

# 电脑化 人脑

● 神经网络——第六代计算机(普及)



73.87  
65

# 电脑人脑化

神经网络——第六代计算机

(普及本)

[美] 斯 华 龄



北京大学出版社

9310210

新登字(京)159号

电 脑 人 脑 化 (普及本)

[美] 斯 华 龄 著

责任编辑: 翟 定

\*

北京大学出版社出版发行

(北京大学校内)

北京大学印刷厂印刷

新华书店经售

\*

850×1168 毫米 32 开本 4.125 印张 100 千字

1993 年 3 月第一版 1993 年 3 月第一次印刷

印数:0001—3,000 册

ISBN 7-301-02077-5/TP·172

定价:4.50 元

JS/31/03

### 内 容 简 介

本书作者目前任国际神经网络协会主席。神经网络的研究是借鉴人脑的结构与工作原理以设计和建造具有一定智慧的新一代计算机,即第六代计算机。这是一个正在迅速发展的新兴研究领域。作者曾在欧美各国并专程来华举办系统讲座,介绍有关神经网络的基本知识和最新进展,受到广泛欢迎。本书是在这些讲座的基础上整理而成的,内容包括:神经网络概论,联想记忆神经网络,神经网络的结构与算法,神经网络的学习方式,视觉系统,第六代计算机的体系结构,以及动态与展望等。

0129182

## 目 录

导 言 .....	(1)
第一章 神经网络概论 .....	(4)
第二章 联想记忆神经网络 .....	(22)
第三章 神经网络的结构与算法 .....	(37)
第四章 神经网络的学习方式 .....	(54)
第五章 视觉系统 .....	(89)
第六章 动态与展望 .....	(105)
第七章 第六代计算机的体系结构:神经网络、 超级计算机和光学数据库机 .....	(115)

## 导 言

本书呈献给读者的是一个正在迅速发展的、全新的科学研究领域。神经网络(neural networks)的研究是借鉴人脑的结构与工作原理以设计和建造具有一定智慧的机器。人们希望这种神经网络能够做人所能做的各种事情,例如,通过视觉、听觉或触觉识别事物,把大量的信息浓缩在一起,实现自动化,特别是要它具有学习的能力和能够创造发明。这种神经网络不同于通常的人工智能机,我们称之为第六代计算机。神经网络科学研究的进展无疑会在人类生活的各个领域引起巨大的影响。

在神经网络的研究工作中,要特别强调各种科技人材的通力合作,至少要与生物学、物理学、数学和无线电电子学这四门学科的人紧密结合,并且是与工业、经济等各领域有密切关系的。本书的目的就是希望能吸引更多的科技人材关心这个研究领域,投身到我们的行列中来。

根据目前的认识,神经网络至少应由三个基本部分组成:前端感知器件,相当于眼、耳等;中心处理机,相当于大脑;终端执行机,相当于手、足等。设计神经网络时,应根据每部分的职能分工,分别地实现。每部分的规模也应根据具体问题的需要来确定。神经网络没有统一的结构,必须针对各种任务分别进行设计。

设计神经网络时,无须先对它编程序,只要赋予它一定的学习功能,规定它的学习方式,让它在实际工作中不断学习、逐步适应。神经网络的学习方式应尽量接近人的学习方式。能够进行学习,这是神经网络的一个显著特点。

在本书中,读者可找到下列问题的答案:

- 哪些图像识别问题能用神经网络解决?

9310210

- 神经网络对图像处理、知觉器等有何影响?如何提高容错性(fault tolerance)?
- 神经网络与联结机(Connectionist machines)有何异同?
- 神经网络最少要多少层才能具有一定的智慧?
- 如何确定所需要的层数及每层的单元数?
- 如何选择学习方式?
- 如何根据具体任务确定神经网络的结构?如何确定系统的动力学?
- 在近期及在将来,如何减少建造神经网络的成本?
- 利用大规模集成电路或光学元件制造神经网络的研究现状如何?
- 神经网络研究的国际动态与趋势如何?
- 把神经网络、光学、数字计算机等技术结合起来促进神经网络发展的可能性与前景如何?

以上所有这些方面的问题,将分别在本书第一部分中作系统的阐述.

在结束导言之前,作为一段插曲,笔者想讲一个有关创造发明的故事.由于我们希望神经网络具有创造发明的本领,这个故事并不是题外的话,而是与我们研讨的主题有很密切的关系.

我曾向我的老师,统计物理学家乔治·乌仑贝克(J. Uhlenbeck)请教创造发明的秘诀是什么?创造发明是以什么样的方式进行的?怎样才能训练出更加聪明的学生?乌仑贝克经过长时间的考虑,并和他的老师艾仑菲斯特(Ehrenfest)以及丹尼斯·葛波(D. Gabor)等人讨论之后,提出了以下四个要点.

第一,要发表自己的想法,特别是不要害怕说出第一句话.几百年前,有一个希腊人日复一日地躺在沙滩上看日出日落.到了年终,学校里让他写篇文章介绍一年的成就.在老师的鼓励下,这位希腊人终于写出一篇文章阐述他的一个观点:太阳只有我的脚趾

那样大. 一个偶然的机 会, 另一个希腊人看到了这篇文章, 但他认为太阳只有半个脚趾大. 于是两人发生了激烈的争论. 最后, 经与第三个希腊人共同讨论, 终于得出一致的看法: 两人结论的差异是由身高不同引起的, 太阳的大小只能以其张角来衡量, 说它线度的大小并无意义. 并且, 进而根据阳光下阴影的移动作出太阳绕地球旋转的结论. 从而, 他们发现了天体的运动. 这就是天文学的开始. 由此可见, 敢于说出自己的想法, 不要怕讲错, 这是非常可贵的.

第二, 要注意完整性. 麦克斯韦(Maxwell)把有关电磁现象的四个定律放到一起, 才发现了有关位移电流的规律. 由此, 才能发现电磁波, 才能出现收音机和电视机. 这个典型的例子充分说明了完整性之重要.

第三, 要注意找出对立物, 有阴必有阳, 相反相成. 常常有这样的情形: 当一条路总走不通时, 会发现实际上相反的一面是正确的.

第四, 要重视潜意识的作用. 必须努力工作. 假定每周工作 40 小时, 成果为百分之百, 则若再工作 10 小时, 成果便可达到百分之二百. 这是潜意识在思考、在帮你的忙的结果. 神经网络之奥妙正在于此.

以上这几点, 初看起来似乎浅显, 但放在一起才能看出全局. 这些思想对我的科学道路有重大的影响. 本书也是根据这样的精神编写的. 希望它能激发起读者潜意识中的创造能力, 产生出新的思想、新的成果.



# 第一章 神经网络概论

## 一、大脑与神经网络

人的大脑是至今自然界所造就的最高级产物,研究和制造具有智慧的计算机无疑应该以大脑为模拟对象.

人的大脑约由  $10^{10}$  个神经细胞组成,这是一个天文数字,大体相当于天空中星星的数目. 细胞之间通过树突与轴突互相连接,构成纵横交错的网络结构. 脑细胞通过突触互相交换信息,每个脑细胞平均有  $10^4$  条通路与其它脑细胞连接. 因此,这是一种极其复杂的通讯网.

按传统观点,一个人降生时,就已经具有了全部数量的脑细胞,其数目终生不再增加,随着岁月的流逝它们将成批地死亡. 但是,人们的记忆并不因为脑细胞的减少而逐步丧失,与此相反,人们的智慧却能越来越发展. 那么,奥秘在哪里呢? 原来,人们的记忆和智慧并不是储存在单个脑细胞内,而是储存在脑细胞之间互相连接的网络之中,这称为分布式存储方式. 这样一种“集体智慧”,并不因单个脑细胞的死亡而消失. 人们通过学习可不断地改变脑细胞之间的连接形式,使这个网络的功能不断提高,这就是人的智慧发展的生理学过程.

基于冯·诺依曼思想的传统计算机,其结构与大脑有根本的差异. 传统计算机由存储器和运算器组成,两者之间通过被称作总线的一组连线连接,传递信息. 总线如同非常拥挤的高速公路,其信息通过量及通过率限制了传统计算机的速度. 这种卡脖子现象被称为冯·诺依曼瓶颈. 对比大脑神经网络的结构,不难看出后者的优越性. 神经网络的结构有如大量村镇由纵横交错的乡间

小路连接在一起,这种交通网不易发生交通堵塞,并且可使大量的运输过程同时进行。

在神经网络的研究中,人们借鉴了大脑结构的上述特征,采用大量的、比较简单的元件作为系统的基本单元,依靠单元之间复杂的连接构成具有良好功能的网络,这些连接可以按照一定的方式改变,这就使得神经网络具有一定的学习能力。

目前的神经网络研究仅仅是对大脑作较简单的模拟,大脑的结构和功能当然要复杂得多。目前的神经生理学研究已经发现神经元及其间的连接(突触)至少有以下几种不同的行为。

神经元的生物学行为:

- (1) 可处于两种不同的状态:抑制与兴奋;
- (2) 可出现爆发和平台两种情形;
- (3) 可出现抑制后的反冲;
- (4) 具有适应性。

突触的生物学行为:

- (1) 可进行信息综合;
- (2) 可出现渐次变化的传输;
- (3) 可有电接触及化学接触等多种连接方式;
- (4) 可出现延时激发。

人工神经网络模拟大脑不应是简单的仿效,更不可能让人工神经网络包含大脑的全部功能和活动。例如,不一定要求人工神经网络保留新陈代谢行为,但大脑中对产生智慧有本质意义的结构或属性则应尽量吸取。上述神经元及突触的生物学行为看来对大脑的智慧都是很重要的。目前人工神经网络的研究只对神经元的第一种行为和突触的第一种行为进行了模拟,其余几条均未涉及。这是研究工作的发展阶段所决定的,只能一步一步走。但大脑极其复杂的结构和功能为神经网络的研究指示出大有可为的广阔前景。总起来讲,神经网络应能解决人所能解决的各种工作,包括数

据处理、知识加工及自动控制等各个方面。

## 二、分布式存储与容错性

容错性是指根据不完全的、有错误的信息作出正确、完整结论的能力。大脑的容错性是它的一种极为重要的智慧形式。

传统计算机的容错性是很差的，若干元件的损坏、几条指令或几个数据的错误就可能导致整个计算工作无法正常进行。人的大脑则有高度的容错性，我们常常能够仅由某人的一双眼睛、一个背影、一个动作或一句话的音调，便能把这个人认出来。大脑的容错性与大脑的分布式存储的组织方式有极密切的关系。

人工神经网络采用分布式存储，因而有较好的容错性。图 1.1

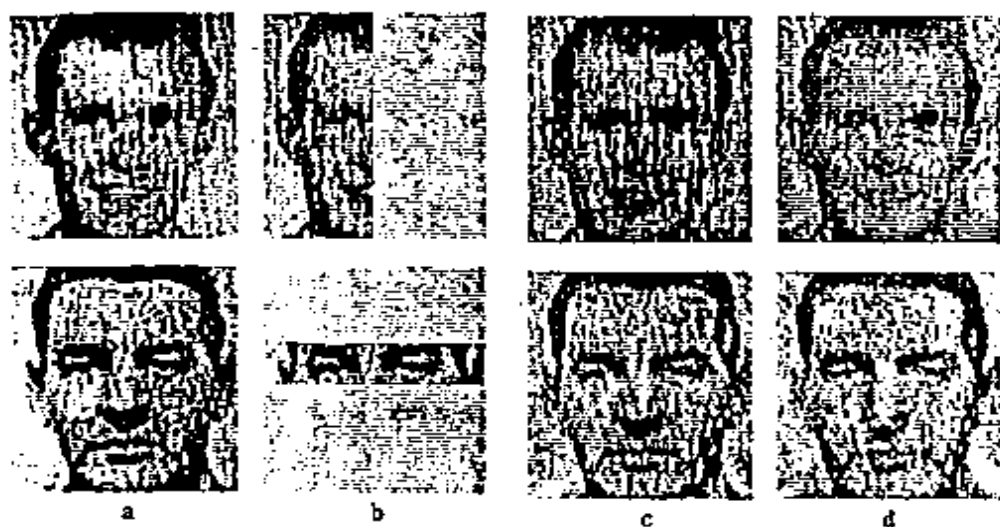


图 1.1 联想记忆实验结果

是用神经网络模拟的从部分信息恢复完整记忆的一个实验结果。图 1.1(a)是已经存储在网络中若干幅图当中的两幅图，图 1.1(b)是对应于(a)中每幅图的一部分，作为不完整的输入信息。图

1.1(c)是总共预存储 100 幅图像时,网络根据不完整输入而得到的输出.图 1.1(d)是预存储 500 幅图像时的相应输出.可以看到,所得到的输出图像相对于存储的标准图像有一定的畸变,但大体上还是恢复了完整的记忆.

熟悉全息现象的读者可能会发现这里所描述的容错性与全息有某种相似之处.全息底片也是采用分布式存储方式,由底片的一部分可以大体恢复原物的完整形象.全息技术的发明者,诺贝尔奖获得者,D. 伽柏(D. Gabor)曾经提出大脑的作用就像是一幅全息图.这种观点后来受到另外两位诺贝尔奖获得者,沃森和克瑞克的批评.今天的研究表明,大脑的作用确实是一幅全息图,不过它不像光学中的傅里叶全息那样存在于频域中,而是存在于空域(Co-ordinate Space)中.

现在,我们以一个最简单的神经网络模型为例,演示它如何实现分布式存储和它的容错性.考虑一个由两层神经元组成的网络,每层有  $N = 3$  个神经元,一层为输入层,一层为输出层,两层之间有 9 条连接线使两层单元彼此互相连接.输入和输出均由 3 分量矢量表示,其分量取值为 0 或 1.第  $i$  个输入单元与第  $j$  个输出单元之间的连接强度记为  $w_{ji}$ ,全部连接可由  $3 \times 3$  的矩阵  $W$  表示.输入矢量  $s^{\wedge}$  与输出矢量  $s^{\text{出}}$  之间的关系可表达为

$$s^{\text{出}} = \theta(W \cdot s^{\wedge}) \quad (1.1)$$

或

$$s_j^{\text{出}} = \theta\left(\sum_{i=1}^3 w_{ji} s_i^{\wedge}\right), \quad (1.1')$$

其中,  $\theta$  是一种非线性函数,在不同问题中可取相应的适当形式.这里,取成简单的阶跃函数:

$$\theta(u) = \begin{cases} 1, & \text{当 } u > 0; \\ 0, & \text{当 } u \leq 0. \end{cases} \quad (1.2)$$

神经网络所记忆的内容储存在各神经元间的连接网络中,这

意味着它的记忆内容由连接矩阵  $W$  表示. 连接矩阵  $W$  可以通过适当的学习过程建立. 一种比较简单的学习方式是由待存储态矢量的外积构成连接矩阵. 若  $s$  是待存储的态矢量, 则这时的连接矩阵为

$$W = ss^T = \begin{bmatrix} s_1s_1 & s_1s_2 & s_1s_3 \\ s_2s_1 & s_2s_2 & s_2s_3 \\ s_3s_1 & s_3s_2 & s_3s_3 \end{bmatrix}. \quad (1.3)$$

这里,  $s^T$  是矢量  $s$  的转置,  $ss^T$  是并矢运算. 如果有多个矢量待存储, 则要求网络依次学习, 数学上表达为将每一个形如 (1.3) 式的外积相加而构成连接矩阵.

例如, 若存入一个状态  $x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ , 则连接矩阵为

$$W = xx^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (1.4)$$

在这种情况下, 当输入为  $x$  时, 输出亦为  $x$ . 这种输出等于输入的状态称为不动点. 存储态为不动点表示, 当以正确完整的记忆作为输入时, 神经网络可给出正确的存储态输出. 如果输入的态与  $x$  不完全一致, 而是相对于完整的记忆内容有一些偏离, 仍然可能得到正确的存储态输出. 例如, 若输入的是对于  $x$  有 1 个或 2 个分量的偏差的态:

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}; \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}; \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

由 (1.1) 式易看到输出仍为  $x$ . 这就是容错性, 它是由连接矩阵 (1.4) 式的形式决定的.

网络的容错能力与存储状态的数目有关. 当存入 2 个态:

于  $N = 3$  的情形, 状态空间  $2^N = 8$  个可能状态中只有 3 个互相独立的状态. 前面写出的  $x, y$  和  $z$  就是这样一组互相独立的状态, 其中只有  $1/3$  可用作存储态, 而保留另外  $2/3$  的态不作为存储态, 以保证网络具有较好的容错性. 存储量增加, 容错性随之变差, 这是由于所存储的状态之间互相干扰造成的.

如果仔细考察上面的例子, 会发现更复杂的问题. 表 1.1 中列出整个状态空间中的流向图, 即分别以状态空间中  $2^N = 8$  个态之一作为输入, 考察其输出情况. 由表 1.1 (见 12 页) 可以得出几点结论: (1) 存储量增加时, 容错性变差; (2) 存储量增加时, 某些

存储状态可能被破坏, 如表中第 5 和 6 两种存储状态中,  $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$  态

不再是不动点, 已从记忆中删除掉, 这种被删掉的记忆态称为不稳定记忆态; (3) 稳定性与容错性不仅与存储量有关, 而且与存储态的分布有关, 将第 2 例与第 5 和 6 两例比较即可得此结论; (4) 确定网络容量时, 首先要保证记忆的稳定, 同时要根据实际需要来确定对容错性的具体要求, 因为不同的容错程度所允许的存储量是不同的.

现在再举一个例子. 设状态矢量的每个分量可取 1 或 -1. 考虑  $N = 4$  情形, 输入层与输出层合并为一层, 存储的知识以单元间

互相连接的强度表示. 令存储状态为  $s = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ , 则连接矩阵为

$$W = \begin{bmatrix} 0 & 1 & -1 & 1 \\ 1 & 0 & -1 & 1 \\ -1 & -1 & 0 & -1 \\ 1 & 1 & -1 & 0 \end{bmatrix}. \quad (1.7)$$

注意,这里连接矩阵是由存储态取外积  $ss^T$  后再令对角元为零而得到的.这在物理上表示每个神经元与自身的连接强度为零,不进行自我对话.在计算上,当维数  $N$  较大时,零对角元有方便之处.同时,对角元为零时,若存储互相正交的状态矢量,存储量可达  $M/N = 1$ ,而若对角元不为零,即使是正交状态矢量,存储量也只能达到  $M/N = 1/2$ .可以证明,当  $M/N > 1/2$  时,容错能力完全消失.

连接矩阵(1.7)式所描写的网络,其所存知识仍靠内积读出,只是输出信号时所使用的非线性函数略有差别,取为

$$\sigma(u) = \begin{cases} 1, & \text{当 } u > 0, \\ -1, & \text{当 } u \leq 0. \end{cases} \quad (1.8)$$

以上介绍的用外积写入知识的方法只是最简单的一种,相当于均匀分布,此外,还可采用加权外积的方法,即非均匀分布,或者,采用扩充空间的外积方法,这相当于采用隐单元的情形,这些在此不作详细介绍.

在传统计算机中,采用局域地址记忆(LAM)有其不方便的地方,20多年来,发展了内容定址记忆(CAM)的方式,这是很了不起的成就.这些记忆都是一维的知识表示,而神经网络的记忆则采用二维知识表示,这是神经网络与前面所述诸种方法在数学表达上的差异.但局域地址记忆和内容定址记忆也可视为二维矩阵表示的特殊情形.实际上,局域存储和非局域存储都是重要的,其用途不同,不能互相代替,而是互相补充的.

### 三、什么是神经网络

为了使神经网络能模拟大脑的一部分智慧,应确定神经网络的基本属性,换言之,应给出一个基本的定义.

我们认为,人工神经网络应具有以下基本属性:(1)非线性;(2)非局域性;(3)非定常性;(4)非凸性;(5)非 ? 性.

表 1.1 不同存储状态下的连接矩阵及输入态与输出态的对应关系

存储态	M	W	输入→输出
1	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$
			$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$
2	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$
			$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$
			$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$
			$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$



3	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$	3	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	8个输出态均分别与相应的输入态相同
4	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	1	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$
5	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$	2	$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$
6	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	2	$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

## 1. 非线性

非线性关系是自然界诸规律的普遍特征,线性关系往往只是为简化问题而作的局部近似,世间许多奇妙的事物都源于非线性. 大脑的智慧就是一种非线性现象. 在前面所介绍的几个人工神经网络的例子中,基本单元模拟了脑细胞的一种生物学属性,即它可处在抑制或激发两种不同的状态. 这种二态行为在数学上表现为一种非线性关系. 这种非线性由神经元的输出与输入信号强度比来表现:小输入信号受较大倍数的放大后输出;较大的输入信号受较小倍数的放大后输出;强输入信号不经放大或衰减后输出,当其超过某个阈值时以饱和值输出.

对某个神经元  $j$ ,若来自其它神经元  $i$  的输入为  $x_i$ ,相应连接强度为  $w_{ji}$ ,则神经元  $j$  的输出通常可表达为

$$y_j = f\left(\sum_{i=1}^B w_{ji} x_i - \xi\right), \quad (1.9)$$

其中,  $f$  是一个非线性函数,常数  $\xi$  称为激励阈值. 图 1.2 中给出了  $f$  函数的几种常用的形式.

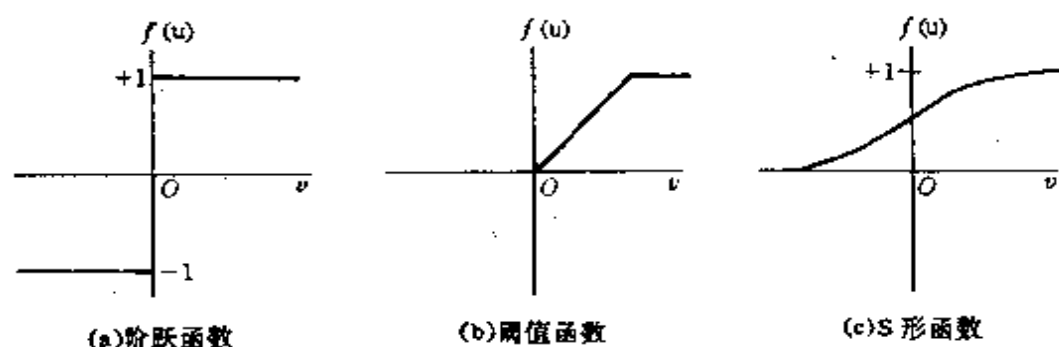


图 1.2 神经元的输入与输出的非线性关系

在某些神经网络模型中,单元的总输入信号与其它单元的輸出信号之间也采用非线性关系,即单元之间采用复杂的非线性连

接而不是简单的线性连接. 这种网络具有更好的性能, 例如可削弱存储态之间的干扰, 从而提高容错性和存储容量等.

