

神经计算中坐标变换的网络模型(CMAC)的泛化特性¹⁾

欧阳楷 陈卉 周萍 周琛

(首都医科大学生物医学工程系 北京 100054)

摘要 在神经计算中神经网络的泛化特性是一个非常重要的内容. 该文简述了小脑模型(CMAC——Cerebellar Model Articulation Controller)的原理和学习算法, 并用仿真方法讨论了在机器人使用的坐标变换关系(输入直角坐标值, 输出机器手的关节角度)下CMAC的泛化性能: 当泛化率为1:100时CMAC仍能正常工作. 系统的精度虽能满足需要, 但是进一步提高却受到限制. 本文还讨论了影响精度的各种因素及可能的改进方法.

关键词 泛化性能, 小脑模型(CMAC), 坐标变换.

1 引言

自Marr的视觉计算理论出版以来, 计算神经科学已经成为神经科学的一个重要分支. 虽然计算神经科学研究的结果只是神经系统信息加工定量描述的一种可能性, 还需要生物实验的验证, 但在目前神经系统所能提供的知识不完全的情况下, 这种方法仍不失为一种科学探索, 特别是由它提出的某些推测已成为生物实验研究的出发点和信息科学进一步发展的新观点、新方法的源泉. 因此它引起了科学家的极大关注, 成为当前科学的前沿之一.

神经计算中坐标变换是常见的一类计算任务. 用数学方法或符号系统串行完成这类任务是不成问题的, 而用网络方法平行地完成就不那么容易了. 其根本原因在于网络完成坐标变换任务时需要给定的网络进行训练, 而且要求在全平面上对所有的点都进行训练, 使之在一定的分辨率下达到精度, 其训练样本集相当大. 例如有人在研究视跟踪系统中的神经计算时, 将转动着的视网膜的中央凹上映射的靶象转换到固定坐标系中, 所用训练集的样本量达到5184, 其分辨率仅为整个区域的 $1/256^2$. 所以这里提出了神经网络的一个重要理论问题——泛化问题, 即用较少的样本进行训练, 使网络能在给定的区域内达到指定的精度.

最早进行泛化理论研究的是Amari^[1], 他认为泛化是将输入集中样本点的给定邻域映射到输出集中映射点(与样本点对应)的某一邻域. 由此可以看出泛化能力除了由精度决定

1) 本文受国家自然科学基金、“八六三”计划、攀登计划支持.

2) 刘林, 灵长类视觉运动感知的神经计算研究; [博士学位论文]. 北京: 中国科学院生物物理所, 1995. 52—56.

外,还取决于映射方式.所以多层感知机的泛化能力是极其有限的,实践也证实了这一点.

本文研究一种非线性输入输出网络,它是受小脑的神经解剖生理启发而建立的,Albus 称之为 CMAC(Cerebellar Model Articulation Controller)^[2].这种网络的非线性映射关系与多层感知机不同.本文在简单介绍了 CMAC 网络的构成后,着重研究其泛化性能.

2 模型简述

2.1 模型的结构

CMAC 中输入输出之间的映射方法是由输入激活 C 个单元,这 C 个单元内存储的数据累加构成输出.这 C 个单元是按“输入相邻则输出相近”的原则来编码的,即相邻的输入(输入值比较接近)所对应的 C 个单元中相同的较多,不相同的较少,由此可以看出, C 选取得越大,对样本的映射关系的影响越大,从而可以形成较好的泛化性能.

简单的 CMAC 模型如图 1 所示.可以看出由输入集合 S 到输出集合 P 的映射分为两部分,即 $S \rightarrow A$ (A 为存放权重的地址空间)和 $A \rightarrow P$.

