,如果设定 $S=0$, 则该基本算式是个封闭超曲面神经元。 F 函数的基设定为定值时,输入点的轨迹是封闭超曲面。它的核心位置由 w 决定;设先考虑所有 w 均相等的情况,该封闭超曲面可以用 p 的值来改变它的形状。以三维空间的情况来说明, p 值为 1/3, 1/2, 1, 2, 3, 4 等不同值时该封闭曲面如图 1~图 6。

此时若使 w 权值具有不同的值,则相当于将封闭曲面在不同方向加以拉伸或压缩。如果 $p=2$, 则超球面将在不同维方向伸或缩形成不同的超椭球面。图 7 与图 8 是举例在三维空间中在不同轴方向有不同拉伸或压缩形成的椭球面。

可以看到式(5)同时符合(2)和(3)二个要求。

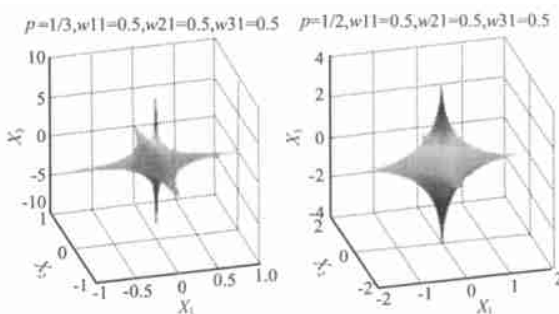


图 1

图 2

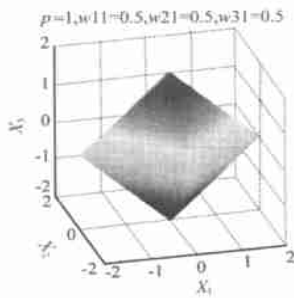


图 3

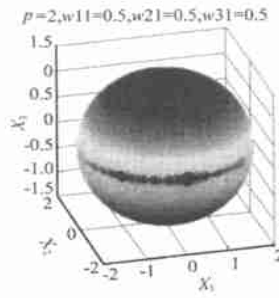


图 4

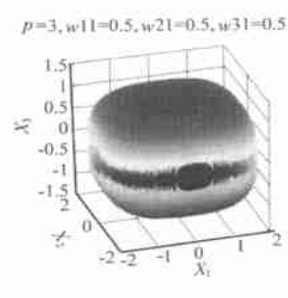


图 5

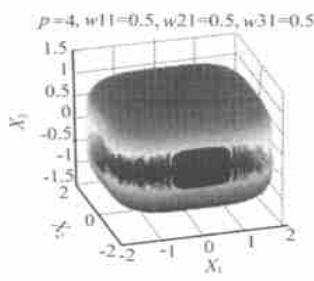


图 6

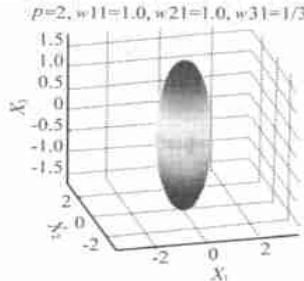


图 7

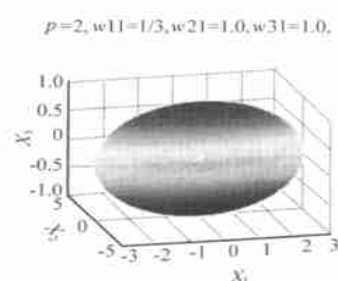


图 8

4 基于通用超曲面神经元的神经网络硬件通用计算式

基于通用超曲面神经元基本数学模型式(5)作者研制了同时适用于传统 BP 网络、RBF 网络以及各种高阶超曲面神经元网络通用的 CASSANDRA- 神经计算机硬件。

其通用计算式如下:

$$O_{mi}(t+1) = F_{k_i} \{ \sum_{j=1}^n C_i (I_{mj} - W_{ji}) \} \quad (6)$$

其中:

$$\mathfrak{N} = \left[\sum_{j=1}^n \left(\frac{W_{ji}}{|W_{ji}|} |W_{ji}(I_{mj} - W_{ji})|^p \right) + \sum_{g=1}^n \left(\frac{W_{gi}}{|W_{gi}|} |W_{gi}(O_{mg}(t) - W_{gi})|^p \right) \right]$$

式中: $O_{mi}(t+1)$ 为输入空间输入第 m 个对象时第 i 个神经元在 $(t+1)$ 时间的输出状态值, 其中 m 最大不超过 128, i 神经元数量最大不超过 1024; F_{k_i} 为第 i 个神经元的输出非线性函数, 其下标 k_i 为第 i 个神经元所用非线性函数在函数库中的序号 (1 至 8); I_{mj} 为第 m 个输入对象中的第 j 个 (即第 j 维) 输入值 (输入至网络第 j 个输入节点); W_{ji} 与 W_{ji} 分别为由第 j 个输入节点至第 i 个神经元的“方向”权值和“核心”权值; W_{ji} 与 W_{ji} 分别为由第 g 个 ($1 \leq g \leq 256$) 神经元输出联接至第 i 个 ($1 \leq i \leq 1024$) 神经元的“方向”权值和“核心”权值; p 为幂参数 (1/3, 1/2, 1, 2, 3, 4); S 为单项正负符号规则 (0 或 1); $O_{mg}(t)$ 为当输入为第 m 个对象时第 g 个 ($1 \leq g \leq 256$) 神经元在时间 (t) 的输出状态值; i 为第 i 个 ($1 \leq i \leq 1024$) 神经元的阈值; C_i 为神经元输入规模比例因子; i 为神经元非线性函数坐标比例因子。