## 2. 非局域性

非局域性是自然界中事物间普遍联系的一种表现. 一个系统的许多整体行为不仅取决于系统单元的个性, 而且可能主要由单元之间的相互作用、相互连接所决定. 以水结冰为例, 当水温降低到冰点时水会结冰, 这种现象并不是单个水分子的效应, 而是由水分子之间的相互作用所决定的集体行为. 大脑的智慧是大脑的整体行为, 决定于神经细胞的整个连接网络, 这就是大脑的非局域性特征.

在人工神经网络的设计中, 不再采用传统计算机的局域性记忆方式, 而是以单元之间的大量连接模拟大脑的非局域性. 这些连接通常是可变的, 使神经网络能够进行学习和训练.

前面所介绍的联想记忆神经网络是展示非局域性的典型例子, 它们采用分布式存储, 记忆内容存储在连接网络中. 由输入的局部信息或畸变信息可得存储在网络中的正确的、完整的信息, 这种联想记忆方式称为自联想记忆. 另一种联想记忆方式是异联想记忆. 这种记忆方式是存储两组不同形式、不同内容的样板及两组样板之间的对应关系. 当输入第一组中某样板的部分或畸变的信息时, 可以得到第二组中对应样板的正确、完整的信息输出.

异联想记忆的一个例子是触觉机器手. 手的表面排列大量探测器, 可得到大量输入数据, 构成很长的输入矢量. 我们感兴趣的并不是全部探测到的细节, 而仅是有关整体性质的少数特征参量, 如速度、表面光滑程度等, 因而所要求的输出是只包含几个分量的短矢量. 这种问题若用传统计算机处理, 则要进行大量计算, 难以做到实时处理. 在人工神经网络中, 可由几组典型的数据矢量及相应的信息矢量的外积构成狭长的并矢, 作为连接矩阵, 由实测的数据矢量与连接矩阵作内积运算便可得到相应的信息输出. 这是一

种大规模并行处理极大量数据的方法,效率很高,显示了神经网络非局域性的优点,具有实时性和容错性.

### 3. 非定常性

宇宙万物处于永恒的运动之中,大脑的思维活动也在不断地变动演化.神经网络要具有智慧、能够创造发明,就不能处于定常状态,而必须是能在某种程度上模拟思维运动的一种动力学系统.

神经网络的动力学过程的一种典型的实现方式,是采用迭代过程.在联想记忆模型中,系统的状态  $x(t)$  按下述动力学方程演化:

$$x(t+1) = f(W \cdot x(t)). \quad (1.10)$$

这里,  $t$  是间隔为 1 的离散时间,  $W$  是连接矩阵,  $f$  是非线性函数, 状态  $x$  是一个矢量,  $W \cdot x$  是内积运算.  $x(t+1)$  是  $t$  时刻的输出矢量, 并作为  $t+1$  时刻的输入矢量. 于是出现以迭代过程描写的动力学演化过程.

通常, 迭代过程将趋向于状态空间的一个不动点, 即满足方程

$$x = f(W \cdot x) \quad (1.11)$$

的状态. 当  $M/N$  足够小时, 即存储量很小时, 诸存储态均为不动点. 若从一个与标准态有某些偏离的状态出发, 经过一系列迭代过程, 终将趋向于这个存储态. 状态空间中, 在每一个存储态周围都有一个确定的区域, 凡以此区内的态为初始态, 经迭代过程, 必将趋向原存储态. 这样的区域称为该存储态的吸引域. 吸引域的大小和形状定量地刻画了网络的容错性. 从吸引域中一状态出发, 趋向不动点的过程则描述了自联想记忆的搜索过程, 当存储量增大时, 将会出现各种复杂的情况. 例如, 非存储态的不动点增多, 存储态的吸引域缩小、边界变得参差不齐, 以及存储态失稳即存储态不再是不动点而从记忆中消失等.

### 4. 非凸性

世界万物丰富多采, 而且越变越复杂, 这与非凸性有密切关

系. 一个系统的演化方向, 在一定条件下将取决于某个特定的状态函数, 例如能量, 函数的极值相应于系统比较稳定的状态. 非凸性是指这种函数有多个极值, 故系统具有多个较稳定的平衡态, 这将导致系统演化的多样性. 例如, 已经发现雪花有几百种结构, 每种结构对应于能量函数不同的极小值, 但出现哪种结构的雪花则与形成过程中的降温方式有关.

优化问题一般也是非凸性的, 即评估函数有多个极值. 在一个局域范围内可以找到一个局域极值, 这是局部优化. 全局优化则需在全局范围内找到最佳极值. 要实现全局优化必须防止系统落入局部极值中, 如能有效地克服局部极值的束缚, 便可解决全局优化问题. 解决全局优化问题的典型方法是模拟退火方法. 这种方法的精神是: 允许系统在演化过程中有一定的概率向任何方向随机变动, 逐步减小这种随机变动的概率, 系统终将落入全局极值. 这种过程相当于统计分布的演化过程. 通常的材料退火过程大抵与此相仿, 故而得名.

适有童话一则可引喻退火过程. 昔有小盒一只, 其内底部凹凸不平如山峦起伏, 于最低谷处有一小孔与外界相通, 并有花生仁若干匿于盒中低洼处未能由深谷小孔漏出. 有猴欲食花生而急摇之, 诸花生飞漱盒间无几落入深谷漏出. 猴怠, 遂有几枚花生仁由底孔漏出. 猴大喜, 急摇而渐缓其势, 终饱一餐.

在神经网络进行学习或训练, 以改变内部连接时, 也可以定义某种评估函数以描述学习效果. 这种评估函数一般也是非凸性的. 因此, 神经网络的学习过程也是一种全局优化问题, 也要避免系统落入局部极值, 故而可采用模拟退火方法.

G. E. Hinton, T. J. Sejnowski 和 D. H. Ackley 等人提出的模拟退火方案称为玻耳兹曼机, 其概率分布采用高斯分布. 这种方法速度较慢. 作者曾提出一种快速模拟退火方法, 称为柯西机, 采用柯西分布作为概率分布函数. 这种方法的有关内容将在第三章中

作系统介绍.

#### 5. 非 ? 性

我们把神经网络定义的最后一条用一个疑问号表达,其目的是强调迄今人们对大脑思维过程以及如何用人工神经网络进行模拟的问题了解仍然很肤浅,研究还很不充分,我们所面临的是一个充满未知的新领域.例如,关于神经网络如何进行学习的问题还没有得到令人满意的答案.因此,神经网络的研究不仅有许多技术上的难点有待克服,而且根本的问题是在基本原理方面也还必须在更开阔的视野上进行更深刻的探索.

### 四、设计神经网络的要点

我们希望神经网络能做人所能做的各种事情,如传感器数据处理,知识加工和自动控制等.传感器数据处理包括自然语言识别、图象识别和某些军事用途;知识加工包括自动知识提取、并行逻辑推理和假设判断等;自动控制包括机器人控制、适应机控制和优化过程控制等.为了使神经网络具有这些方面的智慧,在设计时必须解决下述几个关键问题.

#### 1. 确定神经网络的规模

人的脑神经细胞约有  $10^{10}$  个,细胞间的连接约有  $10^{10} \times 10^4$  条.就目前的工艺水平,还无法制造如此复杂的神经网络.采用现今可行的工艺方法,可以制造每层包含几百到上千个单元的神经网络.目前可行的两种工艺均有其缺点和各自的长处.光学神经网络,其通讯通过光路联系,较方便,但单个神经元体积较大.电学神经网络,其集成度很高,但其间的连接线很复杂.若有  $N$  个单元,则需  $N^2$  条连接线,这一困难称为“连接线瓶颈”难题.

就目前的工艺水平而言,所能制造的神经网络绝对无法与人脑相比拟.因此,人工神经网络只能是一种智慧很低的系统.低智能的系统完全能够很优秀地工作在适合于它的领域.如果能对网

神经网络研究的新潮流. 这是两个联系紧密又有原则区别的研究方向, 有必要在此对两者作一简单比较.

人工智能系统是指在传统计算机硬件基础上发展的图灵过程软件, 以使传统计算机更加有用, 更有智慧. 通常所说的人工智能系统主要是指机器人和各种专家系统. 人工智能系统被称为第五代计算机.

人工神经网络则是第六代计算机, 是把算法和结构统一为一体的系统, 这是一种硬件与软件的混合体, 由于它模拟大脑的结构, 故也可称为湿件(wetware). 这种系统有更高的智慧并可能有更快的计算速度.

人工智能与神经网络的运算逻辑实际上是相同的. 例如, 识别事物均采用等价类模型. 以识别苹果为例, 当水果、红色及美味三要素重叠时, 便判定为苹果.

神经网络与人工智能在鼓作原理上的区别可概括为以下两点:

(1) 知识表象不同. 人工智能用一维串表示知识, 而神经网络则用二维或高维矩阵表示知识. 高维知识表示有下列优点:

第一, 由于自由度增加, 使存储空间扩大, 因而有更大的存储容量;

第二, 高维空间中每一状态有更多的近邻, 使多体效应更加复杂和显著, 这使容错性大大提高. 另外, 多维空间会有相变, 这对知识的存储和学习有重要意义;

第三, 高维空间更易于分类, 分类就是识别事物, 其原则是使同类事物更加聚集、异类事物更加分离, 这在高维空间中更容易实现.

(2) 学习方式不同. 人工智能要求预编程序, 它是在指定规则的基础上进行工作的. 这种系统称为闭集, 只能进行推论; 神经网络无须预先编程, 不必制定工作规则, 而是通过学习建立和改变知

识.它具有推广和抽象能力.这种系统称为开集,可以进行引证.

虽然神经网络比人工智能有更高的智慧,但并不抹煞人工智能的功用,因为一旦查明大部分规则,那么用人工智能系统工作会更有效率.神经网络与人工智能有如人之左右手,由神经网络探索规律,然后人工智能依此规律去办事,二者相辅相成.

## 六、神经网络研究的历史概况

早在 40 年代就已出现了神经网络的初步设想. 50 年代, Rosenblatt 设计了知觉器,当时以为这已解决了模拟大脑的全部问题,过分乐观了一阵,后来人们发现了很多当时无法解决的困难.与此同时,出现了人工智能的研究,成为主流,而神经网络的研究一直停滞不前.近 30 年来,一直有少数人仍坚持研究神经网络.近年来,一方面人们已经认识到传统计算机的局限性,另一方面由于神经网络的研究取得了一些重要成就,人们开始把注意力逐渐投向神经网络的研究.当今世界上正在悄悄地酝酿着一个新的突破.我们确信,神经网络研究的进展必将对整个人类社会带来广泛而深刻的影响,因此必须给予充分的重视.同时,还应当吸取历史的教训,始终保持清醒的头脑,对神经网络研究的现状、神经网络与传统计算机及其它科学技术的关系等给予适当的评价.只有这样,才能保证神经网络的研究工作顺利发展.



## 第二章 联想记忆神经网络

### 一、联想记忆神经网络的特点

在第一章中,初步介绍了如何用神经网络实现联想记忆.我们指出,这是一种分布式存储的记忆方式,最简单的方法是由矢量的外积构造连接矩阵,记忆便存储在这个连接矩阵中;记忆的搜索过程则是通过输入矢量与连接矩阵的内积以及输出单元对信息的非线性加工而实现的.神经网络的联想记忆有一定的容错性,这与它的非局域存储方式有密切关系.在本章中,将更详细地讨论神经网络联想记忆的特点,它的容错性与存储容量,以及如何用它来进行图像识别.

把神经网络用于联想记忆至少有以下优点:

(1) 如果已完全确定了标准输入矢量及相应标准输出矢量,则可采用非常简单的神经网络即静态网络实现联想记忆.这种网络中单元间的连接是固定不变的,由矢量外积构成的连接矩阵确定,可以不必再反复进行学习.这种方式虽然十分简单易行,但仍可达到相当的容错性.当然,在必要时,也可让神经网络能够学习,通过学习调整连接矩阵,以达到所要求的联想记忆功能.

(2) 这种联想记忆方式无须对输入矢量进行预处理,可以直接进入搜索,省去了编码和解码工作.而且,实现联想记忆并不要先找出输入、输出矢量之间的内在关系.例如,如果我们不了解张三与李四的关系,但要求由张三能联想到李四,则只要在连接矩阵中加上张三与李四的外积,便可达到目的.

(3) 这种联想记忆采用并行处理的方式,将输入矢量的全部信息同时进行加工,当信息量很大时,这种大规模的并行处理比通

常的串行处理效率高得多,也就易于达到实时性的要求.

## 二、联想记忆神经网络的容错性

假定状态矢量分量的可能值为 1 或 -1. 我们将引入一种简化的符号表示状态矢量:把矢量的分量看作一个二进制数的各位数,当某分量为 -1 时,取二进制的相应位数的值为 0,否则取为 1,并以相应的十进制数作为该矢量的名称. 例如

$$\begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix} = |13\rangle, \quad (2.1)$$

因为  $1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 13$ . 在(2.1)式右端用符号  $|$  表示列矢量,称为右矢. 将列矢量转置为行矢量时,用符号  $\langle$  表示,称为左矢:

$$\begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}^T = (1, 1, -1, 1) = \langle 13|. \quad (2.2)$$

矢量的外积记为  $|13\rangle\langle 13|$ , 构成一个矩阵.

考虑存储一个状态矢量  $|13\rangle$  的自联想记忆情形, 并取连接矩阵的对角元为 0, 则连接矩阵为

$$[13] = |13\rangle\langle 13| - [\delta] = \begin{pmatrix} 0 & 1 & -1 & 1 \\ 1 & 0 & -1 & 1 \\ -1 & -1 & 0 & -1 \\ 1 & 1 & -1 & 0 \end{pmatrix}, \quad (2.3)$$

其中 $[\delta]$ 是分量为

$$\delta_{ij} = \begin{cases} 1, & \text{当 } i = j, \\ 0, & \text{当 } i \neq j \end{cases}$$

的矩阵, 输出信号采用的非线性函数取为

$$\sigma(u) = \begin{cases} 1, & \text{当 } u > 0, \\ -1, & \text{当 } u \leq 0. \end{cases} \quad (2.4)$$

现考察这个自联想记忆模型的性能.

当输入矢量为 $|13\rangle$ 时, 输出仍为 $|13\rangle$ :

$$[13]|13\rangle = |13\rangle, \quad (2.5)$$

这表明所存储的状态 $|13\rangle$ 是状态空间中的不动点.

当一个状态矢量的全部分量均取相反符号时, 所得到的矢量代表的状态称为原状态的互补态. 易看出,  $|13\rangle$ 的互补态矢量为 $|2\rangle$ . 我们发现

$$[13]|2\rangle = |2\rangle, \quad (2.6)$$

即 $|2\rangle$ 也是一个不动点. 当存入一个状态矢量时, 其互补态矢量也同时成为存储矢量. 事实上, 容易证明一个状态矢量构成的连接矩阵与其互补态矢量构成的连接矩阵相同. 例如, 有

$$[2] = [13].$$

现考察上述自联想记忆模型的容错性.

先考虑将标准矢量的若干分量取为 0 作为输入矢量的情形. 将一个分量取为 0, 例如:

$[13](0, 1, -1, 1)^T = (3, 2, -2, 2)^T \longrightarrow (1, 1, -1, 1)^T = |13\rangle$ ,  
这里, 非线性运算以 $\longrightarrow$ 表示. 两个分量取为 0 时, 有

$[13](0, 0, -1, 1)^T = (2, 2, -1, 1)^T \longrightarrow (1, 1, -1, 1)^T = |13\rangle$ .  
三个分量取为 0 时, 有

$$[13](0, 0, 0, 1)^T = (1, 1, -1, 0)^T \longrightarrow |12\rangle,$$

而

$$[13]^2(0, 0, 0, 1)^T \equiv [13]|12\rangle \longrightarrow |13\rangle.$$

其中,  $[13]^2$  表示  $[13]$  连续作用两次, 即以一次作用所得的输出  $|12\rangle$  再作为输入矢量. 四个分量全部为 0 时, 有

$[13](0,0,0,0)^T = (0,0,0,0)^T \longrightarrow (-1,-1,-1,-1)^T = |0\rangle$ ,  
并有

$$[13]^2(0,0,0,0)^T \longrightarrow |2\rangle,$$

$$[13]^3(0,0,0,0)^T \longrightarrow |2\rangle.$$

即四个分量全部为 0 的矢量作为输入时, 经两次迭代, 输出为  $|13\rangle$  的互补态  $|2\rangle$ . 这个结果与非线性函数 (2.4) 式的定义有关, 当定义改为  $f(0) = 1$  时, 则两次迭代后输出为  $|13\rangle$ .

现考虑输入矢量的若干分量与  $|13\rangle$  有相反符号的情形. 一个分量不同, 有

$$[13](-1,1,-1,1)^T = [13]|5\rangle = (3,1,-1,1)^T \longrightarrow |13\rangle. \quad (2.7)$$

两个分量不同, 有

$$[13](-1,-1,-1,1)^T = [13]|1\rangle = (1,1,1,-1)^T \longrightarrow |14\rangle, \quad (2.8)$$

$$[13]^2(-1,-1,-1,1)^T = [13]|14\rangle = (-1,-1,-1,1)^T \longrightarrow |1\rangle. \quad (2.9)$$

三个分量不同, 有

$$[13](-1,-1,1,1)^T = [13]|3\rangle = (-1,-1,1,-3)^T \longrightarrow |2\rangle. \quad (2.10)$$

四个分量不同, 有

$$[13](-1,-1,1,-1)^T = [13]|2\rangle = (-1,-1,1,-1)^T \longrightarrow |2\rangle. \quad (2.11)$$

以上表明, 当输入与存储矢量  $|13\rangle$  有一个分量不同的矢量时, 输出仍为  $|13\rangle$ ; 当输入的矢量与  $|13\rangle$  有两个分量不同时, 出现周期为 2 的循环:

$$|14\rangle \rightleftharpoons |1\rangle, \quad (2.12)$$

而二者互为互补态： $|1\rangle = -|14\rangle$ ，还可出现循环

$$|7\rangle \rightleftharpoons |8\rangle = -|7\rangle \quad \text{及} \quad |4\rangle \rightleftharpoons |11\rangle = -|4\rangle$$

等；当输入矢量与 $|13\rangle$ 有三个不同分量时，输出为 $|13\rangle$ 的互补态 $|2\rangle$ ；当输入矢量全部四个分量均与 $|13\rangle$ 不同时，即为 $|13\rangle$ 的互补态 $|2\rangle$ ，是不动点。

从以上的讨论可以看出，当输入矢量是将存储矢量中若干分量变为零时，其容错性高于将这些分量变为其相反值的情形。因此，当应用联想记忆神经网络时，如果不能确定输入矢量的某些分量，最好的办法是令它们为 0，如果任意赋予 1 或 -1，其效果会更差。

以下，对(2.7)~(2.12)式作进一步的讨论。在这个例子中， $|13\rangle$ 与 $|2\rangle$ 是两个不动点，二者的全部分量的符号均相反，故 $|2\rangle$ 是状态空间中与 $|13\rangle$ 相距最远的状态。在状态空间中，与 $|13\rangle$ 有一个分量不同的状态与 $|2\rangle$ 有三个分量不同，输入这种矢量时输出为 $|13\rangle$ ；而与 $|13\rangle$ 有三个分量不同的状态与 $|2\rangle$ 只有一个分量不同，输入这种矢量时输出为 $|2\rangle$ 。这表明，在状态空间中，不动点 $|13\rangle$ 与 $|2\rangle$ 附近均存在一个区域，该区域中的点将“流向”相应的不动点。这种区域称为相应不动点的吸引域，而这个不动点则称为不动点吸引子。

与 $|13\rangle$ 有两个分量不同的状态与 $|2\rangle$ 的差别也是有两个分量不同，它们与两个吸引子 $|13\rangle$ 与 $|2\rangle$ 是等距离的，处于两个吸引域的分界处。这种状态既不被吸引到 $|13\rangle$ ，也不被吸引到 $|2\rangle$ ，(2.12)式表明，它们与其互补态之间构成周期为 2 的循环。事实上，这些状态是与 $|13\rangle$ 或 $|2\rangle$ 正交的状态，即它们与 $|13\rangle$ 或 $|2\rangle$ 的内积为零，如 $\langle 13|14\rangle = 0$ 。容易看出，维数  $B$  为偶数时，两个矢量若有一半的分量互不相同，则二者的内积为 0 且互相正交。在上面  $B = 4$  的例子中，有两个分量互不相同的矢量是互相正交的。

现在，我们给出一个一般性的定理。假定在自联想神经网络

中存储  $M$  个状态矢量  $|A_m\rangle, m = 1, 2, \dots, M$ , 连接矩阵取为

$$[W] = \sum_{m=1}^M |A_m\rangle\langle A_m| - M[\delta]. \quad (2.13)$$

则对任一状态矢量  $|x\rangle$ , 有下述定理:  $|x\rangle$  与全部存储矢量  $|A_m\rangle (m = 1, 2, \dots, M)$  正交的充分必要条件是

$$\begin{aligned} f([W]|x\rangle) &= -|x\rangle, \\ f([W](-|x\rangle)) &= |x\rangle, \end{aligned} \quad (2.14)$$

其中,  $f$  函数的定义见(2.4)式. (2.14)式意味着, 状态矢量  $|x\rangle$  及其互补态  $-|x\rangle$  构成由连接矩阵  $[W]$  决定的周期为 2 的循环.

利用这个充分必要条件可以有效地判断一个新的矢量与全部存储矢量之间的正交性.

这里, 只给出(2.14)式作为正交性的必要条件的证明. 假定  $|x\rangle$  与全部  $|A_m\rangle$  正交, 即  $\langle A_m|x\rangle = 0$ , 则有

$$[W]|x\rangle = \sum_{m=1}^M \langle A_m|x\rangle |A_m\rangle - M|x\rangle = -M|x\rangle,$$

由此可直接得到(2.14)式.