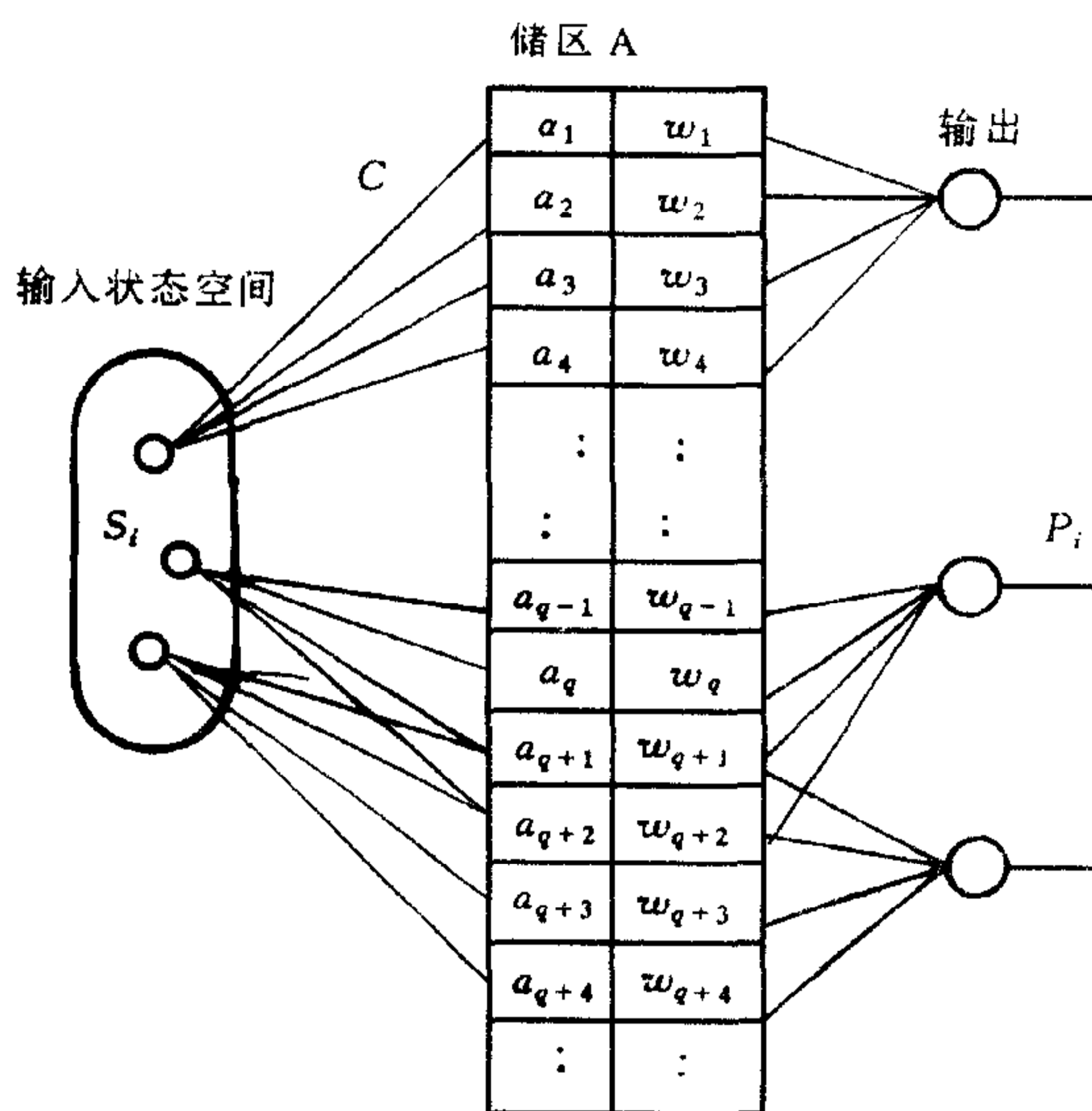


图 1 简单的 CMAC 模型

首先应确定参量 C ,有的文献称之为概括能力,有的称为综合能力(generalization).它实质上代表输入样本间相互影响的能力. C 值越大,CMAC 的泛化能力越强,但学习时间势必加长.因此确定 C 的数值是很重要的,我们将在下面详细讨论泛化能力与 C 的关系.