由式 (6) 可知 CASSANDRA- 神经计算机可以模拟含有

各种不同特性神经元的 (包括超平面, 超球面以及各种超椭球面和超立方体等) 任意结构的神经网络。其一次模拟网络规模的限制为神经元总数量 1024 个, 其中可作为中间层 (即输出可作为其他神经元或自身输入的) 神经元不超过 256 个。故每个神经元的输入突触限为 512 个。对于前馈网络, 一次可计算输入空间中 128 个以下对象。对于有反馈网络, 一次可迭代计算 128 次。

CASSANDRA- 的权值与状态输入输出的数据格式采用了 16 位浮点格式, 以保证其等效运算精度。

5 结论

在美国 IEEE 1992 年出版的神经网络理论基础与分析论文集的引言^[9]中对词汇“人工神经网络”含义作了如下描述: “用大量简单的“神经元”计算并行叠联而成的任何计算结构”, 即“神经网络”, “更确切地说是人工神经网络”。其发展的驱动力主要来自这样一个事实, 即在图像识别、模式识别方面人类的能力远远高于任何数字计算机”。

在人工神经网络用于模式识别时, RBF 网络等以封闭超曲面特性神经元为基础的网络, 显然优于传统的以超平面特性神经元为基础的神经网络。这可能会导致象飞机发展过程中那样很早就超脱了鸟类飞行所用的翅膀上下运动的机制。虽然超平面特性神经元的数学模型更接近于单个脑细胞的功能。

本文提出了并在作者所研制的神经计算机中实用了灵活可变超曲面特性神经元作为网络基础, 希望引起有共识的人工神经网络研究者和实用者们给予神经元的基本数学模型的发展以更多的关注。

参考文献:

- [1] Clark S. Lindsey and Thomas Lindblad. Review of Hardware Neural Networks:A User's Perspective [A]. Proceedings of the Neural Networks [C],From Biology to High Energy Physics,International Journal of Neural System,Supp. 1995,6:215 - 224.
- [2] Yuzo Hirai. VLSI Neural Network Systems [A]. Proceedings of the Neural Networks [C],From Biology to High Energy Physics,International Journal of Neural System,Supp. 1995,6:203 - 213.
- [3] W. S. McCulloch, and W. Pitts. A logic Calculus of the Ideas Imminent in Nerves Activity [J]. Bulletin of Mathematical Biophysics, 1943, 5: 115 - 133.
- [4] M. A. Cohen, and S. Grossberg. Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks [J]. IEEE Trans. Syst. Man Cybern., 1983, SMC-13:815 - 826.
- [5] R. L. Hardy. Multiquadric equations of topography and other irregular surfaces [J]. J. Geophys. Res., 76:1905 - 1915.
- [6] M. J. D. Powell. Radial basis functions for multivariable interpolation: A review [A]. in J. C. Mason and M. G. Cox, editors, Algorithms for Approximation. Clarendon Press, Oxford, 1987.
- [7] D. R. Collins and P. A. Penz. Considerations for neural network hardware implementations [J]. IEEE Symp. Circuit and System, 1989:834 - 836.
- [8] Wang Shoujue, Shi Jingpu, Chen Chuan, Li Yujian. Direction-Basis-

Function Neural Networks [A]. # 58 Session:4.2, IJCNN 99 [C].

- [9] Clifford Lau edited. Neural Networks: Theoretical Foundations and Analysis [J]. IEEE PRESS, A Selected Reprint Volume, Neural Networks Council, Sponsor, 1992.

作者简介:

王守觉 中国科学院半导体研究所研究员,中科院院士. 1925 年生于上海市,早年就读于西南联大和同济大学,毕业后在北平研究院镭学研究所从事氧化亚铜研究,解放后改为中国科学院应用物理所结晶学室. 1960 年成立半导体所后历任器件室主任、副所长、所长等职务. 1980 年当选为学部委员(院士). 现为中国电子学会副理事长,电子学报主编. 他是中国半导体学科奠基人之一,现从事半导体超高速电路与神经网络算法、模型、硬件和应用的研究.

李兆渊 1972 年生,1995 年毕业于清华大学,后保送中科院半导体所师从王守觉院士攻读博士学位,主要从事神经网络算法和应用研究.

陈向东 1971 年生,1994 年毕业于西安交通大学,1994 年至 2000 年一直在半导体所师从王守觉院士攻读硕士、博士学位,主要从事人工神经网络算法、应用和硬件实现的研究.

王柏南 1973 年出生,1997 年毕业于北京大学,后保送中科院半导体所师从王守觉院士攻读硕士学位,现已留所. 主要从事人工神经网络理论研究.