现给出几个例子, 对维数  $N = 4$  的矢量, 有

$$f([13]|11\rangle) = -|11\rangle, \quad f([13](-|11\rangle)) = |11\rangle,$$

$$f([13, 11]|7\rangle) = -|7\rangle, \quad f([13, 11](-|7\rangle)) = |7\rangle,$$

$$f([13, 11, 7]|14\rangle) = -|14\rangle, \quad f([13, 11, 7](-|14\rangle)) = |14\rangle.$$

这里,  $[13, 11]$  表示存储两个矢量  $|13\rangle$  与  $|11\rangle$  时的连接矩阵, 而  $[13, 11, 7]$  表示存储  $|13\rangle, |11\rangle$  与  $|7\rangle$  时的连接矩阵. 运用上述定理, 易看出

$$|14\rangle = (1, 1, 1, -1)^T,$$

$$|13\rangle = (1, 1, -1, 1)^T,$$

$$|11\rangle = (1, -1, 1, 1)^T,$$

$$|7\rangle = (-1, 1, 1, 1)^T,$$

是一组互相正交的矢量; 同时, 其互补矢量

$$\begin{aligned}
|1\rangle &= -|14\rangle = (-1, -1, -1, 1)^T, \\
|2\rangle &= -|13\rangle = (-1, -1, 1, -1)^T, \\
|4\rangle &= -|11\rangle = (-1, 1, -1, -1)^T, \\
|8\rangle &= -|17\rangle = (1, -1, -1, -1)^T,
\end{aligned}$$

也组成一组正交的矢量。

当存储矢量互相正交时,可以使神经网络联想记忆的存储容量显著提高.令矢量维数为  $N$ ,共存储  $M$  个状态矢量  $|A_m\rangle$ ,  $m = 1, 2, \dots, M$ ,其连接矩阵由(2.13)式给出,存储容量以  $M/N$  衡量.确定容量的首要标准是保证所有存入的状态矢量都是局域稳定的不动点,即当某个存储矢量或对它有微小差别的矢量作为输入时,得到这个存储矢量的正确、完整的输出。

考察将一个存储状态矢量  $|A_n\rangle$  作为输入矢量的情形,连接矩阵  $[W]$  ((2.13) 式) 与  $|A_n\rangle$  的内积为

$$\begin{aligned}
[W]|A_n\rangle &= \sum_{m=1}^M \langle A_m | A_n \rangle |A_m\rangle = M|A_n\rangle \\
&= N \left[ \left(1 - \frac{M}{N}\right) |A_n\rangle + \sum_{m \neq n} \frac{\langle A_m | A_n \rangle}{N} |A_m\rangle \right]. \quad (2.15)
\end{aligned}$$

这里利用了  $\langle A_n | A_n \rangle = N$ , 即一个矢量与其自身的内积为  $N$ 。

为了保证  $|A_n\rangle$  是不动点, (2.15) 式右端的每一个分量均应与  $|A_n\rangle$  的相应分量符号相同. (2.15) 式右端由两项组成, 当  $M/N < 1$  时, 第一项能够满足上述要求, 第二项是  $|A_n\rangle$  与其它存储矢量互相“重叠”而引起的干扰效应, 因若两个矢量的内积  $\langle A_m | A_n \rangle \neq 0$  则表示这两个矢量间有互相“重叠”的部分. 假定所存储的矢量在状态空间中是随机分布的, 则其互相干扰的效应类似于某种静态噪声, 由于它们共由  $M - 1 \approx M$  项组成, 用统计的方法可估计出这种随机干扰的相对量级为  $\sqrt{M/N}$ . Hopfield 曾给出容量的量级为  $M/N \sim 0.15$ , 这相当于要求存储矢量间的互相干扰引起的相对误差限制在  $1/3$  左右。

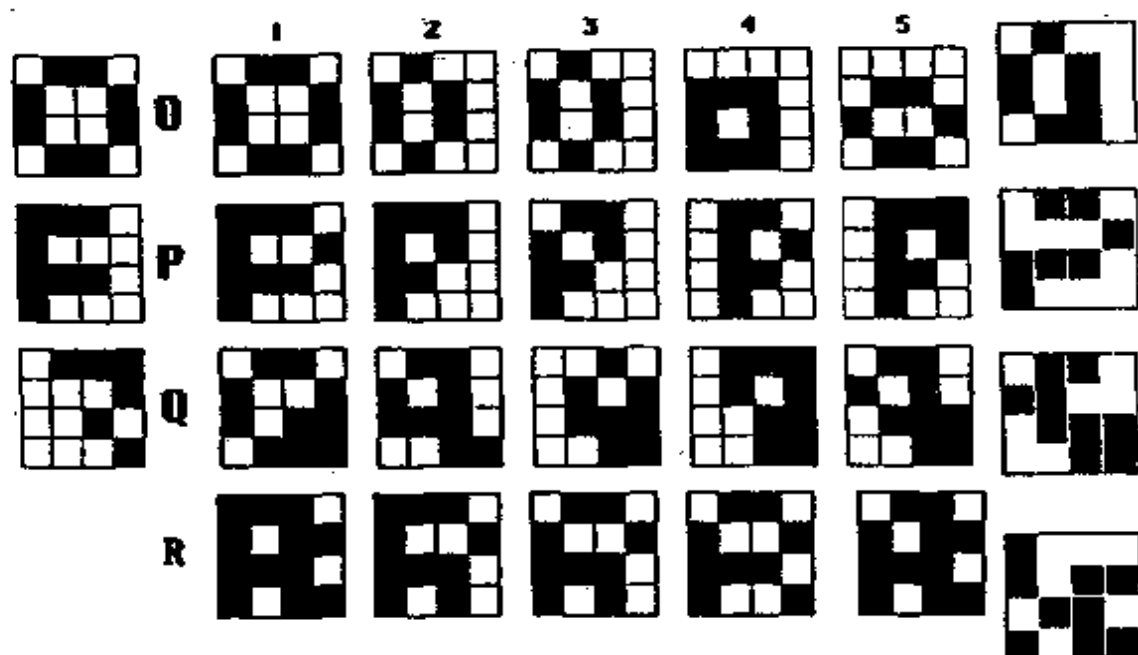


图 2.1 手写体英文字母的识别

O, P, Q 三组字母图像主要特征的特征图形, 我们所选取的图形如图 2.1 中第一列所示. 这些特征图形有两个重要特点:

(1) 吸取了它们所代表的样板的主要特征, 如: O 的特征是一个闭合圆, P 的特征是左侧有一竖, Q 的特征是右下角有一笔等;

(2) 这一组特征图形相应的矢量是互相正交的, 存储正交矢量是有利于联想记忆的. 事实上, 并不要求特征矢量严格正交, 只要它们之间的内积足够小, 近似正交就够了.

第二步要对网络进行训练. 我们采用一个二层网络. 第一层的任务是对样板进行特征提取, 这是一种异联想记忆. 由输入图形经第一层单元所得到的特征图形不一定是我们所存储的标准特征图形, 可能有一些畸变. 第二层单元的任务是找出与第一层的输出最靠近的特征图形作为输出, 从而最终给出识别结果, 这是一种自联



想记忆.

实现特征提取的连接矩阵可表达为

$$[FE] = (1/NS) \sum_{f=1}^F \sum_{s=1}^S [|F_f\rangle \langle E_{fs}|] - \text{对角元.} \quad (2.19)$$

其中,  $N$  为矢量的维数, 这里样板矢量与特征矢量的维数均取为  $N$ ;  $|F_f\rangle$  是特征矢量,  $F$  为特征矢量的数目;  $|E_{fs}\rangle$  是样板矢量,  $S$  为每一种特征矢量对应的样板数. 图 2.2 是相应于图 2.1 的  $[FE]$  矩阵. 这是一个  $16 \times 16$  的矩阵, 每个矩阵元的数值由图中相应位置的小方块表示, 方块的面积代表矩阵元的绝对值, 黑方块代表正值, 白方块代表负值.

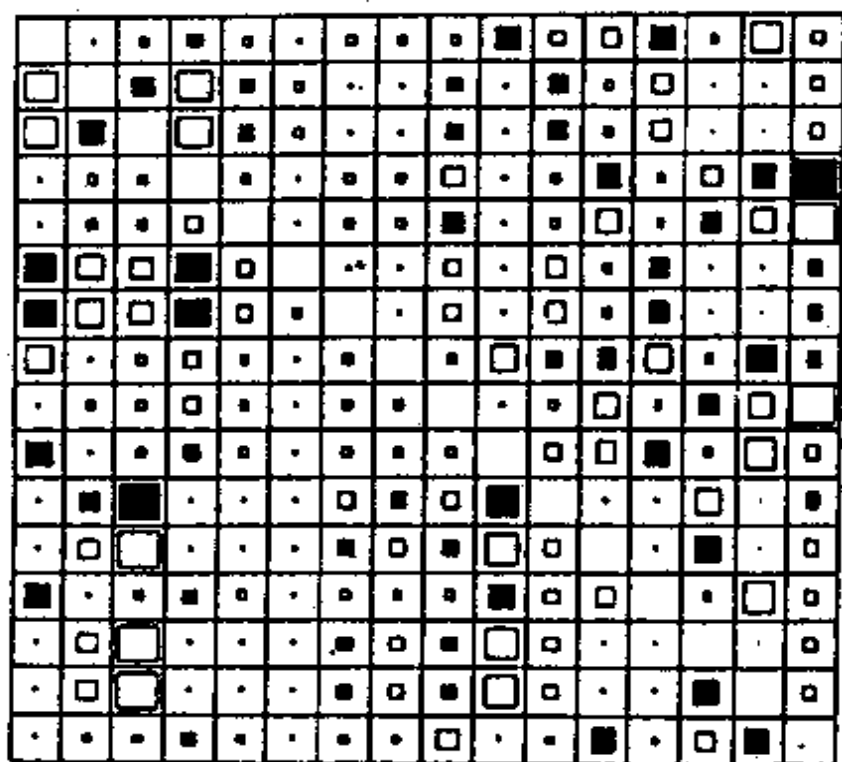


图 2.2  $[FE]$  矩阵

决定特征矢量输出的自联想记忆矩阵可表达为

$$[AM] = (1/N) \left\{ \sum_{f=1}^F |F_f\rangle \langle F_f| - F[\delta] \right\}, \quad (2.20)$$

相应于图 2.1 例子的  $[AM]$  矩阵如图 2.3 所示.

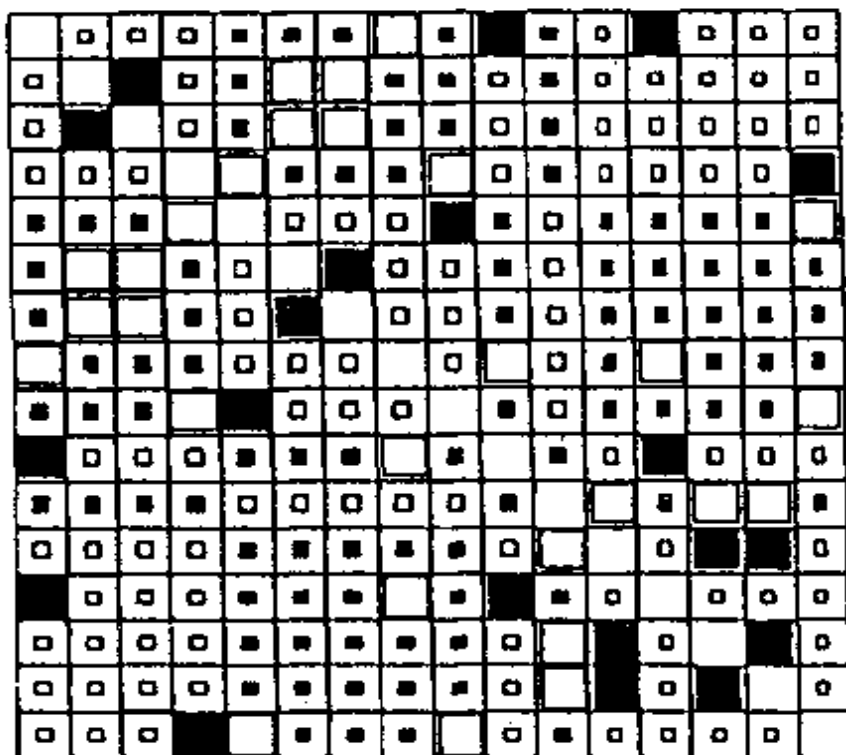


图 2.3  $[AM]$  矩阵

我们发现,分别由图 2.2 与图 2.3 所表示的连接矩阵  $[FE]$  和  $[AM]$  所组成的二层网络,能够实现图 2.1 所要求的图像识别,全部 15 个样板均能被识别出来,未出现混淆和存储容量达到饱和的现象.

但是,这并不表示能实际应用的手写体字母的识别问题已经解决了,这里还有很大的差距.例如,当我们把 26 个英文字母全部包含进来时,由于手写体字母变化很大,每个字母相应的样板数很多,使得完全靠人力选取正交特征矢量几乎是不可能的.是否有系

统的方法呢？答案是肯定的，这将在下一个例子中说明。

## 2. 手写体阿拉伯数字的识别

为了能利用神经网络读支票和邮政编码，必须解决手写体阿拉伯数字、小数点及逗号的识别问题，这里介绍作者的一项工作。

采用  $N = 8 \times 8 = 64$  的二维图像如图 2.4 所示，包括从 0 到 9 的数字、小数点及逗号共 12 种字符。每种字符有 4 种不同的样板（第一行可视为标准图形，下面三行为畸变图形），共有 64 个样板。

我们采用一种系统的方法确定特征矢量，其步骤如下：

第一步，将同一字符的几种样板进行平均，得到一个比较模糊的图形。

第二步，对平均图形作 Walsh 变换，这种变换是按方形波（只取 1 或 -1 的波形）展开，得到平均图形的频谱。

第三步，将频谱作非线性处理，然后分成低频、中频及高频三支；或者，先分成低频、中频、高频三支，然后分别作非线性处理。三个频段的划分界限是由人来确定的。

第四步，将 Walsh 谱的低频、中频和高频三支分别进行 Walsh 逆变换，得到三个图形。这三个图形用来作为相应字符的特征矢量，每个字符有三个不同的特征矢量。

确定了特征矢量之后，便可建立异联想记忆矩阵，按本节第一个例子的方法便可进行字符识别。

作为一个例子，图 2.5 给出图 2.4 中数字 0 的 Walsh 谱，图中三种符号分别相应于字符 0 的三种不同样板。

在图 2.6 中，给出了几个字符的特征矢量。图中第 1 行是样板图形，第 2 行是由低频分支得到的特征矢量，第 3, 4 行分别是由中频分支和高频分支得到的特征矢量。

总起来说，联想记忆可用来进行字符识别，但联想记忆并非直接用于字符的图像，而是首先要进行特征提取，然后才能运用

联想记忆.

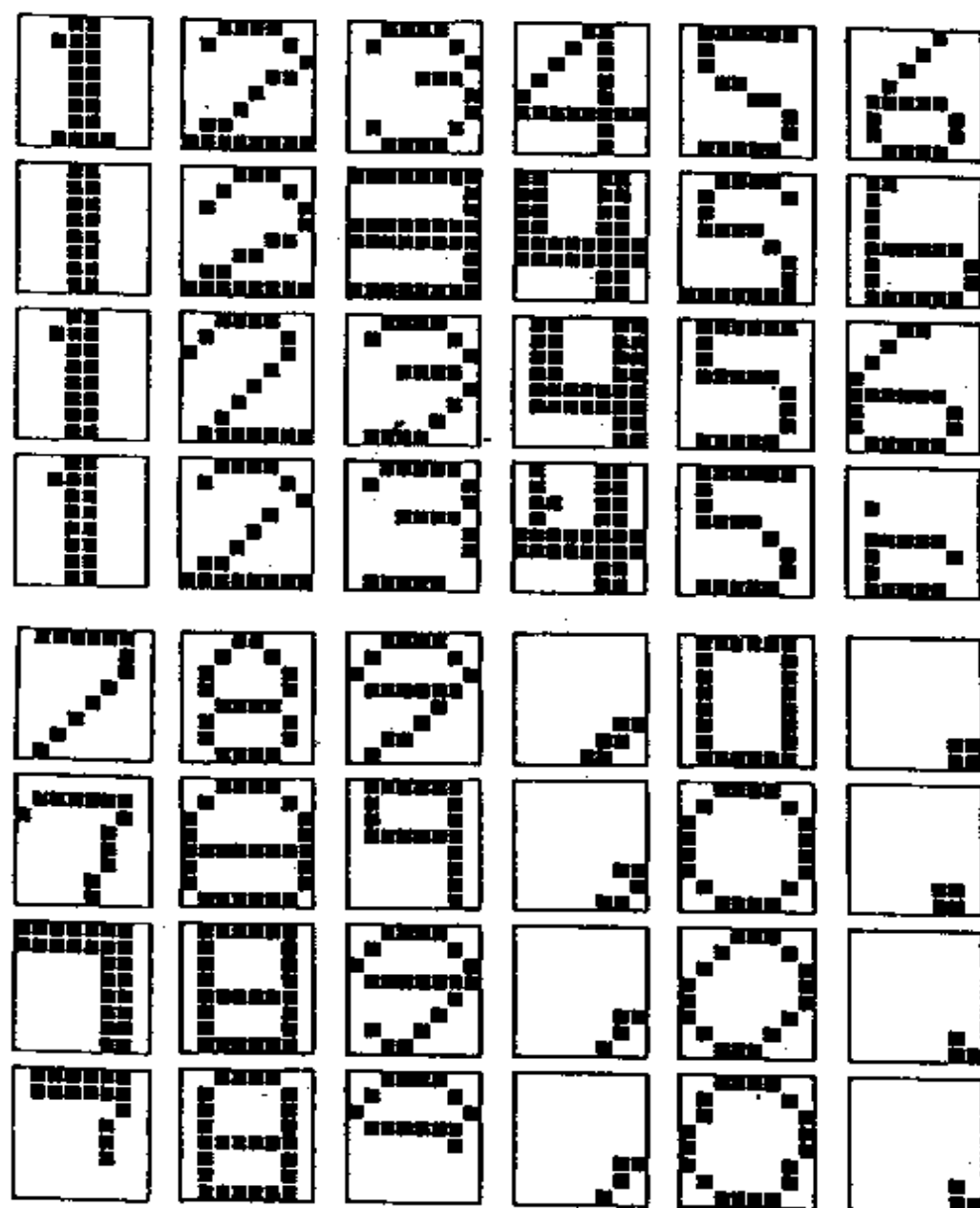


图 2.4 手写体阿拉伯数字的识别

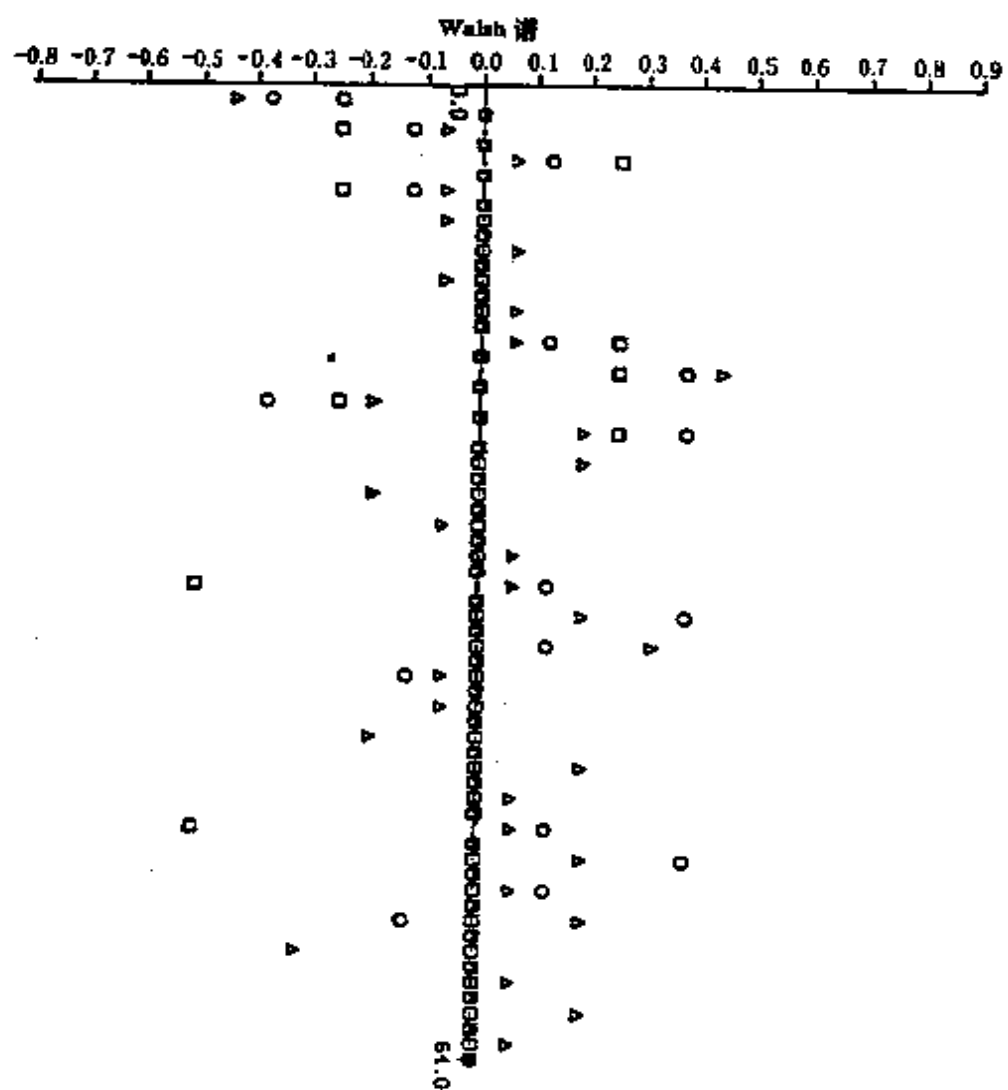


图 2.5 数字 0 的 Walsh 谱

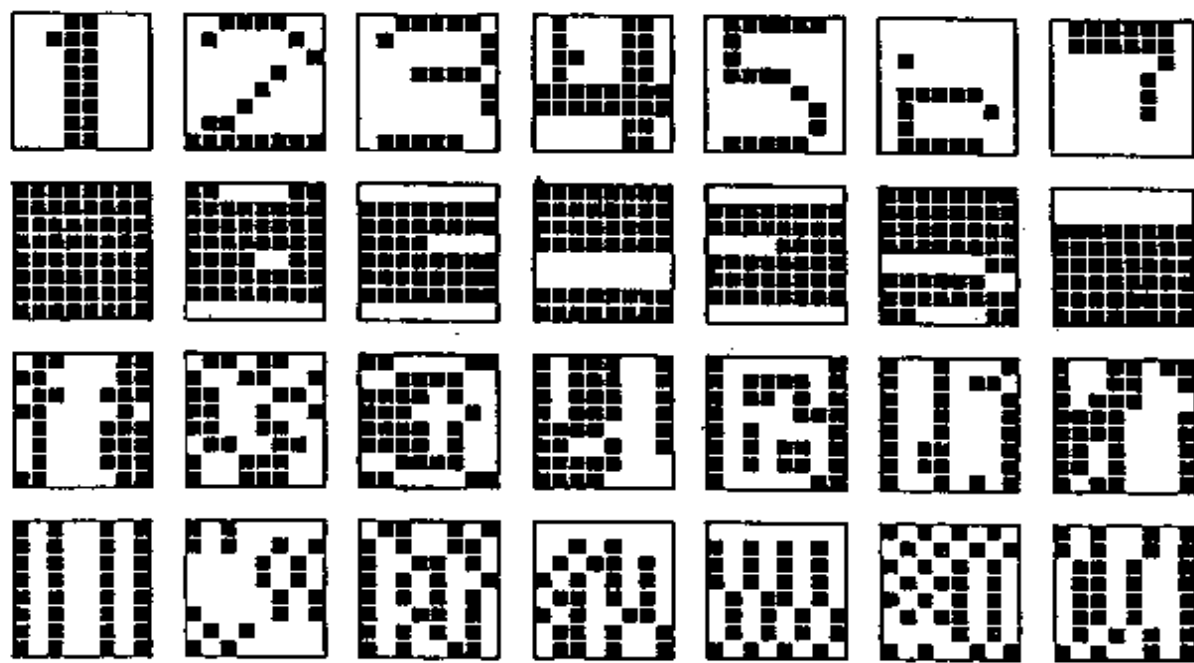


图 2.6 特征矢量,第 2,3,4 行分别相应于由低、中、高频得到的特征矢量

## 第三章 神经网络的结构与算法

### 一、神经网络的结构与用途

我们必须针对各种不同的目的去设计和制造相应的神经网络,不可能制造一种通用的神经网络,这是神经网络与传统计算机的一个重要差别.但是,不同用途的神经网络,其性能通常应符合一些共同的基本要求,主要包括容错能力、自纠正能力、快速反应能力以及在一定范围内实现全局优化等.这些性能大体上分别与第一章中关于神经网络的定义相对应,即非局域性、非线性、非定常性以及非凸性等.