$S \rightarrow A$ 的映射方法是对量化后的输入样本 S 的每个变量 S_i ,根据它的值的大小,按一定规则,将其映射成 C 个虚拟单元地址 $V_{i1}, V_{i2}, \dots, V_{iC}$,它们代表输入变量 S_i 激活的单元.将这些虚拟单元重新组合成 C 个符号序列 $a_1^*, a_2^*, \dots, a_C^*$ (记为 A^*),将它们作为存储权重的地址,其中 a_i^* 为 $V_{1i}, V_{2i}, \dots, V_{Ni}$ 的某种组合.要注意:

(1)将样本的输入变量 S_i 映射成虚拟单元地址 V_{ij} 的方法(或称规则)一般对最终结

2.2 激活单元的确定

2.2.1 数据的量化

由于输入量 S 多为 N 维模拟量, $S = (S_1, S_2, \dots, S_N)$, CMAC 对这些连续量要进行量化,使之成为离散整数.例如,对于 N 维输入向量的每一分量 S_i ,将其量化为 $1 \sim R_i$ 范围内的一个整数 S_i' :

$$\Delta S_i = \frac{S_{i\max} - S_{i\min}}{R_i},$$

$$S_i' = \text{int} \left(\frac{S_i - S_{i\min}}{\Delta S_i} \right) + 1.$$

其中 $S_{i\min}$ $S_{i\max}$ 分别为 S_i 的最大值和最小值, ΔS_i 为量化分辨率.

2.2.2 将输入向量映射到存储器 A 的地址

果没有影响,但一旦确定下来,则各变量映射的方法都应该是一致的.

(2)要按照数值相近、编码相近的原则来确定虚拟单元,如果两个样本的第 i 个变量 S'_i 较 S''_i 差一个单位,则对应的 V'_{i1} 与 V''_{i1} 相差一个单元,即相互错开一个单元.

2.2.3 由 A 映射到输出向量 P

在存储器 A 中,每个单元中都存放着 L 个权重(L 为输出向量的维数),输入向量激活 C 个单元后,输出向量 P 可以按下式计算:

$$p_j = \sum_{k \in A^*} w_{jk}, \quad j = 1, 2, \dots, L. \quad (1)$$

2.2.4 学习算法

2.2.4.1 权重的确定——有监督的学习

设样本为 (S^*, P^*) , CMAC 的输出为 P , 给定一个误差限 ϵ_j , 若 $|p_j^* - p_j| \leq \epsilon_j$ ($j=1, 2, \dots, L$), 对应的权不修改; 否则按下式调整权:

$$w_j(t+1) = w_j(t) + \beta \cdot \frac{p_j^* - p_j}{C}. \quad (2)$$

其中 β 为学习率, t 为学习次数. 此公式则意味着权重一次均摊到 C 个激活的单元上, 而且不是迭代计算.

2.2.4.2 多个样本的学习

给定样本集 S^1, S^2, \dots, S^M , 按训练集中样本的顺序逐个学习. 每次按(2)式训练一个样本, 这样对于有公共地址单元的两个样本, 当第二个样本被满足时就会破坏第一个样本的输出精度. 因此要对样本集进行多次循环训练, 以达到各个训练样本的输出精度都得到满足.

3 仿真研究

3.1 输入输出关系

神经计算中的坐标变换有很多种, 繁简不一. 为研究泛化问题的一般性, 我们选择了较复杂的, 即小脑运动控制中的多自由度的关节控制所表示的坐标变换. 以 $X(x, y, z)$ 表示肢体要达到的空间位置坐标, 以 $\theta(\theta_1, \dots, \theta_N)$ 表示各关节应转过的角度, 则有 $X=f(\theta)$. 运动控制的坐标变换就是已知 X 求 θ (或已知 θ 求 X). 设空间某点坐标为 (x, y, z) , 三自由度手臂的角度为 $(\theta_1, \theta_2, \theta_3)$, 手臂长度为 L , 则

$$\begin{cases} x = l[\cos\theta_1 - \cos(\theta_1 - \theta_3)]\cos\theta_2, \\ y = l[\cos\theta_1 - \cos(\theta_1 - \theta_3)]\sin\theta_2, \\ z = l[\sin\theta_1 - \sin(\theta_1 - \theta_3)]. \end{cases} \quad (3)$$

CMAC 的坐标变换任务就是已知 (x, y, z) , 求解 $(\theta_1, \theta_2, \theta_3)$.

3.2 样本集的确定

坐标变换的输入集应遍历全部三维坐标的空间, 这里为简单起见, 令 z 为常数, 则输入限于平面 xy . 不失一般性, 可取平面内第一象限. 设手臂长度为 30, 量化分辨率为 0.1, 则 x, y 坐标各取 300 个点, 故输入集内点的个数为 70686 (1/4 圆内的点). 训练样本取 13 根辐射线 (射线间夹角为 7.5°) 上的点, 每线取 60 点, 则样本集的数量为 $13 \times 60 = 780$. 由

于这些辐射线上的点经量化后有一部分是要相同的,因而实际上样本数量为 703.那么在 703 个样本训练条件下,要求系统能在 70686 输入集下正常运行,泛化率定义为 $703 : 70686 \approx 1 : 100$.

3.3 精度

为便于研究误差,在本文的仿真研究中,坐标变换 (X, θ) 的关系由 (3) 给出,教师信号 θ 由 (3) 的逆运算给出,即 $\theta = f^{-1}X$,

$$\begin{cases} \theta_1 = -\arccos \frac{z}{\sqrt{x^2 + y^2 + z^2}} + \arcsin \frac{\sqrt{x^2 + y^2 + z^2}}{2l}, \\ \theta_2 = \operatorname{arctg} \frac{y}{x}, \\ \theta_3 = 2\arcsin \frac{\sqrt{x^2 + y^2 + z^2}}{2l}. \end{cases} \quad (4)$$

无论是输入集还是训练集,都由 CMAC 从 X 求出 θ ,再由 θ 按 (3) 式算出 \tilde{X} ,以 X 和 \tilde{X} 的差作为误差,结果如表 1 所示.

表 1 误差分析 (以 x 坐标为例)

	最大误差	最小误差	平均误差	均方差
样本集	2.212	0.001	0.605	0.457
输入集	3.649	0.001	0.677	0.577

以圆和正弦曲线为例,观察只训练输入集的 1% 的情况下系统的泛化能力(见图 2).