现阶段,研制神经网络大体有三种途径:电子学方法,采用大规模集成电路;光学方法,采用光学器件;数字模拟方法,利用传统计算机进行数字模拟研究.总的说来,无非是软件模拟、硬件制造以及将算法与结构统一于一体的所谓“湿件”(wetware)或称“网件”(netware)的研制等.

然而,当我们真的去模拟大脑时,却遇到了极大的困难.已经指出,人的大脑是极复杂的系统,约有  $10^{10}$  个神经元,每个神经元平均与  $10^4$  个神经元连接,其单元及连接不仅数量极大,而且有极高的密度.按当前的技术水平,大约只能制造  $10^2$ — $10^3$  个单元组成的网络,与大脑的水平相差悬殊,因此,近期内还只能制造低智能的神经网络系统.但是,我们可以依靠人力帮助神经网络提高性能,如第二章中所介绍的如何为网络选取特征矢量、提高容量等.只要训练得当,这种低智能的神经网络仍可完成很出色的工作.

目前,已有很多单位在用电子学或光学的方法研制神经网络系统.实际上,在美国几乎没有一所大学不在从事这方面的研究.

具体的方案主要是直接模仿生物学上的神经结构. 例如, 电子学方法中, 用非线性元件模拟神经元, 树突与轴突对应于连接单元的导线, 而突触则相应于连线的电阻, 即为连接的权重. 由于电子技术已发展得很成熟, 电子元件可做得很小, 用电子学方法制造神经网络似乎是一条捷径. 但是, 当单元数增多时, 单元间的连接数将按单元数的二次方增加. 稍大一点的网络, 其导线就像一锅面条一样乱成一团, 在技术上极难实现. 因此, 连接困难是电子学方法所面临的一个限制.

用光学方法模拟生物学上的神经网络可采用激光传输信号系统. 由于光束本身无法表示正负, 故应采用双层结构, 各层分别代表正信号与负信号. 神经元用非线性放大器模拟, 而树突和轴突相应于光束和透镜, 以光束作为通信线路可避免电子学方法中的连接困难, 但因光学元件体积比电子元件大得多, 要在一个网络中包含大量的光学元件也是极为困难的.

上述电子学和光学的神经网络系统实际上已经有人在做了. 例如, 采用二维或三维连接结构的大规模集成电路, 可达到较高的连接密度, 连接权重也可做成可变的. 光学网络可输入一维空间的光学信号, 连接矩阵是二维的, 进一步还可研究输入二维空间光学信号的网络, 为此要实现四维的连接.

以上讨论的是神经网络的结构, 现在进一步考察一下神经网络的处理单元, 这对应于生物学上的神经元.

在生物学中, 神经元的类型十分复杂, 我们大体上把它们划分为大型的、中型的和小型的三类, 各具有不同的功能. 用于不同目的的神经网络可采用不同类型的处理单元.

大型神经元具有记忆、逻辑和通信三种能力, 相当于一个微型计算机. 可用光学-电子学相结合的方法制造这种大型处理单元. 包含大型处理单元的神经网络可用于机场及工厂的调度和安全管理, 化工厂中管道阀门的自动控制, 以及核反应堆的监督控制等.



面,实际上在很多领域都开展了研究工作.至于它的应用前景则更是极为广阔.凡是人凭直觉解决的问题是无法用必须预先制订规则的人工智能系统来完成的,而这正是神经网络系统可用武的广阔天地.

关于今后几年进展的估计,若以网络内每秒通讯数代表运算速率,以网络内连接数表示规模,则光学网络看来比较有前途,在若干年内,其规模可达  $10^{10}$  ( $10^5$  个单元),速率可达  $10^{12}$ .当然,这个水平比人的大脑还相差很远,大脑的规模约为  $10^{14}$ ,而速率为  $10^{15}$ .

## 二、模拟退火算法

在本节中,将介绍神经网络的一种重要算法——模拟退火方法.任何自动化均要求一定约束条件下的全局优化,模拟退火算法是实现约束全局优化的一种基本途径.

在第一章中曾讲过一个猴子如何从盒中取花生仁的故事,这是说明模拟退火方法基本精神的一个生动的比喻.花生仁只能从盒子底部最深的坑的中心小孔漏出,但盒底有很多高低不同的坑,花生仁在盒内运动的自然趋势是移向最近的坑底,因此,要让花生仁漏出就要使它们越过壁垒找到最深的坑底.这正是一个约束全局优化问题.摇动盒子使花生仁产生随机运动,以便跳出浅坑,这相当于给系统一定的温度以越出局域极小.只有使摇动的强度按一定方式逐渐减弱,才能使花生仁最终从深坑底部漏出,这相当于温度逐渐下降的过程,也就是退火过程.如果能设计出适当的退火过程一般总能达到目的.

读者可能会发现模拟退火方法与应用极为广泛的蒙特卡罗方法很相似.实际上,我们可以把前者看作是后者的一种推广.这两种方法中都引入了随机运动和概率分布函数,但蒙特卡罗方法中概率分布不随时间变化,相当于“恒温”情形,而模拟退火方法中概

率分布则是随温度下降而不断变化的。

模拟退火方法是 1953 年 N. Metropolis 等人研究二维相变时提出来的。1983 年, S. Kirkpatrick 等人用这种方法设计大规模集成电路(VLSI)。G. E. Hinton, T. J. Sejnowski 与 D. H. Ackley 于 1983 年把模拟退火方法用于神经网络, 提出了玻耳兹曼机(Boltzmann Machine)。1984 年 S. Geman 和 D. Geman 给出退火率(温度随时间的变化率)为  $T(t) \propto 1/\log t$ 。这个退火过程太慢, 因而效率很低, 几乎没有多大实用价值。1985 年, 本书作者提出了一种快速模拟退火方法, 称为柯西机(Cauchy Machine), 使这种方法有了实际应用的可能性。

这里, 我们将介绍快速模拟退火算法(柯西机), 并与玻耳兹曼机进行比较。

优化问题在数学上常可表达为使某个评估函数取极小(或极大), 评估函数一般由很多变量决定, 把这些变量用一个矢量  $x$  表示, 则评估函数可写成  $C(x)$ 。

如果评估函数  $C(x)$  只有单个极小, 则称为凸性(convex)情形, 可以用任何梯度下降法找到它的极小而实现优化。但在很多情况下,  $C(x)$  有多个极小, 称为非凸性(nonconvex)情形。约束全局优化问题就是要在一定范围内找到最低的极小, 这个最低的极小称为一定范围内的全局性极小, 而其余的极小则称为局域极小。非凸性情形的全局优化不能采用使  $C(x)$  单调下降的方法, 必须保证系统能从局域极小内爬出来, 以便搜寻到全局性极小。换言之, 要使  $C(x)$  有一定的概率向增加的方向变化, 即引入一定程度的随机性扰动。引入随机扰动的方式一方面要使系统能克服局域极小的约束, 另一方面又要保证系统最终达到全局极小后不再受扰动的干扰和破坏, 这就是模拟退火方法所要解决的问题。模拟退火方法可以看作是有噪声的梯度下降法, 当标志噪声强度的温度为零时, 便成为通常的梯度下降法。

模拟退火算法基本上由三部分组成:

(1) 以一定的概率密度跃迁到新的状态,这个概率密度函数称为生成函数(generating function);

(2) 以一定的概率密度容忍评估函数的偶然上升,这个概率密度函数称为容忍函数(acceptance function);

(3) 以一定的冷却程式降低温度,这个等效温度是生成函数和容忍函数中的控制参量,确定所引入的随机扰动(噪声)的强度.

经典的模拟退火(CSA;classical simulated annealing)算法中采用高斯型的生成函数,

$$G_G(x) \approx \exp[-x^2/T(t)], \quad (3.1)$$

其中  $T(t)$  是温度,决定了概率密度分布的特征宽度.这种分布函数远处是指数型衰减的,因而代表一种局域型搜寻过程.为了能找到全局极小,温度下降要很慢. Geman 兄弟证明了它所要求的退火方式为

$$T(t)/T_0 \propto 1/\ln t, \quad (3.2)$$

其中  $T_0$  为初始温度,要求  $T_0$  取足够高的值.

在快速模拟退火算法中生成函数采用柯西型生成函数(在一维情形为洛伦兹函数)

$$G_C(x) = T(t)/[T^2(t) + x^2]^{(D+1)/2}, \quad (3.3)$$

其中  $D$  是维数.这种函数在其峰值附近与高斯型分布(3.1)式相同(两种分布函数在峰值附近的展开式相同),但(3.3)式具有较高的两翼,这使它与高斯分布在远离中心的地方有明显的差别.高斯分布所决定的二次矩是有限值,而柯西分布的二次矩是发散的: $\langle x^2 \rangle = \infty$ .柯西型生成函数所决定的状态跃迁会偶然出现大尺度的跳跃,因此,这相当于一种半局域的搜寻.下面将会证明,采用柯西分布时,可以有很快的退火率:

$$T(t)/T_0 \propto 1/t. \quad (3.4)$$

模拟退火算法的过程大致如下:从一个随机选取的状态出发,

依据生成概率在每一步产生一个新的状态. 如果这个新状态的评估函数值比原来的状态低, 则令它为系统的新状态; 如果它比原先状态的评估函数值高, 则它成为新状态的概率由容忍函数确定. 若系统不进入这个新状态, 则它仍保持原先的状态, 生成函数与容忍函数均按照规定的冷却方式变化.

图3.1中, 以一维情形为例给出快速模拟退火算法的流程图. 具体说明如下: 令  $t-1$  时刻处于状态  $x_{t-1}$ , 其评估函数为  $C_{t-1} \equiv C(x_{t-1})$ . 现根据生成函数确定  $t$  时刻的状态

$$x_t = x_{t-1} + \Delta x, \quad (3.5)$$

生成函数取为柯西函数

$$G_C(\Delta x) = \frac{1}{\pi} \cdot \frac{T}{T^2 + (\Delta x)^2}. \quad (3.6)$$

将  $G_C(\Delta x)$  作积分, 得

$$F(\Delta x) = \int_0^{\Delta x} G_C(\Delta x) d\Delta x = \frac{1}{\pi} \arctan \frac{\Delta x}{T}. \quad (3.7)$$

$F(\Delta x)$  的取值范围为

$$-0.5 \leq F(\Delta x) \leq 0.5. \quad (3.8)$$

在均匀分布的区间  $[-0.5, 0.5]$  中, 取一个随机数  $y$

$$-0.5 \leq y \leq 0.5, \quad (3.9)$$

并由

$$y = F(\Delta x) \quad (3.10)$$

确定  $\Delta x$ , 即

$$\Delta x = T \tan(\pi y). \quad (3.11)$$

对均匀分布的  $y$ ,  $\Delta x$  的概率密度函数即为柯西函数(3.6)(因

$F' = G_C$ ). 将(3.11)代入(3.5)便得到新的状态  $x_t$ . 由  $x_t$  计算  $C_t \equiv C(x_t)$  及

$$\Delta C = C_t - C_{t-1}. \quad (3.12)$$

如果  $\Delta C < 0$ , 则说明新的状态  $x_t$  比原先的状态  $x_{t-1}$  有更低的评估

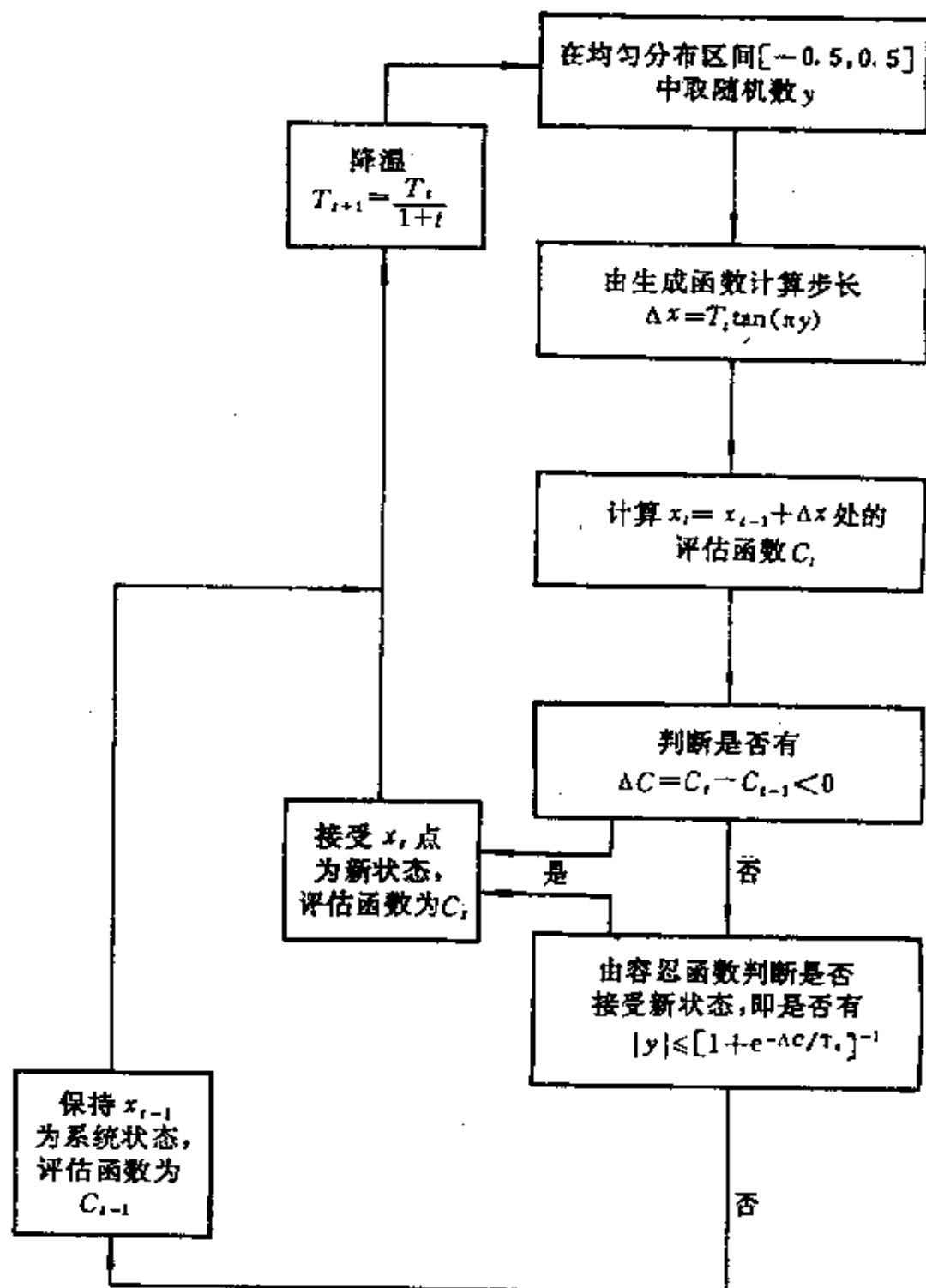


图3.1 快速模拟退火算法流程(一维情形)

函数,接受此状态为系统的新状态,并按退火率(3.4)式降温

$$T_i = T_0/t, \quad (3.13)$$

然后作下一步搜寻. 如果  $\Delta C > 0$  则要由容忍函数判断是否接受新的状态. 采用与玻耳兹曼机相同的容忍函数:

$$A = \frac{e^{-C_i/T}}{e^{-C_i/T} + e^{-C_{i-1}/T}} = [1 + e^{\Delta C/T}]^{-1}, \quad (3.14)$$

在  $\Delta C > 0$  的情形,  $A$  的范围为

$$0.5 \geq A \geq 0, \quad (3.15)$$

当  $\Delta C = 0$  时,  $A = 0.5$ ;  $\Delta C \rightarrow \infty$  时  $A = 0$ . 容忍函数的作用是对评估函数的增大作概率性限制,不妨仍取开始时选取的随机数  $y$  ( $|y| \leq 0.5$ ) 来达到这一点,如果

$$|y| \leq [1 + e^{-\Delta C/T}]^{-1}, \quad (3.16)$$

则容忍这种  $C$  的增加,接受  $x_i$  为新状态,然后降温并作新的搜寻;如果不满足(3.16)式,则不接受  $x_i$  作为新状态,仍保留原先的  $x_{i-1}$  为系统的状态,然后降温并作进一步的搜寻. 每一次搜寻,均由产生新状态、计算新状态的评估函数、判断是否接受新状态及降温等四个步骤组成.

上述快速模拟退火算法与经典模拟退火算法的步骤基本相同,区别只在于生成函数及退火率不同. 两种生成函数的傅立叶变换均可写成  $\exp\{-|k|^\alpha\}$  的形式,当  $\alpha = 2$  时得到高斯分布;而当  $\alpha = 1$  时即为柯西分布. 高斯分布相应于无规行走,而柯西分布作为生成函数则相应于无规行走再加偶然的跳跃,因而后者能较快地搜索到目标. 现用一个比喻来说明两种生成函数的差别:一个男孩子请女友喝咖啡,结账时提出要为女友付钱,却突然发现钱包不见了. 于是温度一下子升得很高,脸涨得通红,两手在各个衣袋里乱翻,仍是找不到钱包. 如果按照高斯型概率分布进行搜索,他将从身边开始在附近一步步乱摸,逐渐扩大范围,可能最终在他的汽车里找到钱包. 但若按柯西型概率分布寻找,他会很快冷静下来,

并且可能一步跨到桌子旁寻找,如仍未找到,他会迅速奔向汽车并找到钱包,这种方法可以更快地搜索到目标。

现在,我们以一个具体的例子比较两种模拟退火算法.取评估函数形式为

$$C(x) = x^4 - 16x^2 + 5x \quad (3.17)$$

证个函数有两个极小值,分别为  $C_1 = -50$  和  $C_2 = -78.3$ . 生成函数分别为高斯函数

$$G_G(\Delta x) \propto \exp[-(\Delta x)^2/10^{-2}T(t)] \quad (3.18)$$

和柯西函数

$$G_C(\Delta x) \propto \frac{10^{-2}T(t)}{10^{-4}T^2(t) + (\Delta x)^2} \quad (3.19)$$

当状态点处于局域极小  $C = -50$  附近时,生成函数  $G_G$  与  $G_C$  如图

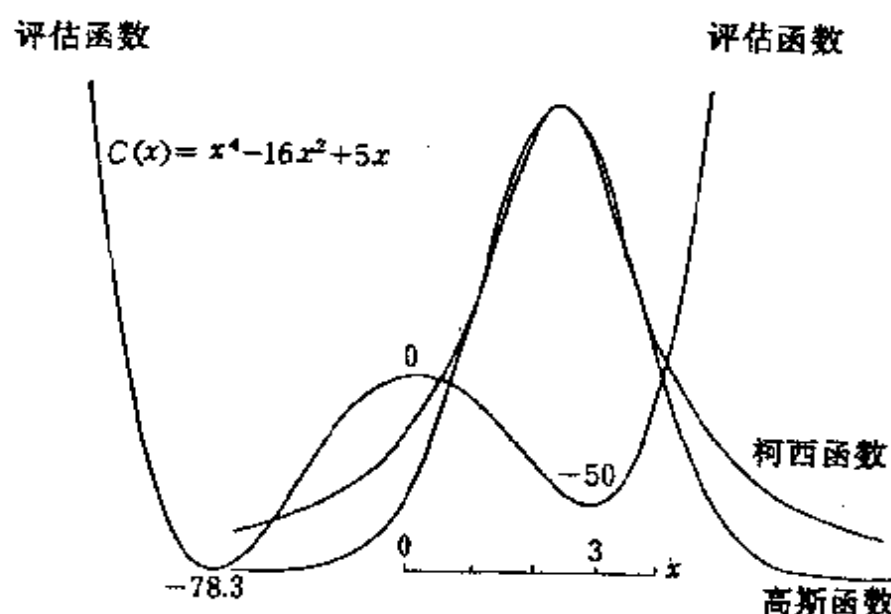


图 3.2 评估函数(3.17)式及生成函数(3.18)与(3.19)式

3.2 所示.退火率分别取为

$$T(t) = 25[\ln(1 + 10^{-2}t)]^{-1} \quad (3.20)$$

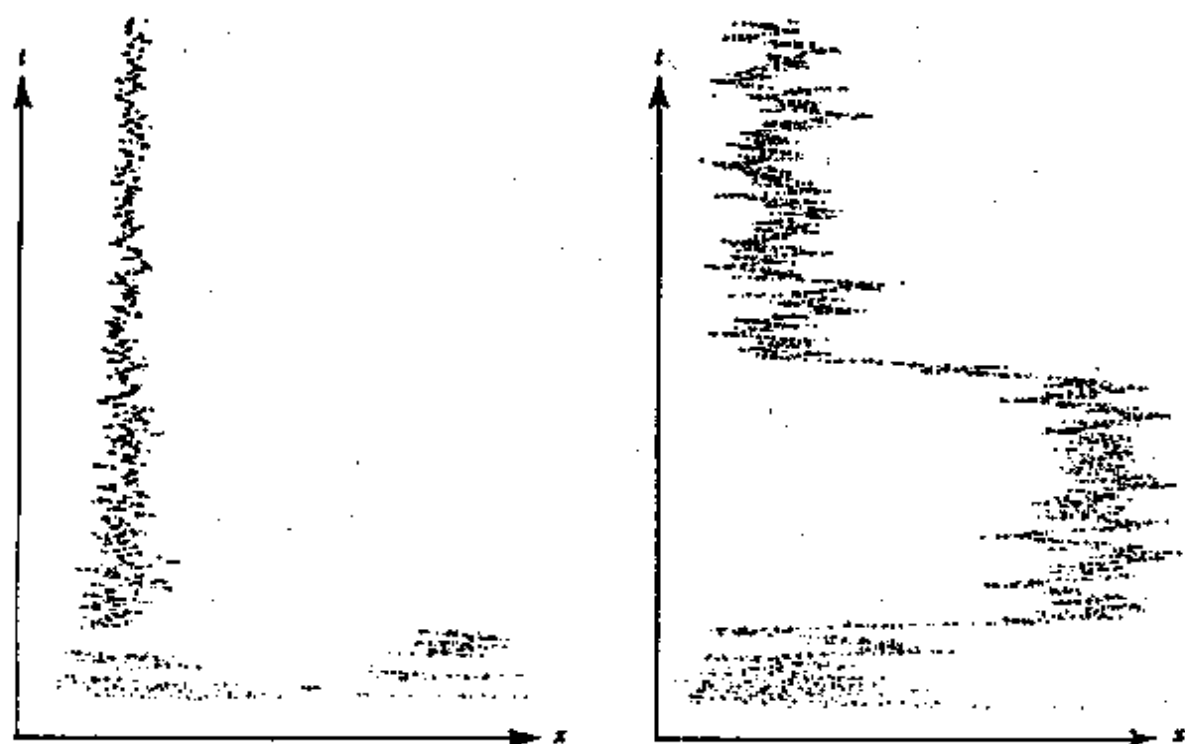
和

$$T(t) = 100[1 + 10^{-2}t]^{-1}. \quad (3.21)$$

容忍函数的形式均取为(3.14)式:

$$A = (1 + e^{\Delta C/T(t)})^{-1}. \quad (3.22)$$

图3.3给出了状态点随时间的变化,图3.3(a)相应于柯西机,图3.3(b)相应于玻耳兹曼机,从图中可看到搜索全局极小的整个过程.