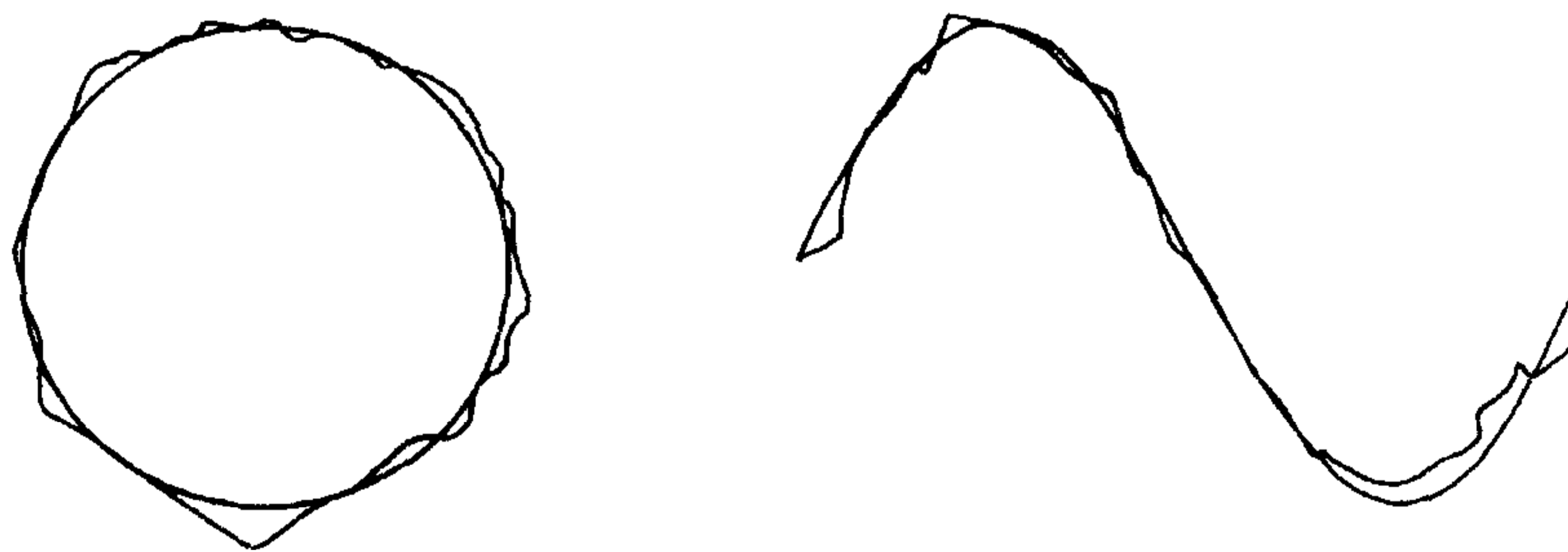


图 2 训练后 CMAC 泛化能力

(注:光滑曲线为样本曲线,不光滑曲线为 CMAC 输出的曲线)

3.4 影响精度的因素

由图 2 可以看出系统具有一定的泛化能力,但是比较粗糙.为提高精度,需要研究它的影响因素,从而找到提高泛化精度的办法.

1) 提高训练精度

由于 CMAC 与多层感知机不同,其学习算法不是在最小乘方意义下的优化计算,其收敛性并无严格的证明.仿真的结果表明,在 C 较大、各样本间有较大交叠的情况下,训练误差在达到一定值时,将不再继续减小,表 1 的误差就是在此条件下列出的.

对 CMAC 学习算法中解的存在性与收敛性,目前尚无条件进行较为严格的讨论,文 [3,4] 中的讨论不适宜 C 较大、泛化能力较好的情况.根据仿真的结果可以指出,误差虽

不能趋于 0,但是可以减小到一定程度.

2) 误差分析

根据本文所用 CMAC 模型及仿真方法,误差分为以下各部分:

(1) 由于式(3)和(4)是非线性计算,故存在一定的误差;在编程中选用双精度浮点数时,此误差可以忽略(大约小于 10^{-6}).

(2) 样本训练时遗留的误差. 这与很多因素有关,如 C 的大小、样本的分布、输入与输出的对应关系、量化级数以及学习算法等.

(3) 泛化误差. 由于 CMAC 是利用各样本激活单元的重叠来获得泛化性能的,实质上是利用相邻样本的相互关系以获得输出,它显然不会是完全恰当的输出,这就带来了泛化误差. 在表 1 中输入集较样本集的误差为大就是这个原因.

(4) 样本序列始端误差. 由于 CMAC 的学习算法是把样本排成序列逐个训练,训练后序样本时会改变与前序样本激活的单元交叠的单元的值,因此最后一轮样本训练完毕时,序列前面几个样本的误差就较大,这就是序列始端误差. 我们把样本的顺序以辐射线为单位,将辐射线与 X 轴的夹角从 0° 变化到 90° 、从 90° 变化到 0° 构成不同的样本序列,其误差分布情况都是前序样本的误差较大(最前面两条辐射线上的样本中,训练时未达到训练精度的样本分别占 40.3% 和 48.9%).

当角度变化从 45° 向两边分别达到 0° 和 90° 时,由于样本数值的变化,还会引起较大的学习振荡,误差增大.

(5) 量化误差. CMAC 对模拟输入要进行量化,然后再送入网络进行映射,因此即使网络训练得非常好,也会存在由量化引起的误差.

3) 每个样本激活的单元数 C 对泛化的影响

C 是 CMAC 泛化能力产生的基础, C 的大小与泛化能力有很大关系: C 较小泛化能力较低; C 增加泛化能力增加,但训练的时间要增加,并且训练遗留的误差以及始端误差也增加,因此也不是 C 越大泛化性能越好(见表 2).

表 2 不同情况下系统的误差(以圆曲线 x 坐标为例)

		最大值	最小值	平均值	均方差
C 的大小	15	20.001	0.076	3.712	4.905
	30	2.800	0.003	0.877	0.756
	50	2.373	0.004	0.625	0.579
	80	1.984	0.005	0.696	0.555
	120	2.657	0.027	1.207	0.747
样本量	703	1.984	0.005	0.696	0.555
	553	2.178	0.008	0.646	0.564
	385	3.551	0.052	1.186	0.872
样本分布	射线	1.984	0.005	0.696	0.555
	方格	2.149	0.013	0.699	0.516
	圆周	1.532	0.018	0.661	0.412

我们的仿真研究经验表明:

(1) C 有一个合理的数值范围, 过大过小都不好, 如本实验中 C 在 50~80 之间为宜;

(2) 从经验上讲, 确定 C 的大小以样本点所能影响的范围能覆盖非样本点、并能反映输入输出的关系为原则. 表 2 列出了 C 值相同 ($C=80$)、样本数不同时系统的误差.

4) 样本点的选择

本文利用辐射线分布构成样本集(主要考虑到输出变量为角度), 当然也可以有其它分布的样本集, 如方格分布、圆周分布、随机分布等. 样本分布对系统精度的影响见表 2.

4 讨论

从以上 CMAC 对较复杂的非线性输入输出关系映射的仿真研究中可以看出, CMAC 较多层感知机有更大的泛化能力, 这对于神经网络是一个重要的、值得重视的性能. 但是由于它的学习算法较为粗糙, 因此系统的映射能力比较粗糙, 映射的精度不高, 只能说大体上具有一定的泛化非线性映射的能力.

所以在保证 CMAC 计算速度和泛化能力的前提下, 进一步提高其精度就成为神经网络 CMAC 理论研究领域的一项挑战. 作者认为, 师法生物实际, 可以提出以下几个改进的方向, 这方面的工作也正在进行中.

1) 加上背角反射能力. 脊椎动物运动系统脊椎神经中的背角具有某种固定的反射能力, 它可以在某一小的范围内改进系统的精度. 我们的初步实验表明这是有效的.

2) 基于小脑解剖与生理的 CMAC 模型的改进. 人的小脑对运动控制具有较复杂的结构和机理, 可以完成较为精细的、可学习的运动控制, 师法生物实际将是重要的改进方向.

3) 改进学习算法与信息存储方法.

4) 根据映射关系确定样本集的大小与分布也是把经验知识用于学习的有效措施之一. 同时 C 的结构及数值的分配也是值得研究的.

参 考 文 献

- [1] Amari S L. Mathematical foundations of neurocomputing. Technical Reports Faculty Engineering University of Tokyo, Japan, 1989. 1—48.
- [2] Albus J S. A new approach to manipulator control: Cerebellar Model Articulation Controller (CMAC). *IEEE Trans. ASME*. 1975. 9. 220—227.
- [3] 张立明. 人工神经网络模型及其应用. 上海: 复旦大学出版社, 1993. 170—184.
- [4] Wang Yin fai. A Sideris learning convergence in the Cerebellar Model Articulation Controller (CMAC). *IEEE Trans. Neural Network*, 1992, 3(1): 115—121.

GENERALIZATION OF NEURAL NETWORK MODEL(CMAC) FOR COORDINATE TRANSFORMATION IN NEURAL COMPUTATION

OUYANG KAI CHEN HUI ZHOU PING ZHOU CHEN

(Dept. of BME., Capital University of Medical Science, Beijing 100054)

Abstract Generalization of neural network is a very important topic for coordinate transformation in neural computation. In this paper, we describe the principle of Cerebellar Model Articulation Controller(CMAC) including its learning algorithm, and discuss the generalization of CMAC through simulation of coordinate transformation (the input is position coordinate values and the output is articulation degrees of robot). The CMAC may still run well at generalization rate 1 : 100. Several factors affecting the accuracy are also discussed.

Key words Generalization, coordinate transformation, cerebellar model articulation controller (CMAC).

欧阳楷 男, 62岁, 教授, 1959年毕业于清华大学. 目前是北京生物医学工程学会理事, 全国生物医学工程学会、全国生物物理学会、全国自动化学会三个学会的生物信息与控制分科学会委员, 《生物医学工程学报》、《北京生物医学工程》、《人工脏器与透析》三个杂志的编委. 目前还承担着攀登计划、863计划课题任务. 另有专著三本, 论文三十余篇.

陈卉 女, 29岁, 1990年毕业于中国人民大学, 现为首都医科大学生物医学工程系讲师. 目前参加了攀登计划、863计划课题及其它一些科研项目, 主要从事人工神经网络应用及生理系统建模仿真的研究. 在国际会议上发表论文一篇, 国内学术刊物上发表论文两篇.