(a) 柯西机

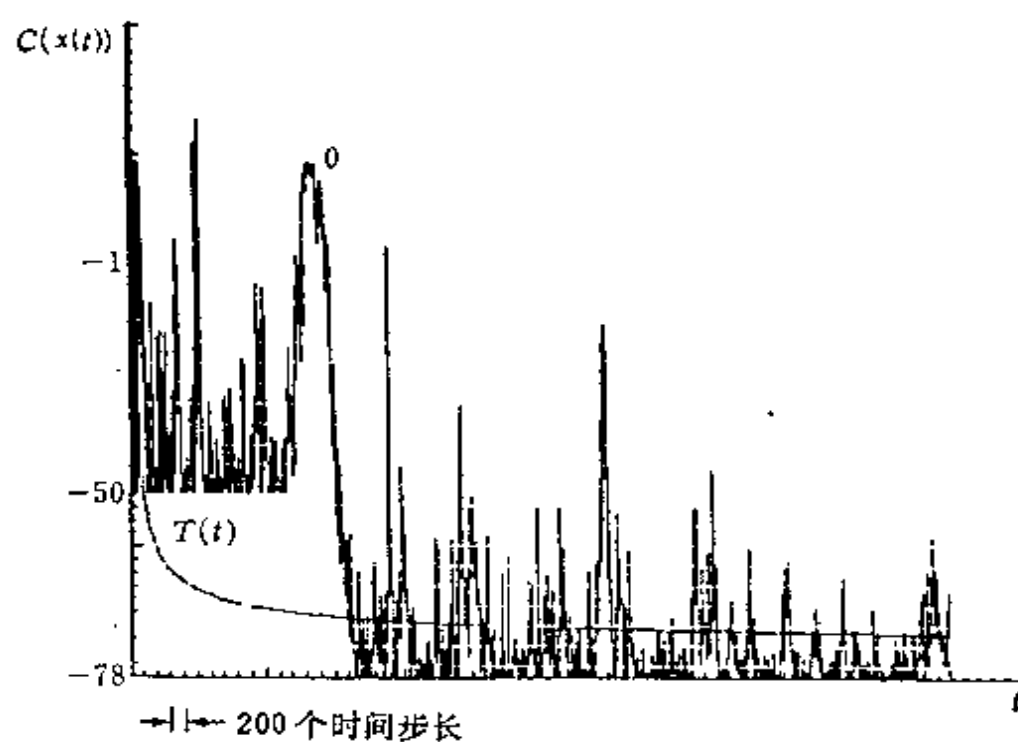
(b) 玻耳兹曼机

图3.3 状态点随时间的变化

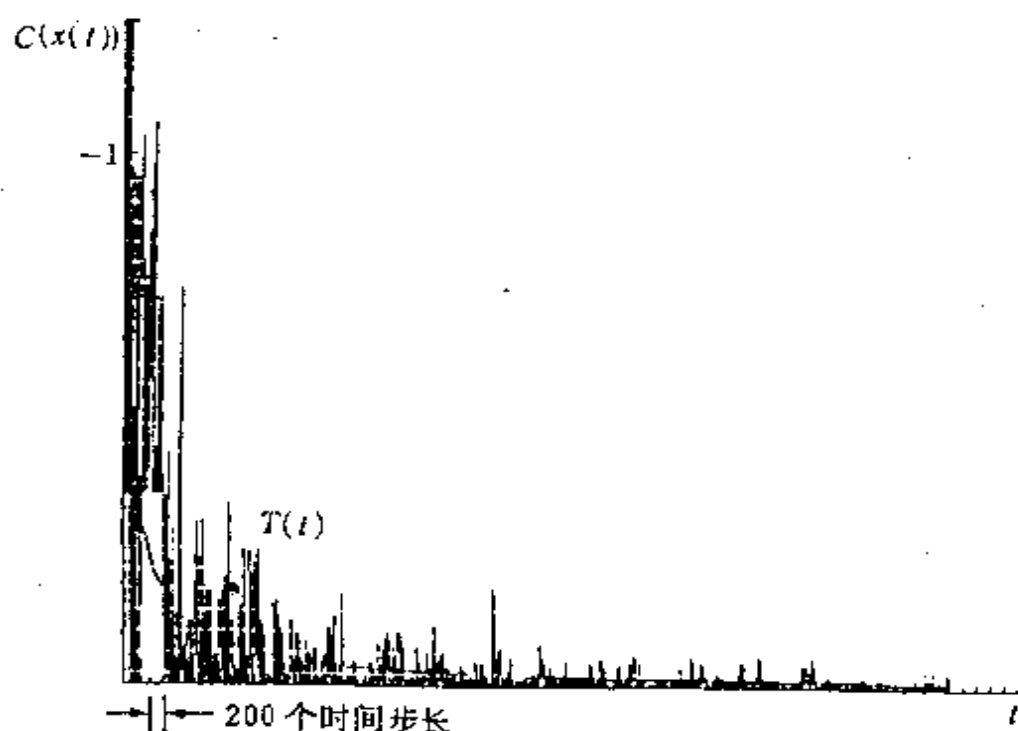
图3.4是相应评估函数随时间的变化,图3.4(a)为柯西机,图3.4(b)为玻耳兹曼机.

从图3.3与3.4可以看出,柯西机进入全局极小 ( $C = -78$ ) 比玻耳兹曼机快得多. 在图3.4中,虚线代表温度随时间的变化





(a)



(b)

图3.4 评估函数随时间的变化

生成函数可实现全局优化.

对快速模拟退火算法,有

$$g_i \approx \frac{T(t)}{[T^2(t) + (\Delta x)^2]^{(D+1)/2}} \approx \frac{T(t)}{[\Delta x]^{D+1}}, \quad (3.28)$$

取退火率为

$$T(t) = T_0/t, \quad (3.29)$$

则有

$$g_i = \frac{T_0}{t|\Delta x|^{D+1}}.$$

由此得

$$\sum_{t=t_0}^{\infty} g_i \approx \frac{T_0}{|\Delta x|^{D+1}} \sum_{t=t_0}^{\infty} \frac{1}{t} \approx \infty \quad (3.30)$$

所以,对柯西型生成函数,只需与时间成反比的退火率即可满足我们的要求.

我们可以严格证明,有了这一步长条件,我们最终一定可以实现全局优化.考虑极限情况,假设温度已经很低而没有实现全局极小,由于容忍函数已经较小(平均说来),状态点一定在某一局域极值附近作小幅动振动,同时也不断向周围探测,而且探测步长在足够长时间内可取到任意值,所以系统也就能探测到任意点,而一旦探测到评估函数较低的位置立即发生“跃迁”.这样,最终一定能达到全局极小值点.

当然,这只是一种理想化的数学证明,实际中的模拟退火要在有限时间内完成,而且对于具体问题还有参数调节等许多事情要办,这些多半要靠实验中去摸索了.

尽管快速模拟退火算法效率较高,它的应用仍然受到一些限制.评估函数在状态空间中可具有不同的形态,如图3.5所示,其中图3.5(a)显示的是只有单个极小的凸性函数.前面已经指出,在这种情形下只需采用任何一种梯度下降法便可实现优化.图3.5(b)

是非凸性情形,全局极小形如口很大的深坑,即占据相空间中相当大的范围,这种情形下的全局优化用快速模拟退火算法是十分有效的,图3.5(c)也属于非凸性情形,但全局极小形如口很小的深井,它的范围只是相空间中很小的一个区域.对于这种情形,快速模拟退火是无能为力的.这是因为快速模拟退火算法所用的概率密度函数宽度较大,相当于一种较粗的扫描,难以刻画很细小的结构.打个比方,这就好比想把高尔夫球打进无标志的球洞一样,虽然你不断击球,把它从各种浅坑中驱赶出来,但它滚入球洞的可能性微乎其微.

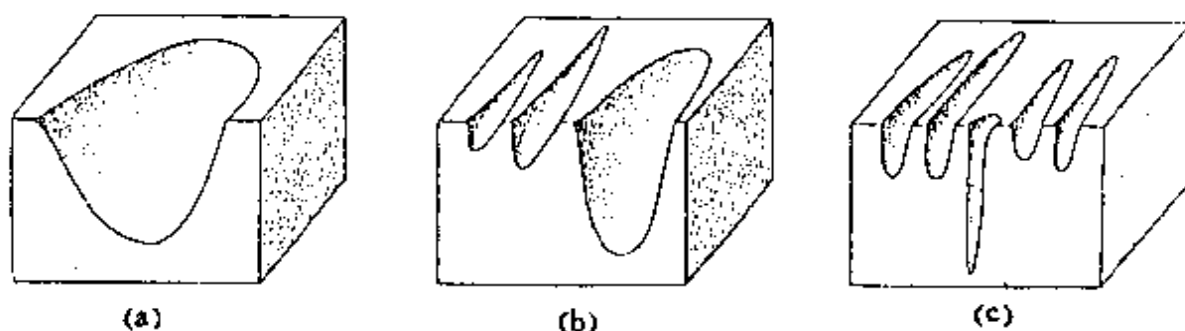


图3.5 不同类型的评估函数

还有值得提到的一点是我们可以用光学手段实现柯西分布,图3.6是示意图.图的左侧是一个可以随机旋转的镜子,令转角为 $\theta$ ;右侧是一个平面屏幕,中心有一狭缝,它与镜的转轴的距离为 $y$ ,光线从狭缝入射到镜子上,然后反射到屏幕上.令镜子的角度为 $\theta$ 时,反射光投射到屏幕上高为 $x$ 的地方,则有下列关系:

$$2\theta = \arctan \frac{x}{y} \quad (3.31)$$

或

$$2d\theta = \frac{ydx}{y^2 + x^2} \quad (3.32)$$

若旋转镜各种角度的出现是等概率的,由(3.32)式便可看到反射光投射位置  $x$  的概率密度分布函数即为柯西分布函数.利用这种性质,我们可以由光学组件排成阵列,制造光学柯西机.

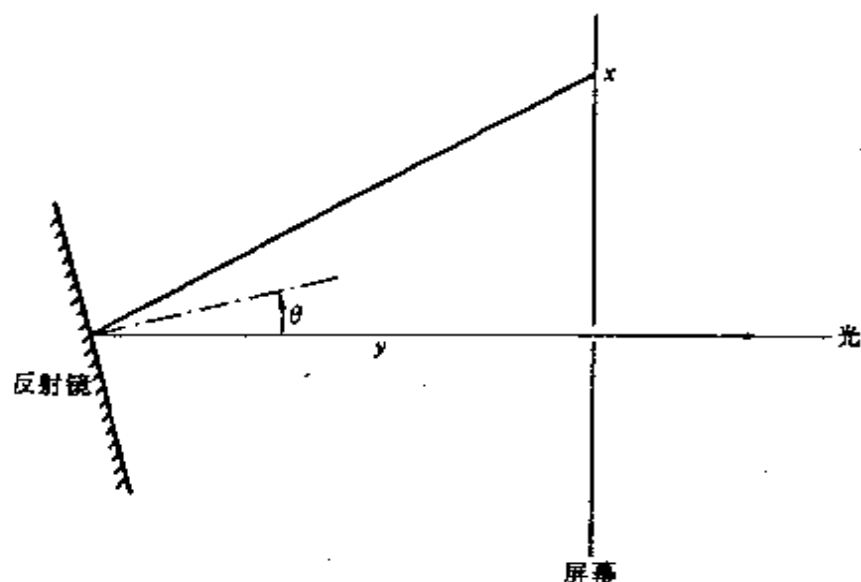


图3.6 用光学手段实现柯西分布的示意图

模拟退火算法的“模拟”二字可以理解为对自然界中普遍存在的各种退火过程的模拟.实际上,整个宇宙正是处于一个退火过程之中.宇宙自大爆炸以后,温度不断下降,出现了质子、中子、电子等各种粒子,形成了各种原子核、原子和分子,出现了等离子体、气体、液体和固体等千姿百态的物质聚集状态,最后还出现了有生命的物质以至于人类.这些现象可以看作是宇宙在退火过程中跑到了状态空间中相对稳定的能量极值点附近.这个进程与宇宙的退火率有密切关系,如果宇宙具有另一种退火率,宇宙将可能是完全不同的另外一副面貌.

用模拟退火方法解决约束全局优化问题有十分广泛的应用领域,例如:贷款分析,支票识别,保险估价,国防应用(雷达、声纳的图像处理——消除噪声、数据压缩、特征提取、图像识别、对抗模

拟、武器瞄准等),气象预报,地震预报,医学应用(药物作用、手术方案等),市场预测,汽车自动驾驶仪,运输调度管理,通信(语言及图像压缩、实时翻译等),商场地址选择,股票市场分析,机器人(视觉系统、机械手、触觉反馈夹具控制等),产品检测,VLSI 芯设计、制造和检验,航空、航天控制系统等.

我们可以大胆地预测,在不远的将来,如再过十年、八年,神经网络系统将在很多领域中开始大显身手.

## 第四章 神经网络的学习方式

神经网络具有一定的智慧,这突出表现在它能够进行学习.神经网络的学习过程主要是按一定的方式调整单元之间的连接,使网络形成所要求的功能.至今,已经提出了若干不同的学习方式.我们要强调指出的是:神经网络的学习方式必须与其结构相配合,否则便不能达到应有的效率.

本书前面介绍过的由标准输出、输入矢量的外积确定连接矩阵的方法实际上可以看作是神经网络的一种学习方式.这是一种最简单易行的学习方式,当我们需要训练一个神经网络时,首先可以试用这种方法.如果效果不好,还可试验本书第三章中所介绍的模拟退火方法.一个网络的学习或训练也是一个在一定约束下的全局优化问题,即通过调整单元间的连接使网络达到最简性能.因此,模拟退火方法也是一种学习方式,可用于训练多层网络.如果上述两种学习方式效果均不理想,则再试採用其它的学习方式.

本章中,主要讨论三类学习方式.第一类是以 Hopfield 网络为典型的单层网络的学习方式,我们将指出,对于这类网络实际上只需用外积法确定连接矩阵就够了;第二类是 D. E. Rumelhart 等人提出的反向传播法(back-propagation)或称为误差传播法(error propagation),可用于训练多层网络,这是一种有导师(supervised)的学习方式;第三类是 S. Grossberg 与 G. A. Carpenter 提出的适应共振(adaptive resonance)法,通常认为这是一种无导师(unsupervised)的学习方式,但这种方法并不真正是无师自通,实际上也仍然是有导师的.此外,本章还将讨论一个典型的约束全局优化问题;巡回推销员问题(TSP:travelling salesman problem).

## 一、Hopfield 网络的学习方式

J. J. Hopfield 于1982年和1984年分别提出了两种神经网络模型,一种是离散的随机模型,一种是连续的确定论模型.这里,我们仅简单介绍 Hopfield 网络(Hopfield Net)的要点.

Hopfield 网络是一种单层网络.令网络由  $N$  个单元组成,它们既是输入单元,又是输出单元.输入或输出图形以  $N$  维矢量表示:

$$X = x_i, \quad 0 \leq i \leq N-1, \quad (4.1)$$

每个分量  $x_i$  的可能取值为  $+1$  或  $-1$  (或  $0, 1$ ), 两个单元  $i, j$  之间的连接权重记为  $t_{ij}$ ,  $t_{ij}$  构成一个矩阵

$$[t_{ij}], \quad 0 \leq i, j \leq N-1, \quad (4.2)$$

即连接矩阵.  $t_{ij}$  可以通过网络的学习过程(即用反馈法逐步改变  $t_{ij}$ ) 来确定.但是,对于这种简单的单层网络,我们实际上已经知道了学习的最终结果,因而无须重复学习过程便可直接写出它的连接矩阵.

例如,对自联想记忆的 Hopfield 网络,令网络储存  $M$  个样板,

$$X^s = x_i^s, \quad 0 \leq i \leq N-1, \quad 0 \leq s \leq M-1, \quad (4.3)$$

则连接矩阵可以确定为

$$t_{ij} = \begin{cases} \sum_{s=0}^{M-1} x_i^s x_j^s, & i \neq j, \\ 0, & i = j, \end{cases} \quad 0 \leq i, j \leq N-1, \quad (4.4)$$

注意,这里取连接矩阵的对角元为零.(4.4)式给出的连接矩阵其实就是本书已介绍过的外积法确定的连接矩阵的形式.

Hopfield 网络是一个动力学系统,在确定连接权重以后,若输入某个图形(矢量),网络将不断演化.一般情况下,系统将趋向某一个定态,称为状态空间中的不动点吸引子(fixed point

attractor), 这个定态便是网络的输出图形(矢量). 具体算法如下:

输入一个矢量  $X = \{x_i, 0 \leq i \leq n-1\}$  作为网络的初始状态 ( $x_i$  的可能值为 +1 或 -1):

$$\mu_i(0) = x_i, \quad 0 \leq i \leq N-1, \quad (4.5)$$

式中  $\mu_i(0)$  同时也就是单元  $i$  在  $t=0$  时刻的输出, 它通过单元之间的连接, 乘以相应的连接权重, 相加后, 成为下一个时刻其它单元的输入.

令  $t$  时刻单元  $i$  的输出为  $\mu_i(t)$  ( $0 \leq i \leq N-1$ ), 则  $t+1$  时刻单元  $j$  的总输入为

$$\sum_{i=0}^{N-1} t_{ij} \mu_i(t), \quad 0 \leq j \leq N-1 \quad (4.6)$$

而  $t+1$  时刻单元  $j$  的输出由下式决定:

$$\mu_j(t+1) \Rightarrow f_h \left( \sum_{i=0}^{N-1} t_{ij} \mu_i(t) \right) \quad 0 \leq j \leq N-1, \quad (4.7)$$

其中,  $f_h$  是一个非线性函数, 这里取为符号函数  $\sigma$

$$f_h(u) = \sigma(u) = \begin{cases} 1, & \text{当 } u > 0, \\ -1, & \text{当 } u \leq 0. \end{cases} \quad (4.8)$$

(4.7) 式是一个迭代方程, 描写 Hopfield 网络的动力学演化过程. 若迭代若干次后, 收敛到某个定态, 于是便给出一个确定的输出.

当储存的样板数与网络的单元数之比  $M/N$  不太大时, 所有储存的样板均为不动点吸引子, 有一定的吸引域 (basin of attraction). 当输入的图形对样板有一定差异时, 网络可通过动力学过程找到相应的样板, 这就是联想记忆的搜索过程.

在 Hopfield 网络中, 由两个状态矢量的方向余弦

$$\cos \alpha = \frac{X \cdot X'}{N} \quad (4.9)$$

度量这两个状态的接近程度. (4.9) 式中

$$X \cdot X' = \sum_{i=1}^{N-1} x_i x'_i \quad (4.10)$$



是矢量  $X$  与  $X'$  的内积. 两个矢量越靠近, 则其方向余弦越大. 且  $X \cdot X = |X|^2 = N$ , 故一个矢量与其自身的方向余弦为 1. 由连接矩阵(4.4)式, 对于一个任意状态  $X = (x_i, 0 \leq i \leq N-1)$ , 有

$$\begin{aligned} \sum_{i=0}^{N-1} t_{ij} x_i &= \sum_{j=0}^{M-1} x_j' \left( \sum_{i=0}^{N-1} x_i x_i' \right) \\ &= N \sum_{j=0}^{M-1} x_j' (X' \cdot X) / N. \end{aligned} \quad (4.11)$$

由动力学方程(4.7)式可以看出, 一个状态一般将趋向于它最接近的样板. 当然, 由于其它样板的干扰, 上述趋势只适用于比值  $M/N$  足够小的情形. 对于随机分布的样板, Hopfield 网络的存储上限为  $M/N \leq 0.15$ .

Lippmann 等人(1987年)提出了一种与 Hopfield 网络大同小异的网络, 称为 Hamming 网络(Hamming Net).

这是一种由上、下两层组成的网络. 令下层单元数为  $N$ , 每个单元输出的可能值为  $x_i = +1$  (或  $-1$ ), 上层单元数为  $M$ , 两层之间连接权重为  $w_{ij}$  ( $0 \leq i \leq N-1, 0 \leq j \leq M-1$ ), 上层单元之间的连接权重为  $t_{kl}$  ( $0 \leq k, l \leq M-1$ ). 此外, 对上层单元而言, 来自下层单元的总输入具有非线性阈值函数的形式, 阈值记为  $\xi_j$  ( $0 \leq j \leq M-1$ ). 上层单元的输出也具有非线性阈值函数的形式, 但取阈值为零. 这里非线性阈值函数的定义为

$$f_j(u - \xi) = \begin{cases} 0, & \text{当 } u < \xi, \\ u - \xi, & \text{当 } \xi \leq u \leq u_c, \\ \text{常数}, & \text{当 } u > u_c, \end{cases} \quad (4.12)$$

其中  $\xi$  为阈值,  $u_c$  是饱和值. 当  $u$  低于阈值时函数值为 0, 高于饱和值时函数值为常数, 而处于二者之间时是线性函数. 在 Hamming 网络中, 假定任何情况下均不会达到饱和.

Hamming 网络的学习问题就是如何确定连接权重  $w_{ij}$  与  $t_{kl}$  以及阈值  $\xi_j$ . 与 Hopfield 网络一样, 它们也是可以直接写出来的.

假定在网络中存入  $M$  个样板

$$X^j = (x_i^j, 0 \leq i \leq N-1), \quad 0 \leq j \leq M-1, \quad (4.13)$$

则上下两层单元之间的连接权重取为

$$w_{ij} = \frac{x_i^j}{2}, \quad 0 \leq i \leq N-1, \quad 0 \leq j \leq M-1. \quad (4.14)$$

阈值  $\xi_j$  取为

$$\xi_j = N/2, \quad 0 \leq j \leq M-1. \quad (4.15)$$

上层单元之间的连接权重取成下述形式:

$$t_{kl} = \begin{cases} 1, & \text{当 } k = l \\ -\epsilon, & \text{当 } k \neq l, \end{cases} \quad 0 \leq k, l \leq M-1, \quad (4.16)$$

其中

$$\epsilon < 1/M.$$

这表示上层单元之间的连接方式是:自身正反馈,相互之间为负反馈,即任一单元均有维持自身并抑制其它单元的趋势.

Hamming 网络的具体运行方式如下:

当下层单元输入一个图形  $X = (x_i, 0 \leq i \leq N-1)$  时,便通过连接  $w_{ij}$  输入信号到上层单元,形成上层单元的初始状态:

$$\mu_j(0) = f_i\left(\sum_{i=0}^{N-1} w_{ij} x_i - \xi_j\right), \quad 0 \leq j \leq M-1. \quad (4.17)$$

上层单元按照下述迭代方程演化:

$$\mu_j(t+1) = f_i\left(\mu_j(t) - \epsilon \sum_{k \neq j} \mu_k(t)\right). \quad (4.18)$$

这个方程是由(4.16)式给出的连接矩阵  $[t_{kl}]$  直接导出的. 上层单元的迭代过程一直进行到收敛于一个定态为止. 这是上层单元之间的一个竞争过程,其定态一般是只有某一个单元  $j$ , 有正的输出, 而其余单元输出为 0. 这可以看成是对下层输入图形的一种分类,若上层信号经过连接  $w_{ij}$  输回到下层单元,便得到标准的存储样板  $X^j = (x_i^j, 0 \leq i \leq N-1)$ .

一个点. 输出由阈值为  $\xi$  的符号函数决定, 输出有  $+1$  和  $-1$  两种可能, 因而可将输入矢量划分为两类, 即把状态空间分成两部分. 这种网络通常以适应(即反馈)方式进行学习, 通过一系列学习过程, 逐步调整连接权重  $w_i$  和阈值  $\xi$ , 使网络能实现所要求的分类. 具体步骤大致如下:

(1) 首先要赋予连接权重和阈值以初值, 初值  $w_i(0)$  ( $0 \leq i \leq N-1$ ) 与  $\xi(0)$  通常取为小的随机数. 当进行第  $t+1$  次训练时, 记连接权重为  $w_i(t)$  ( $0 \leq i \leq N-1$ ), 阈值为  $\xi(t)$  ( $t = 0, 1, \dots$ );

(2) 每次学习时, 取一个新的状态矢量  $(x_i(t), 0 \leq i \leq N-1)$  及所要求的相应输出  $d(t)$ ,  $d(t)$  的可能取值为  $+1$  或  $-1$ , 代表所取矢量应属于哪一类, 例如规定

$$d(t) = \begin{cases} +1, & \text{当输入矢量属于 } A \text{ 类,} \\ -1, & \text{当输入矢量属于 } B \text{ 类;} \end{cases} \quad (4.20)$$

(3) 计算实际输出

$$y(t) = f_h \left( \sum_{i=0}^{N-1} w_i(t) x_i(t) - \xi(t) \right), \quad (4.21)$$

其中  $f_h$  是符号函数,  $w_i(t)$  和  $\xi(t)$  是本次学习开始时的连接权重和阈值;

(4) 如果实际输出  $y(t)$  与所要求的输出  $d(t)$  不一致, 则根据下述规则调整连接权重和阈值:

$$w_i(t+1) = w_i(t) + \eta [d(t) - y(t)] x_i(t), \quad (4.22)$$

$$\xi(t) = \xi(t) - \eta [d(t) - y(t)], \quad (4.23)$$

其中,  $\eta$  是小于 1 的正数. 容易看出, 这种学习规则所导致的效果是: 如果仍输入  $(x_i(t), 0 \leq i \leq N-1)$ , 则 (4.22) 与 (4.23) 式给出的连接权重和阈值的改变具有使实际输出与所要求的输出之间的差距缩小的趋势.

如果恰当选择参加训练的状态矢量, 经过足够多次学习以后, 这种知觉器能够达到所需要的分类要求.

例如,一个  $N = 2$  的单层知觉器,其状态空间是一个二维平面.分类的结果是平面被一条直线

$$w_0x_0 + w_1x_1 - \xi = 0 \quad (4.24)$$

划分为两个区域.容易看出,用这种二维单层知觉器作为分类器划分二维平面时,其分界只可能是一条直线.一个  $N$  维单层知觉器可以把  $N$  维状态空间划分为两部分,其分界是一个  $N - 1$  维的超平面.

单层神经网络作为分类器时,其分类方法与传统的分类方法不同,主要差别可归纳为:

- (1) 神经网络采用并行方式输入数据;
- (2) 神经网络不需要编码和解码;
- (3) 神经网络能进行学习,即用反馈方式对网络进行调整.

但是,总的说来,这种单层网络构成的分类器功能有限,它与传统的分类方法大同小异,更复杂的分类要求利用多层网络才可能实现.

## 二、反向传播法

反向传播法或称为误差传播法是 D. E. Rumelhart 等人于 1986 年提出的一种训练多层网络的学习方式.多层网络是指除输入和输出单元外,还包含不与外界直接连接的隐单元(hidden units)的神经网络.对于很复杂的问题,简单的神经网络往往无能为力,借助于包含若干层隐单元的多层网络是一条可能的途径.

多层网络的功能比单层的网络强得多,但因网络的内部状态不直接受外界控制、不直接向外界显示,这就使网络的训练(学习)变得更加困难.模拟退火方法是训练多层网络的一种方法,下面介绍另一种方法:反向传播法.

反向传播法是用来训练多层前馈式知觉器(multilayer feed-forward perceptron)的一种学习方式.这种网络由输入层、输出层

及若干个隐单元层组成. 前馈式是指每一层单元的状态只影响下一层单元的状态, 即每一层的单元只收到来自下面一层的输入信号, 并只对上面的一层输出信号. 运用反向传播法训练网络, 要求每个单元的非线性函数都是连续可微的, 例如 S 型 (sigmoid) 函数

$$g(\alpha - \xi) = \frac{1}{1 + e^{-(\alpha - \xi)}}, \quad (4.25)$$

其中,  $\alpha$  是单元的输入信号,  $g(\alpha - \xi)$  为输出信号,  $\xi$  是阈值. 令输入层由  $N$  个单元组成, 网络的输入信号是一个  $N$  维矢量  $(x_i, 0 \leq i \leq N-1)$ , 假定每个分量可以连续取值. 令输出层由  $N'$  个单元组成, 输出为  $N'$  维矢量, 其分量可以连续取值, 但若网络用作分类器, 则可取离散值 (如 0 或 1).

反向传播法的基本精神是: 从实际输出与标准输出之间的方差出发, 沿网络内信号传输的相反方向逐层推算出每一层输出的相应偏差, 据此调整各层单元间的连接权重及每个单元的阈值, 经一系列样板的训练, 使网络趋向稳定达到所要求的性能. 具体步骤大致如下:

(1) 给所有的连接权重及阈值赋初值, 一般取为小的随机数.

(2) 选定训练样板的序列, 即一系列输入矢量  $X^p = (x_i^p, 0 \leq i \leq N-1)$  及所要求的相应输出矢量  $(d_j^p, 0 \leq j \leq N'-1)$ , 这里,  $p$  标志参加训练的不同样板. 这个样板序列可以由不同的样板组成, 也可以是若干个样板组成的循环序列.

(3) 在进行某一次训练时, 先由输入矢量  $(x_i, 0 \leq i \leq N-1)$  (这里省去了样板标志  $p$ ) 计算出实际的输出矢量  $(y_j, 0 \leq j \leq N'-1)$ . 以由输入层, 一个隐单元层及输出层组成的网络为例 (包含更多隐单元层的网络可按同样的方式计算), 其计算过程为: 输入层输入矢量  $(x_i, 0 \leq i \leq N-1)$ , 输出矢量为  $(x'_i, 0 \leq i \leq N-1)$ , 有

$$x'_i = g(x_i - \xi_i), \quad (4.26)$$

其中  $\xi_i$  是输入层单元的阈值. 令隐单元层的输入矢量为  $(u_k, 0 \leq k \leq L-1)$ , 有

$$u_k = \sum_{i=0}^{N-1} w_{ik} x'_i, \quad (4.27)$$

其中,  $w_{ik}$  是输入单元与隐单元之间的连接权重. 隐单元层的输出矢量记为  $(u'_k, 0 \leq k \leq L-1)$ , 有

$$u'_k = g(u_k - \xi'_k) \quad (4.28)$$

$\xi'_k$  是隐单元的阈值. 令输出单元的输入矢量为  $(v_j, 0 \leq j \leq N' - 1)$ , 有

$$v_j = \sum_{k=0}^{L-1} T_{kj} u'_k, \quad (4.29)$$

其中  $T_{kj}$  是隐单元与输出单元间的连接强度. 最后可得输出矢量  $(y_j, 0 \leq j \leq N' - 1)$ , 有

$$y_j = g(v_j - \xi_j), \quad (4.30)$$

$\xi_j$  是输出单元的阈值.

(4) 根据评估函数调整连接权重和阈值. 评估函数取为实际输出与标准输出之间的方差

$$E = \sum_p E_p = \sum_p \frac{1}{2} (y_j^p - d_j^p)^2. \quad (4.31)$$

学习过程即为调节各连接权重与阈值使  $E$  趋向其极小值 ( $E_{\min} = 0$ ) 的过程. 在实际训练时, 并不是依次输入全部样板计算出  $E$  以后再进行调整, 而是每训练一次就作一次调整, 即当输入样板  $X^p$  时, 取评估函数为

$$E_p = (y_j^p - d_j^p)^2 / 2. \quad (4.32)$$

(以下计算时省去标志  $p$ .) 当每次修正的幅度很小时, 两种方法的差别不大.

定义

$$\bar{\delta}_j = - \frac{\partial E}{\partial v_j} = g'(v_j - \xi_j)(d_j - y_j) = y_j(1 - y_j)(d_j - y_j), \quad (4.33)$$

$$\delta'_k = -\frac{\partial E}{\partial u_k} = u'_k(1 - u'_k) \sum_j \delta_j T_{kj}, \quad (4.34)$$

$$\delta_i = -\frac{\partial E}{\partial x_i} = x'_i(1 - x'_i) \sum_k \delta'_k w_{ik}. \quad (4.35)$$

在导出这些式子时,用了下述关系

$$\frac{dg}{d\alpha} = g(1 - g), \quad (4.36)$$

$g$  函数的形式由(4.25)式给出. 利用定义(4.33)–(4.35),可以得到评估函数  $E$  随连接权重和 阈值的变化率为:

$$\frac{\partial E}{\partial \xi_j} = \bar{\delta}_j, \quad (4.37)$$

$$\frac{\partial E}{\partial T_{kj}} = -\bar{\delta}_j u'_k, \quad (4.38)$$

$$\frac{\partial E}{\partial \xi'_k} = +\delta'_k, \quad (4.39)$$

$$\frac{\partial E}{\partial w_{ik}} = -\delta'_k x'_i, \quad (4.40)$$

$$\frac{\partial E}{\partial \xi_i} = \delta_i. \quad (4.41)$$

根据(4.37)–(4.41)式,对于第  $t$  次学习,学习规则可表达为

$$\xi_j(t+1) = \xi_j(t) - \eta \bar{\delta}_j, \quad (4.42)$$

$$T_{kj}(t+1) = T_{kj}(t) + \eta \bar{\delta}_j u'_k, \quad (4.43)$$

$$\xi'_k(t+1) = \xi'_k(t) - \eta \delta'_k, \quad (4.44)$$

$$w_{ik}(t+1) = w_{ik}(t) + \eta \delta'_k x'_i, \quad (4.45)$$

$$\xi_i(t+1) = \xi_i(t) - \eta \delta_i, \quad (4.46)$$

其中  $\eta$  是一个小正数,称为学习率. 可以看出阈值的学习规则与连接权重的学习规则十分相似,实际上阈值可以看作是一种特殊的连接权重,相当于与一个输出始终为-1的单元的连接权重.

以上介绍的学习方式相当于一种梯度下降法,当梯度较大或

取较高的学习率时,可能会出现振荡.为了使连接权重和阈值平缓地改变避免出现振荡,可在学习规则中再引入一个“动量项”,例如(4.43)式可改为

$$T_{kj}(t+1) = T_{kj}(t) + \eta \bar{\delta}_j u'_k + \alpha [T_{kj}(t) - T_{kj}(t-1)], \quad (4.47)$$

其中 $\alpha$ 是一常数,“动量项”可以有效地滤掉高频变化, $\alpha$ 值可取为0.9.

对于多层网络,评估函数 $E$ 的形态一般是非凸性的,因此,反向传播法这种梯度下降法能否不被局域极小约束而找到全局极小是一个需要研究的问题,但模拟计算发现局域极小问题实际上并不严重.

还有一个问题是要经过多少次学习才能收敛到稳定的解.发现当训练到一定程度时,网络内部是平缓变化的,但当再引入样板训练时,会出现突变(catastrophe).这是一个尚不清楚的问题.

一个多层网络需要多少隐单元层,每层需多少个隐单元,这要由网络的用途决定,但这并不是唯一确定的,因为采用不同的内部表象时所需要的内部单元数是不同的.

现在介绍一种最简单的多层网络如图4.1所示.这个网络的任务是从输入的两个数中选出较大的数作为输出.网络包含三层单元,输入层由两个单元组成,分别输入数据 $x_0$ 和 $x_1$ ;隐单元层中包含两个单元,其输出为阶跃函数(0或1);输出层由三个单元组成,其中, $z$ 输出的是 $x_0$ 与 $x_1$ 中较大的数, $y_0$ 和 $y_1$ 的可能值为0或1,当 $x_0$ (或 $x_1$ )是较大的数时 $y_0$ (或 $y_1$ )输出为1,否则为0.在图4.1中各单元之间的连线上所标的数即为相应的连接权重,容易验证这些连接权重能保证网络达到所要求的性能.

### 三、适应共振法

1976年,S. Grossberg 提出适应共振理论(ART:adaptive res-



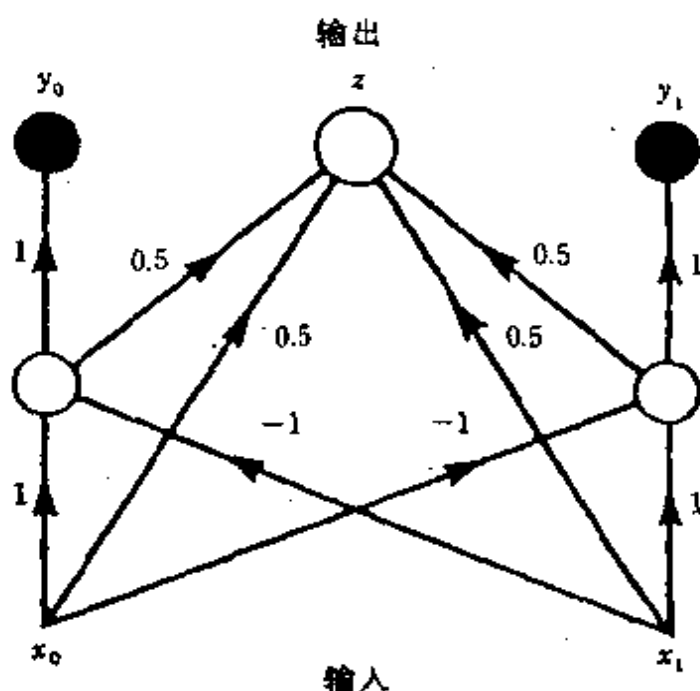


图4.1 从两个数字中挑选较大数字的多层网络

onance theory), 随后, 他又与 G. A. Carpenter 一起研究了基于这种理论的神经网络, 称为 ART 网络或 Carpenter-Grossberg 网络. 这种以适应共振法为学习方式的网络能够识别(或分类)任意多个任意复杂的二元(binary, 即每个分量的可能取值为0或1)输入图像, 其学习过程有自组织和自稳定的特征.

Carpenter-Grossberg 网络可以看作是把由输入层、一个隐单元层和输出层组成的网络的输出层翻下来与输入层合并所构成的网络, 如图4.2所示. 这种网络由底层和顶层组成, 每对顶层单元与底层单元之间的连接由两条连接线组成, 分别代表自上而下(top-down)与自下而上(bottom-up)两种不同的连接权重.

这种网络也可称为 Carpenter-Grossberg 分类器, 底层是特征探测器(feature detectors), 在顶层则给出分类表象(category representation). 一种简单的情形是令一个顶层单元代表一个类别.

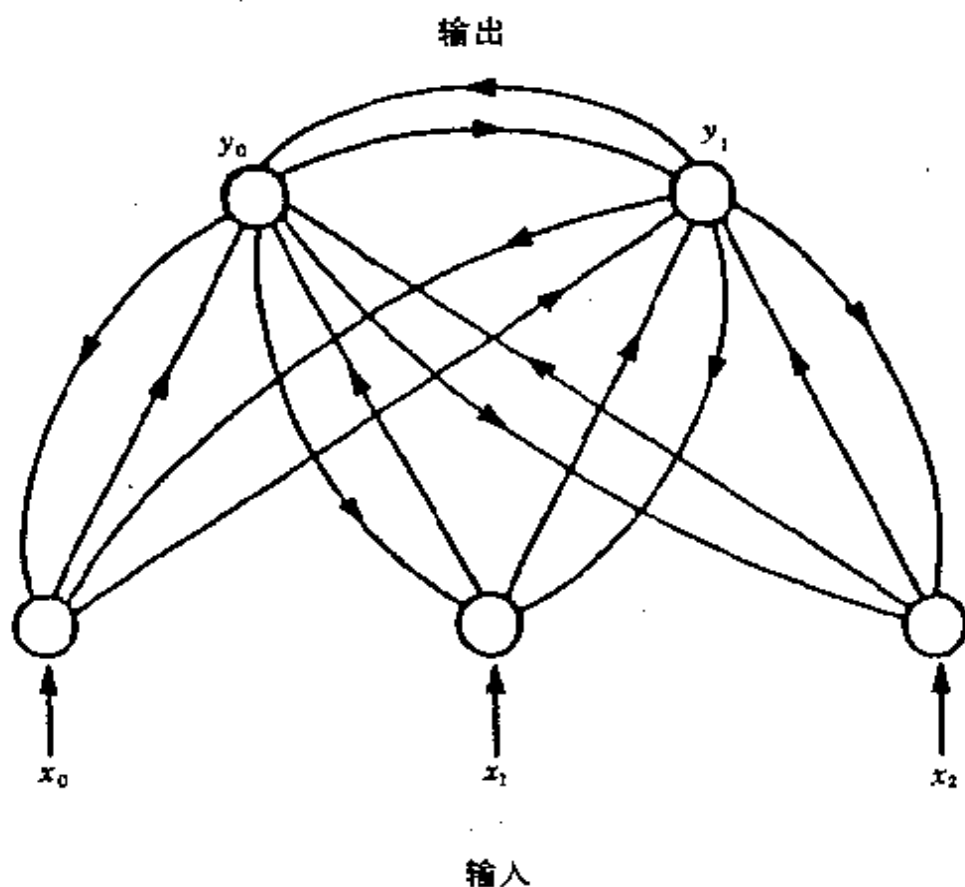


图4.2 Carpenter-Grossberg 网络

现在用一个比较简单的模型说明适应共振法的精神,令底层由  $N$  个单元组成,一个单元  $i$  的输入与输出信号均用  $x_i$  表示,其可能值为 0 或 1;顶层由  $M$  个单元组成,输入信号记为  $\mu_j$ ,顶层单元间互相竞争,只有信号最强的单元对底层有反馈输出,输出信号为 1;自下而上的连接矩阵记为  $B = [b_{ij}]$ ,自上而下的连接矩阵记为  $T = [t_{ij}]$ . 选取一系列样板作为训练的输入矢量,学习过程的步骤如下:

(1) 赋予连接权重初值

$$t_{ij}(0) = 1, \quad 0 \leq i \leq N-1, 0 \leq j \leq M-1, \quad (4.48)$$

$$b_{ij}(0) = \frac{1}{1+N}, \quad 0 \leq i \leq N-1, 0 \leq j \leq M-1. \quad (4.49)$$

选取一个警戒阈值(vigilance threshold)  $\rho$ ; 其范围在

$$0 \leq \rho \leq 1. \quad (4.50)$$

警戒阈值规定了同类样板互相接近的程度, 即只当输入矢量与储存在网络中的某个标准样板互相重叠的程度超过  $\rho$  时才有可能识别为该图形, 否则不能归入此类.

(2) 输入一个新的矢量, 记为  $X = \{x_i, 0 \leq i \leq N-1\}$ ,  $x_i$  的可能值为0或1.

(3) 计算匹配程度(matching scores). 标准样板储存在网络的连接权重中, 每个顶层单元代表一个标准样板. 输入矢量与标准样板的匹配程度由自底层向顶层单元的输入  $\mu_j$  描写.  $\mu_j$  取决于底层单元的输入  $x_i$  及自下而上的连接权重  $b_{ij}$ .

$$\mu_j = \sum_{i=0}^{N-1} b_{ij}(t) x_i, \quad 0 \leq j \leq M-1. \quad (4.51)$$

连接权重  $b_{ij}(t)$  中的时间  $t$  标志训练的次数, 每选取一个新的输入矢量算作一次训练, (4.51)式中出现的是上次学习的结果.

(4) 选择最佳匹配样板

$$\mu_{j^*} = \max_{j \in J} \{\mu_j\}. \quad (4.52)$$

这里,  $J$  是上层单元中未被搜索过的单元的集合(这一点下面将作进一步说明),  $j^*$  是  $J$  中输入  $\mu_j$  最大的单元, 即集合  $J$  的各单元中  $j^*$  所代表的标准样板与输入有最佳匹配.

顶层单元中, 只有  $j^*$  对底层有反馈输出, 并且输出为1. 于是底层收到的来自顶层的反馈信号由自上而下的连接权重  $t_{ij^*}$  决定:

$$Y = \{y_i = t_{ij^*}, 0 \leq i \leq N-1\}. \quad (4.53)$$

$Y$  实际上就是  $j^*$  所代表的样板(注意: 学习规则将保证  $t_{ij}$  的可能值为0或1, 故  $y_i$  的可能值也是0或1).

(5) 警戒测试(vigilance test), 即以警戒阈值  $\rho$  为标准将标准样板  $Y$  与输入信号  $X$  作比较. 输入矢量  $X$  中非零(即  $x_i = 1$ ) 的分

这个学习规则表示只改变与顶层单元  $j$  连接的权重. 在(4.53)式中已指出了自上而下的连接权重  $t_{ij}$  与网络内储存的样板的关系, (4.58) 式反映了经过用输入矢量  $X$  训练以后,  $X$  所属类别的标准样板作了一定的调整: 新的样板中具有非零值(即值为 1)的分量只是那些在原来的样板以及在输入矢量中均有非零值(即值为 1)的分量. 粗略地说, 这相当于以两者的共同特征作为新的标准.  $t_{ij} \cdot (t+1)$  的可能值仍为 0 或 1. (4.59) 式给出的自下而上的连接权重  $b_{ij} \cdot (t+1)$  与  $t_{ij} \cdot (t+1)$  成比例, 其比例系数是为了保证

$$\sum_{i=0}^{N-1} b_{ij} \cdot \leq 1, \quad (4.60)$$

即一个顶层单元与底层全部单元的总连接强度接近于 1, 只有这样, 对不同的  $\mu_j$  进行比较才有意义.

另外, 如果对于输入矢量  $X$ , 搜索了全部储存的样板后均不满足警戒条件时, 网络将按(4.58) 和(4.59) 式建立一个新的类别, 其标准样板即为输入矢量.

(8) 转向步骤(2)以新的输入矢量重新开始搜索和学习.

以上介绍的适应共振学习法包含了快速搜索与连接权重的缓慢调整两个不同时标的过程. 在更复杂的 ART 网络中, 学习过程要用快时标和慢时标两组非线性微分方程描写. 经过一系列学习过程以后, 网络将趋向于自稳定, 分类不再改变. 达到自稳定后, 对任一输入矢量将直接进行分类(识别), 不再经过搜索过程, 因为它将属于响应最强的类别. 这种特征使网络识别一个输入的时间并不因储存的类别增多而变长. 若输入的是一个不能归入任何已知类别的新矢量, 则网络将重新进行搜索和适应性调节.

适应共振学习法通常被认为是无导师的学习, 因为网络是自动进行调整的. 但实际上  $\rho$  的数值的选取是由人确定的, 在这个意义上, 也应看作是有导师的. 所选取的  $\rho$  值越大(越接近于 1) 表示

对分类的要求越精细;相反,很小的  $\rho$  值表示只要求很粗的分类。

通常的控制论一般是利用反馈的方法. 在适应共振法中, 顶层对底层只进行有选择的反馈, 同时又赋予系统一个目标(警戒阈值). 这是一种抑制与激励相结合的办法, 与生物学中的行为比较接近. 我们可以用一个人骑驴子的比喻来说明这种方法: 人和驴子之间并不把各自掌握的全部信息告诉对方, 只交换最必要的信息, 人指挥驴子要靠双重手段, 一方面用鞭子抽(抑制), 另一方面又用胡萝卜在前面诱导(激励). 胡萝卜加鞭子可以更有效地驱赶驴子奔向我们所要达到的目的地。

图 4.3 给出了依次输入 5 个不同矢量时网络的学习过程. 输入矢量是用  $N = 8 \times 8 = 64$  个格子表示的字符, 警戒阈值取为  $\rho = 0.9$ . 当输入第一个矢量(字母 C) 时, 网络内部处于 (4.48) 和 (4.49) 式所表示的初始状态, 学习结果是调整连接把输入作为样板储存在网络内. 当输入第二个矢量(字母 E) 以及第三个矢量(字母 F) 时, 由于与已储存在网络内的样板差别较大(以  $\rho = 0.9$  衡量), 故依次又存入了两个样板. 当输入第四个矢量时, 由于它与已存于网络内的第三个样板差别足够小(此时, 由 (4.54) 与 (4.55) 式, 有  $n_{xy}/n_x = 18/19 > \rho = 0.9$ ), 这个输入矢量被识别为属于第三个样板所代表的类别. 同时, 这个样板也作了相应的调整, 它相当于原来的第三个样板与第四个输入矢量互相重叠部分所组成的图形. 最后, 当输入第五个矢量时, 它与已存储在网络中的 3 个样板差别均较大(例如将它与和它最接近的第三个样板比较, 有  $n_{xy}/n_x = 17/19 < \rho = 0.9$ ), 这个输入矢量不能归入任何一个已有的类别, 结果又建立了第四个样板. 如果取  $\rho = 0.8$ , 则第五个输入矢量便仍属于第三个类别, 同时, 这个样板要作进一步调整. 这个例子说明, 警戒阈值  $\rho$  的选择对分类结果有重要影响。

输入	每次输入后所存的样板
C	C
E	CE
F	CEF
F	CEF
F	CEFF

图 4.3 适应共振法学习过程示例

#### 四、神经网络用于优化问题

几乎在一切实际领域中,我们都会面临各式各样的优化问题.有一些是无条件的优化问题或约束条件很少的优化问题,传统的计算机可以很有效地解决这类问题.但我们常会遇到一些有很多约束条件的优化问题,传统计算机对此往往效率不高,甚至无能为力,而处理这类问题,神经网络有其特有的优势.

在第三章中,曾介绍了如何用模拟退火方法解决全局优化问

题. 通常处理优化问题有几种不同的方法, 至少包括:

(1) 分割——各个击破法. 这是人工智能(AI)经常采用的战术, 其精神是: 把一个大的问题(大系统)分割为若干部分, 而每一部分的优化是凸性优化, 就容易解决. 这种方法的缺陷是: 每一部分分别实现优化并不能保证整个问题达到全局性优化.

(2) 改进的分割——各个击破法. 这种方法不采用固定的分割方式, 而是逐步加以调整以期达到更好的效果.

(3) 空间填充曲线法. 这是数学家 Peano 提出的方法. 例如, 当处理分布在二维空间中的信息时, 可采用扫描法使之与一个一维分布的信息一一对应. 但是, 两种表示的近邻关系并不相同, 在二维空间中互为近邻的点在一维空间中可能相距很远. Peano 的方法就是要克服这种缺点, 把二维中相近的部分扫描到一维空间中仍保持相近. 这里不介绍详细数学内容, 只举一个实际应用例子.

美国有一个城市每天要派司机为老年人送食物. 由于老年人的数目以及住址经常有变化, 司机的数目也会变化, 若每天为每个司机画一张地图就很麻烦. 于是, 他们采用了 Peano 的方法——把地址写成卡片, 按 Peano 法把卡片排成一串. 这样, 二维分布的地址变为一维的, 而且互相靠近的卡片所相应的实际距离也较近. 司机送食物时, 只要按司机的人数把卡片分成若干叠, 便可达到比较合理的分工. 这种方法还可用于处理机器人的空间行走问题.

(4) 模拟退火方法. 这在第三章中已经介绍过, 这种方法在传统计算机和神经网络中均可以运用.

(5) 利用神经网络解决优化问题. 前面已经指出, 它的优点是可以处理有很多约束条件的优化问题.

以下, 我们介绍如何用神经网络解决 TSP 问题, 即巡回推销员问题(Travelling Salesman Problem).

TSP 问题的提法如下: 如果一个人要对  $N$  个城市中每个城市

都访问一次且仅访问一次,那么,怎样安排访问的次序才能使总的行程最短?这显然是一个约束优化问题,约束条件为:每个城市必须且只能访问一次,每次必须而且只能去一个城市,总城市数固定为  $N$ . 优化即为要求总行程最短.

TSP 问题是优化问题中的一种典型,随着城市数目  $N$  的增大,问题的复杂性将急剧增加. 当城市的数目为  $N$  时,各种可能方案的数目为

$$\# = N! / 2N = (N - 1)! / 2. \quad (4.61)$$

这是由于  $N$  个城市有  $N!$  种排列,但两个顺序完全相反的方案行程相同,故要除以 2;以哪个城市开始旅行都是一样的,故还要除以  $N$ . 表 4.1 中列出了  $\#$  随  $N$  的变化:当  $N = 3$  时  $\# = 1$ ,只有一种方案;当  $N = 10$  时,就有 18 万种方案. 如果考虑 30 个城市,则有  $\# \approx 4.4 \times 10^{30}$  种方案,若有 100 个城市,则可能方案数将达到  $\# \approx 4.67 \times 10^{155}$  种. 这种现象称为组合爆炸(combinatorial explosion),是一个典型的 NP-complete 难题,即完全非多项式问题(non-determined polynomial complete problem). 当城市较多时,用通常的搜索方法是极费时间甚至无法实现的.

表 4.1 TSP 问题中可能方案数  $\#$  随城市数  $N$  的变化

$N = 3$	$\# = 1$
$N = 4$	$\# = 3$
$N = 5$	$\# = 12$
$N = 10$	$\# = 181440$

解决 TSP 问题有几种不同的途径,包括:

#### (1) 消除交叉路径法

这是一种基于三角不等式的统计方法,是 S. Liu 于 1965 年提出的. 如图 4.4 所示,有交叉路径的图 4.4(a) 改为无交叉路径的图



4.4(b)时,便使这一部分的行程缩短,因为一个三角形的两边之和大于第三条边长.

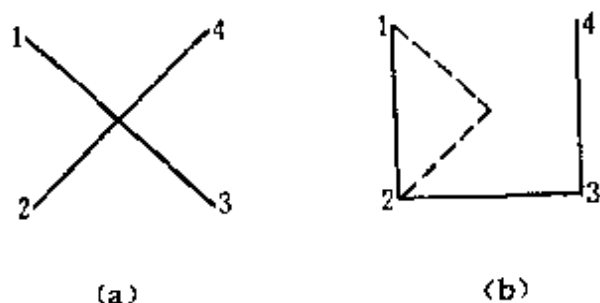


图 4.4 消除交叉路径法

图4.5是一个 TSP 问题的实例.有100个城市杂乱分布在边长为1的方形区域内,城市间的平均距离可估计为 $1/2$ .图4.5(a)是一个任意选取的方案,总行程  $L \approx 0.5 \times 100 = 50$ .图4.5(b)所示的方案交叉路径已大为减少,总行程为  $L = 11.19$ .图4.5(c)已完全没有交叉路径,总行程只有  $L = 8.89$ .

## (2) 模拟退火方法

这是作者本人于1988年提出的方法,在寻找总行程  $L$  的最小值的过程中,允许  $L$  出现一定噪声水平(以等效温度  $T(t)$  描写)之内的增加:

若  $L$  下降,则令其概率  $P = 1$

若  $L$  增加,则令其概率  $P = C \exp[-\Delta L/T(t)]$  这种方法能够克服局域极小而找到最优方案.

## (3) 利用神经网络作区域性搜索

这是寻找局域极小的方法,要求用足够多的搜索初态以达到相当的覆盖范围.用这种方法一般不能找到全局最优解,但它能在很短的时间内达到比较好的解.在处理实际问题时,能够快速求得大体满意的解答是很重要的.实际上,人在复杂的条件下也经常是

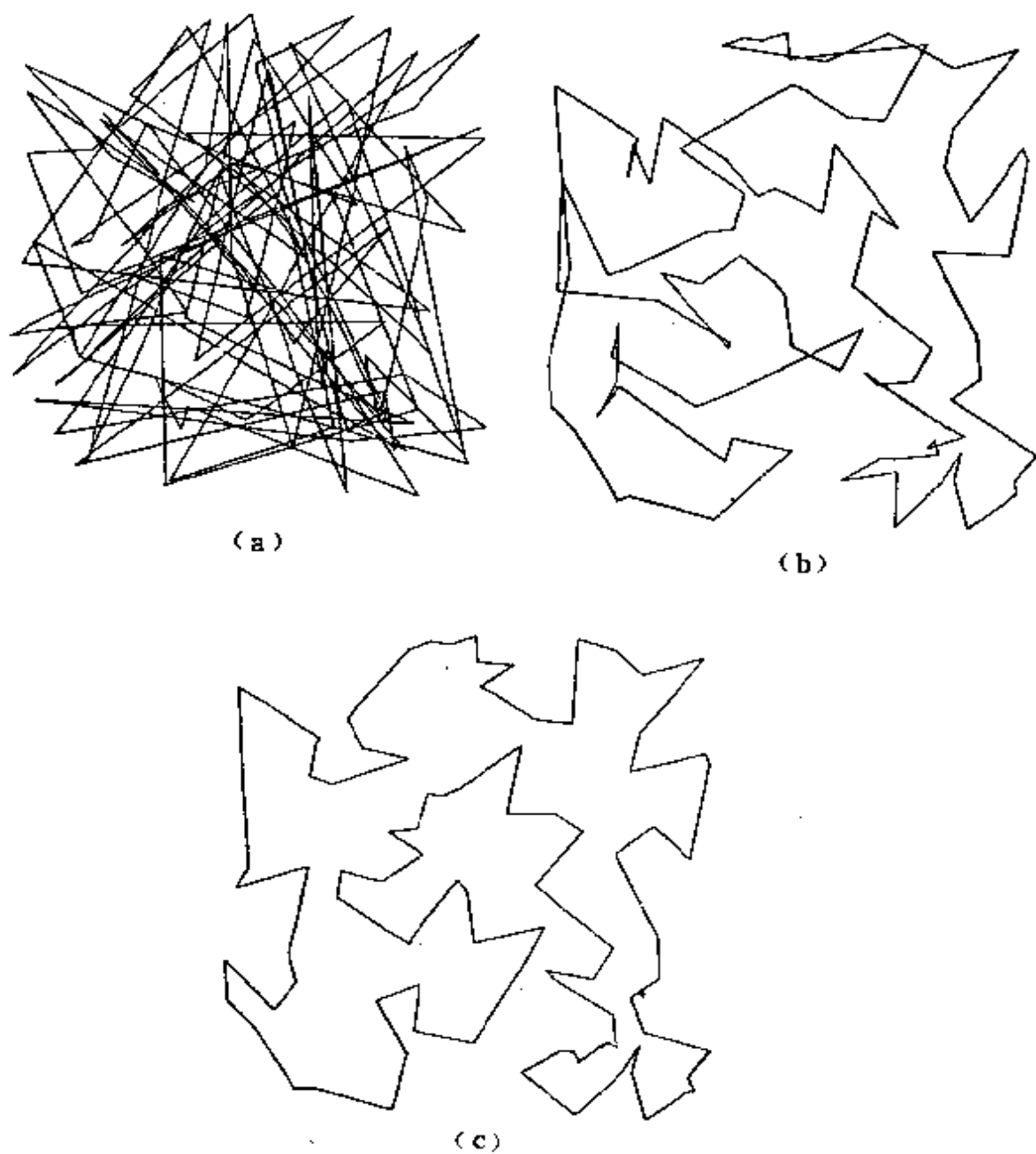


图4.5 TSP 问题的一个例子(100个城市)

很快作出较好的决策,而不一定去追求最佳方案,这对人类的生存和发展有极重要的意义.

用神经网络解决 TSP 问题是 J. J. Hopfield 和 D. W. Tank 于 1985 年提出来的. G. V. Wilson 与 G. S. Pawly 在 1988 年对这种方法的稳定性提出了疑问. 本书作者澄清了这个问题,指出 Hopfield-Tank 的理论没有原则性的错误,只是一些局部性的小错误影响了结果,并且对这个理论作了改进.

Hopfield-Tank 解决 TSP 问题的方法是将 TSP 问题对应到 Hopfield 网络上,通过网络的演化和趋向定态,自动搜索局域优化解. 在本章的开头已对 Hopfield 网络作了简单的介绍. Hopfield 网络是一个动力学系统,其演化过程可以引入一个 Lyapunov 函数或称能量函数来描写:

$$E = -\frac{1}{2} \sum_i \sum_{j \neq i} T_{ij} V_i V_j - \sum_i V_i (I_i - \xi_i), \quad (4.62)$$

这里用了与本章第一节不同的符号,并且写成更一般的形式,(第一节的(4.7)式只相当于(4.62)式右端的第一项), $T_{ij}$  是连接矩阵,假定它是对称的:

$$T_{ij} = T_{ji}, \quad (4.63)$$

$V_i$  是  $i$  单元的输出, $i$  单元的输入为

$$U_i = -\frac{\partial E}{\partial V_i} = \sum_{j \neq i} T_{ij} V_j + I_i - \xi_i. \quad (4.64)$$

而输出与输入的关系为

$$V_i = \text{sgn}(U_i) = \text{sgn}\left(\sum_{j \neq i} T_{ij} V_j + I_i - \xi_i\right). \quad (4.65)$$

这里  $(I_i - \xi_i)$  相当于  $i$  单元的阈值. 容易看到,神经网络的这种演化模型与固体物理学中磁性粒子组成的多体系统的 Ising 模型同构. $V_i$  相当于自旋的取向,(4.62) 式右端第一项描写自旋相互作用,第二项描写局域外场作用,其中  $\xi_i$  为固定外场而  $I_i$  为可调外场.(4.64) 式决定的  $U_i$  即为总局域场,由自治场和外场组成,它

决定了局域自旋的取向.

由(4.62)与(4.64)式,可以得到一次迭代引起的能量改变为

$$\Delta E = - \sum_i U_i \Delta V_i, \quad (4.66)$$

而由(4.65)式容易看出

$$U_i \Delta V_i \geq 0, \quad (4.67)$$

因此,在演化过程中始终有

$$\Delta E \leq 0, \quad (4.68)$$

即系统总是向着能量不断下降的方向变化. 由于  $E$  是有下界的, 因而系统将趋向一个能量为局域极小的定态.

对用于联想记忆的 Hopfield 网络, 令储存样板为  $V^m = (V_i, 0 \leq i \leq N-1), 0 \leq m \leq M-1$ , 则连接矩阵为

$$T_{ij} = \sum_m V_i^m V_j^m, \quad (4.69)$$

假定阈值为0, 则能量函数为

$$\begin{aligned} E &= - \frac{1}{2} \sum_i \sum_{j \neq i} \sum_m V_i^m V_j^m V_i V_j \\ &= - \frac{1}{2} \sum_m (V^m \cdot V)^2 + \frac{1}{2} MN. \end{aligned} \quad (4.70)$$

(4.70)式右端第二项是不重要的常数项, 以下将它舍去, 假定样板在状态空间中的分布接近于互相正交, 则在某一个样板  $m_0$  附近, 其它样板对  $E$  的贡献是不重要的, 有

$$E \approx - \frac{1}{2} (V^{m_0} \cdot V)^2. \quad (4.71)$$

其极小值即在  $V = V^{m_0}$  处, 有  $E_{\min} \approx -N^2/2$ . 因此, 在样板近于正交条件下, 能量函数在样板处为局域极小, 代表一个定态, 而一个与某样板比较接近的状态将趋向于相应的样板态.

上面说的是离散式的 Hopfield 网络, 还有一种是连续式的网络模型, 系统的变化由微分方程描写, 也同样存在相应的能量函数.

极小,而且要保证满足问题的约束条件. 这些约束条件可以用 Lagrange 乘子的方式引入到能量函数中来. Hopfield-Tank 所给出的 TSP 问题的能量函数形式为:

$$\begin{aligned}
 H = & \frac{a}{2} \sum_x \sum_i \sum_{j \neq i} V_{xi} V_{xj} + \frac{b}{2} \sum_i \sum_x \sum_{y \neq x} V_{xi} V_{yi} \\
 & + \frac{c}{2} \left( \sum_x \sum_i V_{xi} - N \right)^2 \\
 & + \frac{d}{2} \sum_x \sum_{y \neq x} \sum_i d_{xy} V_{xi} (V_{y,i+1} + V_{y,i-1}). \quad (4.72)
 \end{aligned}$$

单元  $(x, i)$  的输入信号  $U_{xi}$  满足的方程为

$$\frac{dU_{xi}}{dt} = -\frac{U_{xi}}{\tau} - \frac{\partial H}{\partial U_{xi}}. \quad (4.73)$$

单元  $(x, i)$  的输出信号  $V_{xi}$  为

$$V_{xi} = g(U_{xi}), \quad (4.74)$$

其中  $g$  为 S 型函数(sigmoid), 具体形式取为

$$g(U_{xi}) = \frac{1}{2} (1 + \tanh(U_{xi}/U_0)), \quad (4.75)$$

以上各式中,  $a, b, c, d$  及  $U_0$  是可调参数, 均取正值, 而  $N$  为总城市数, 在计算中也作为可调参数, 不一定等于真正的城市数,  $d_{xy}$  是城市  $x$  与  $y$  间的距离. 由 (4.74) 式决定的  $V_{xi}$  是可以连续变化的模拟量, 其取值范围为

$$0 \leq V_{xi} \leq 1 \quad (4.76)$$

现在对 (4.72) — (4.74) 所给出的 TSP 问题的数学模型作一些说明.

(4.72) 式定义的能量函数  $H$  是一个正函数, (4.73) 与 (4.74) 式决定了  $H$  总是趋向于极小值. (4.72) 右端前三项均是描写约束条件的, 第一项表示每个城市至多只能去一次, 满足此条件时, 此项为 0, 否则将贡献一个正数, 即当这个约束条件被破坏时

会使  $H$  增大. 第二项表示每一次至多只能去一个城市, 满足此条件时, 此项为 0, 否则将贡献一个正数, 即当这个约束条件被破坏时也会使  $H$  增大. 第三项表示总城市数为  $N$ , 如果是对  $N$  个城市中每个城市访问一次, 则此项为 0, 否则也贡献一个正数. 以上是以一种模糊的方式表示约束条件, 能量函数趋向极小与约束条件趋向满足, 二者在趋势上是一致的. 能量函数中的第四项即相应于旅行的总行程, 由于逆向旅行和正向旅行是对称的, 故把此项写成两项之和, 真正满足约束条件的优化解相应于  $H$  函数的前三项为零而第四项取极小.

(4.73) 是输入信号满足的方程,  $U_{xi}$  的变化由两项组成. 第二项是能量函数的斜率, 这一项决定了能量函数向极小移动的趋势, 而第一项则是一个衰减项, 大体上相应于电路中的电阻耗散效应.

(4.74) 式给出了单元非线性行为的具体形式, 选用了 S 型函数是因为它比较接近自然界中神经元的真实行为.

在 Hopfield-Tank 的计算中, 对于 10 个城市情形, 所选择的参数数值为

$$a = b = 500, \quad c = 200, \quad d = 500,$$

$$U_0 = 0.02, \quad N = 15,$$

参数  $a, b, c$  和  $d$  的大小代表能量函数中各项的权重,  $U_0$  描写非线性函数的增益率, 参数的具体数值多半是凭经验选定的.

Hopfield-Tank 对分布于边长为 1 的方形区域内的 10 个城市的 TSP 问题作了数字模拟计算. 在表 4.1 中曾给出,  $N = 10$  时, 可能的旅行方案数为  $\# = 181440$ . 在作模拟计算时, 随机选取了 20 个初始状态进行搜索, 结果有 16 个收敛到合理的结果. 其中, 有一半相应于总行程最短或次短的解. 选择精度达到  $10^{-4} - 10^{-5}$ , 网络趋向定态所需的时间大约为网络特征时间的若干倍. 对于生物神经网络, 收敛时间为 10—100ms, 对于半导体电路组成的网络, 收敛

时间为0.1—1微秒. 因此, 神经网络能在极短的时间内找到 TSP 问题的相当好的解.

但是, G. V. Wilson 和 G. S. Pawley (1988年) 按 Hopfield-Tank 的方法进行计算, 并不能重复其结果. 用 Hopfield-Tank 的参数计算10个城市的 TSP 问题时, 发现从100个不同的初态出发进行试探时, 只有15个不超过1000步迭代就收敛到合理的方案, 而有40个不收敛到合理的解, 另外40个在1000步迭代后仍不收敛, 总的效果仅比无规选择的方案略好一点, 这与 Hopfield-Tank 的结论很不相同.

本书作者重新考察了 Hopfield-Tank 用神经网络解决 TSP 问题的算法, 发现他们并没有原则性的错误, 出现的问题是由一些局部性毛病引起的. 主要是:

(1) 在(4.73)式中的阻尼项是不应当有的, 这个阻尼项导致能量耗散, 影响了搜索过程. 从网络的连接看, 这个衰减项相应于取负值对角矩阵元, 这是神经元的一种负反馈式的自我对话方式.

(2) 用 S 型函数作为单元的非线性函数((4.74)和(4.75)式)其增益太弱. 这种形式的非线性函数通常用于处理有噪声的问题, 而 TSP 问题中根本没有噪声.

(3) 能量函数(4.72)中第三项所代表的约束条件仅仅是一种必要条件而不是充分条件. 例如若每次均去同一个城市, 则仍能使此项为零, 这时第二项也为零, 仅第一项有小的贡献. 这种显然不合理的解从能量上看仍然是一个好的解.

为了克服上述缺点, 我们把(4.72)~(4.74)式稍加改进, 成为下述形式:

$$H = \frac{a}{2} \sum_x \sum_i \sum_{j \neq i} V_{xi} V_{xj} + \frac{b}{2} \sum_i \sum_x \sum_{j \neq x} V_{xi} V_{xj} \\ + \frac{c}{2} \sum_x \left( \sum_i V_{xi} - 1 \right)^2 + \frac{c}{2} \sum_i \left( \sum_x V_{xi} - 1 \right)^2$$

$$+ \frac{d}{2} \sum_x \sum_{y \neq x} \sum_i d_{xy} V_{xi} (V_{y,i+1} + V_{y,i-1}), \quad (4.77)$$

$$\frac{dU_{xi}}{dt} = - \frac{\partial H}{\partial V_{xi}}, \quad (4.78)$$

$$V_{xi} = \theta(V_{xi}), \quad (4.79)$$

其中

$$\theta(u) = \begin{cases} 1, & \text{当 } u > 0, \\ 0, & \text{当 } u \leq 0. \end{cases} \quad (4.80)$$

与(4.72)——(4.74)比较,与 Hopfield-Tank 不同之处有:

(1) 用(4.77)式右端的第三第四项代替了(4.72)式右端第三项,其意义是:每个城市必须而且只能访问一次,每一次必须而且只能访问一个城市 ( $V_{xi}$  矩阵每行之和为1,每列之和也为1). 这种对约束条件的充分必要的表达方式对不合理方案的出现起更严格的限制作用.

(2) 在(4.73)式中舍去了衰减项,变为(4.78)式所给的形式.

(3) 把(4.74)中的 S 型函数换成(4.79)式所给的阶跃函数,这使增益加强.

在作数字模拟计算时,取 Lagrange 乘子  $a=b=c=1$ , 分别对  $N=5, 10, 15, 20, 25$  等情形作了计算,发现这种改进了的模型能很快得 TSP 问题的解. 我们称之为快速 TSP 模型 (fast TSP model). 用计算机进行计算,对5个城市的 TSP 问题在一分钟内便能得到12个合理的优化方案. 计算10个城市的 TSP 问题时,收敛到91个合理的方案(1000次迭代),其中包括最佳方案在内.

图4.7(a)——(j)是对5个城市情形的10个不同搜索过程的结果,每幅图包含三个部分,最左边是最终输出,中间是斜率  $-\partial H / \partial V_{xi}$ , 右边是输入  $U_{xi}$ . 图中,实心方块代表正数,空心方块代表负



数, 方块的面积表示绝对值的大小. 在表 4.2 中列出了对 5 个城市情形的快速 TSP 模型所得总行程的数据.

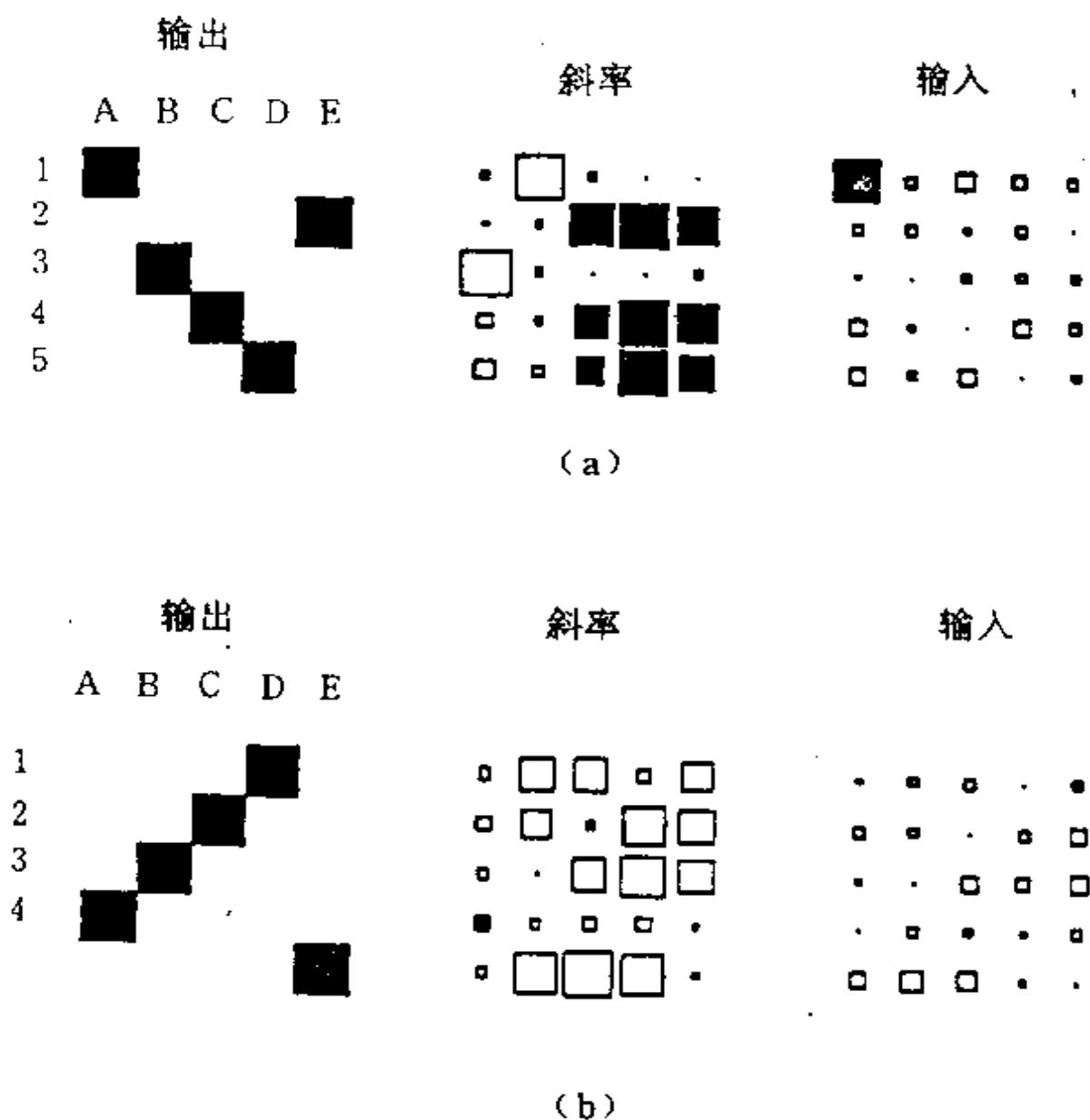
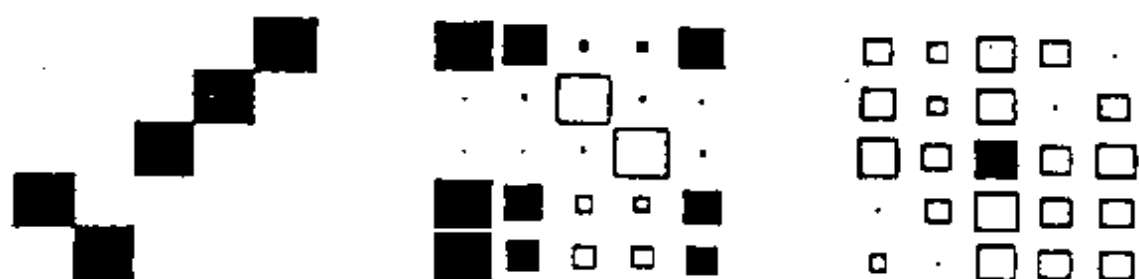


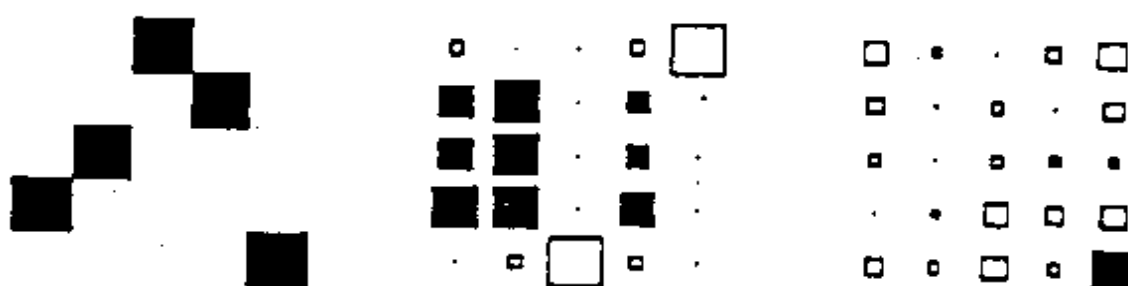
图 4.7



(c)

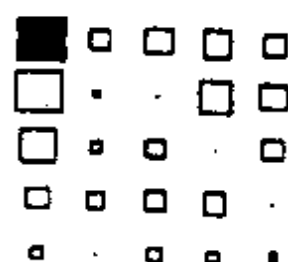
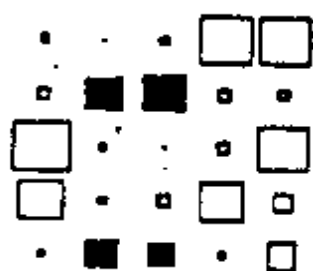
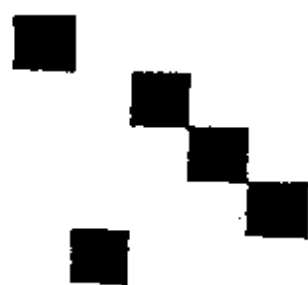


(d)

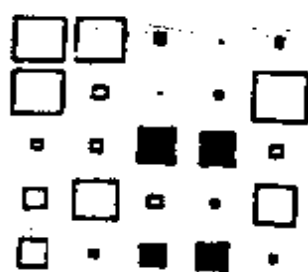
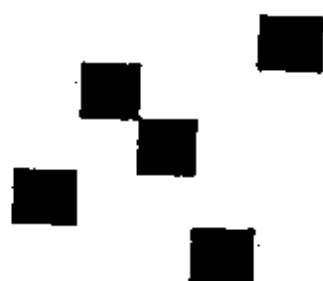


(e)

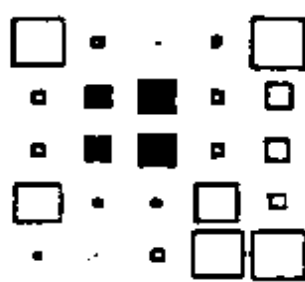
图4.7



(f)

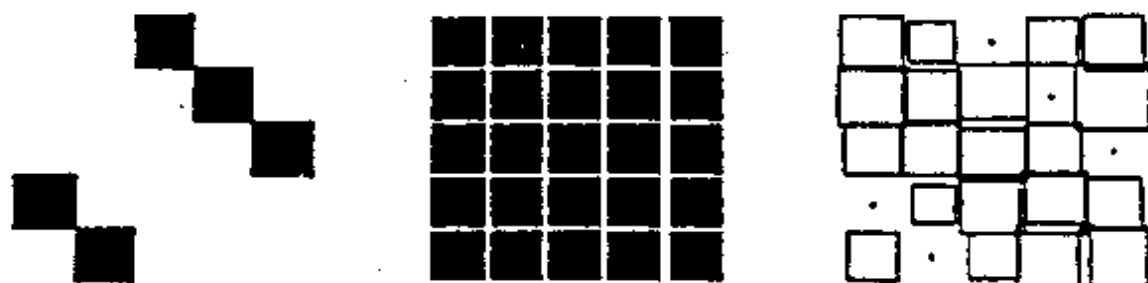


(g)

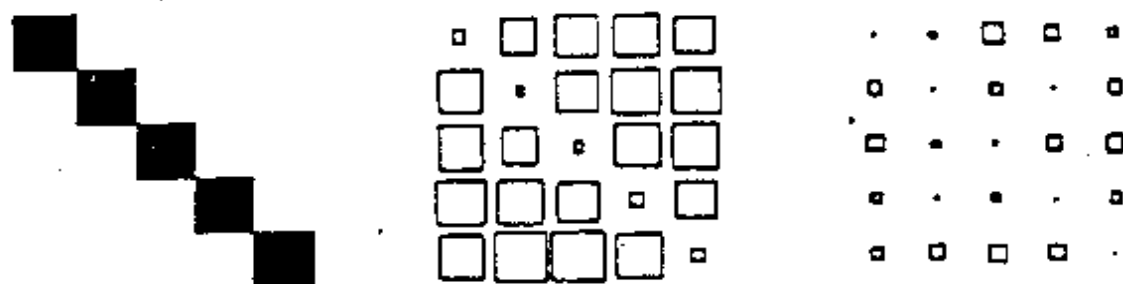


(h)

图4.7



(i)



(j)

图4.7 TSP 问题的解(5个城市情形)

表4.2 对5个城市快速 TSP 模型

城 市	A B C D E	总 行 程
访问次序	1 3 5 4 2	2.39653
	3 1 5 4 2	1.89036
	1 3 2 4 5	1.89036
	2 5 3 1 4	2.92167
	5 2 3 4 1	2.35754
	1 2 3 4 5	1.8324
	4 1 5 2 3	2.35754
	1 4 3 5 2	2.86371
	4 5 2 1 3	2.36542
	5 3 4 2 1	2.78863
	4 2 3 5 1	1.96524

最后简单介绍一下神经网络的另外一种应用：以并行和同时处理的方式实现模 / 数 (A/D) 转换，即输入一个模拟量  $V_a$ ，要求输出为  $N$  个分量的矢量、每个分量的可能值为 0 或 1。例如，输入一个十进制的数 13，要求输出为相应的二进制表示 (1,1,0,1)。这个问题的相应的能量函数可以写成

$$H = - \left( V_a - \sum_{i=0}^N 2^i V_i \right)^2 + \sum_i 2^i |V_i(V_i - 1)| \quad (4.81)$$

(4.81) 式右端第一项是要求输出的二进制数与输入的数尽量接近；第二项是要求输出矢量的分量取值为 0 或 1，这相当于一种约束条件。

的输入信号 $\{e_i\}$ 通过眼睛的作用 $[S]$ 变换为特征 $\{f_i\}$ ,然后通过大脑神经网络的运算 $[T]$ 变换为记忆样板 $\{m_i\}$ .这里, $[S]$ 代表感觉映射(sensory mapping),也就是所需的预处理,而 $[T]$ 则代表联想记忆映射.

人的眼睛内并没有任何运算过程,它的预处理过程完全是靠它的特定的结构实现的.

眼的前部是一个自动调节的光学系统,由角膜,前房水状液,巩膜、瞳孔、晶状体及玻璃体组成,其作用是使外界射入的光聚焦在视网膜上成像.视网膜上密布着感光细胞,包括大约13000万个视杆细胞(rods)和650万个视锥细胞(cones).视杆细胞对视锥细胞对亮度的敏感程度强500倍左右,视锥细胞对颜色的敏感程度则远高于视杆细胞.感光细胞在视网膜上的分布是不均匀的.在视网膜中心,有一个很小的区域称为中央凹,这里高度密集地分布着视锥细胞,大约每平方毫米有15万个,这是视网膜最敏感的区域,有一条直通大脑的传感神经.从中心到边缘视锥细胞逐渐稀疏,密度大约呈指数形式下降(称为PEG分布).光线到达视网膜时,感光细胞内部发生光化学反应,其信号并不直接传到大脑,而是先在感光细胞底部的视神经上进行信号处理,参与处理的最后一层细胞称为视网膜神经节细胞,它们的轴突通过盲斑离开视网膜形成视神经,组成大约包括100万条神经纤维的视神经束.视神经向外侧膝状体提供信息,外侧膝状体是眼与大脑之间主要视觉核团,其轴突称为视放射.对于人类或猴子,外侧膝状体通过视放射传出的信号传到纹状皮层,这是初级视皮层中的接受区.人的眼睛就是依靠这种精密、复杂的结构实现对光信号的预处理.

人眼在进行预处理时有哪些奥妙之处值得我们在研制神经网络的预处理系统时借鉴呢?首先,我们发现眼睛具有时间不变性(time invariant),空间不变性(space invariant)和亮度不变性(brightness invariant).我们曾指出过,不变性就是特征

(feature),是适合于大脑处理的表象.当然,这里所指的不变性并非绝对不变性,而是相对变化很缓,即随时间或空间变化不大及亮度变化不强烈时显示为近似不变.

以下,分别说明眼睛的这三种不变性.

### 1. 时间不变性

我们可以用一个例子说明眼睛的时间不变性.比如你有一个朋友是天文学家,每天晚上看星星,看到后来都快要看疯了.忽然有一天,他告诉你发现了一颗很暗的星.你并不相信,以为他产生了幻觉,让他把照片拍下来.但是,在照片上看不出有这颗星,而天文学家仍坚持确实看到它了.那么,为什么人的眼睛能够发现这颗星而在照片上却显示不出来呢?原来,这是由于大气湍流的干扰所致.大气湍流使折射率发生扰动,星的像在底片上的位置不断闪动,对于很暗的星,底片无法充分形成明晰可辨的图像.但是人的眼睛具有时间不变性,能够克服大气湍流的干扰,因而能够找到这颗星.人眼的时间不变性大体上相当于一种快速跟踪的能力,使得不断闪动的星光能够在视网膜中固定的位置上成像,从而使人觉察到很暗的星星的存在.这方面的问题,R. L. Gregory, Hunt, 与 Bates, 以及本书的作者都进行过研究.下面用一个模拟实验说明人眼的时间不变性.

图5.1是用以说明眼睛时间不变性的一个模拟实验.图5.1(a)是实验装置:下方有一根棒作为观测对象;上方是摄像机用来进行观测;在二者之间放一装着水的缸.不断摇动鱼缸使它产生水波用以模拟大气湍流的效应.我们不直接用大气做实验是因为大气湍流过于复杂,有很多不同尺度和频率的扰动,用每秒拍摄30幅图片的摄像机无法进行研究.在我们这个实验中,由于水波的干扰,摄像机拍下的每幅图片都略有差别.若将它们重叠起来,看到的将是一幅边界模糊的图像.从这里可以了解天文学家的照片上看不到那颗星的原因.人的眼睛是如何进行观测的呢?我们仍用上述实验

中拍摄的图片进行说明. 当把这些图片重叠在一起时, 虽然得到的图像边界模糊, 但它确定了图像的正确位置, 因为统计平均水面是平的. 找到统计平均的中心位置以后, 把不同时刻拍下的照片都令其中心位置与平均中心位置重合. 按这种方式重新重叠在一起的照片便给出了边界清晰的图像, 如图 5.1(b) 的中段所示. 这样一种预处理方式, 大体上模拟了人眼的时间不变性.

## 2. 空间不变性

在几何光学中, 当一个物体移近透镜时, 它的像按与物距成反比的关系增大. 当人观测物体时, 光线通过眼睛前部的透镜系统在视网膜上形成物体的像. 但是人的感受与简单的几何光学分析并不一致. 例如, 当你盯着一个人的眼睛走近他时, 会发现虽然他的眼睛看起来越来越大, 但你并不觉得他身体的轮廓有明显增大. 读者不妨伸出你的一双手立即进行一次实验: 盯着你的手心, 然后把手与你的眼睛的距离缩短一半, 你是否感到手掌的大小增大了一倍? 如果不是这样, 那就是空间不变性在起作用. 空间不变性是人眼在长期生物进化过程中形成的特点, 它在人们的生活中是很重要的. 当你为了仔细观测一个对象而把眼睛靠近它时, 只有你的注视中心部分放大了, 而其它次要部分不发生明显变化以免干扰你的注意力. 又如, 当你追赶一个猎物时, 它的边界能保持比较稳定. 反之, 若你看到猎物的边界一直在变, 就很难确认追赶的对象.

人眼的空间不变性指的就是: 当人向一个尺度较大的观察对象靠近时, 会感到它的尺度近于不变, 其变化远小于由几何关系预言的结果. 根据眼睛的结构, 一个物体被注视的中心部分成像于视网膜的中心区附近, 而物体的边界的像位于视网膜的边缘区. 人眼的空间不变性是与视网膜表面感光细胞的 PEG 分布直接相联系的. 对于这方面问题的研究已有很长的历史. M. Schultze 于 1866 年发表的论文中已描述了由解剖学方法发现 PEG 分布的研究结果. 本世纪 70 年代末以来, E. L. Schwartz, C. Weiman 与



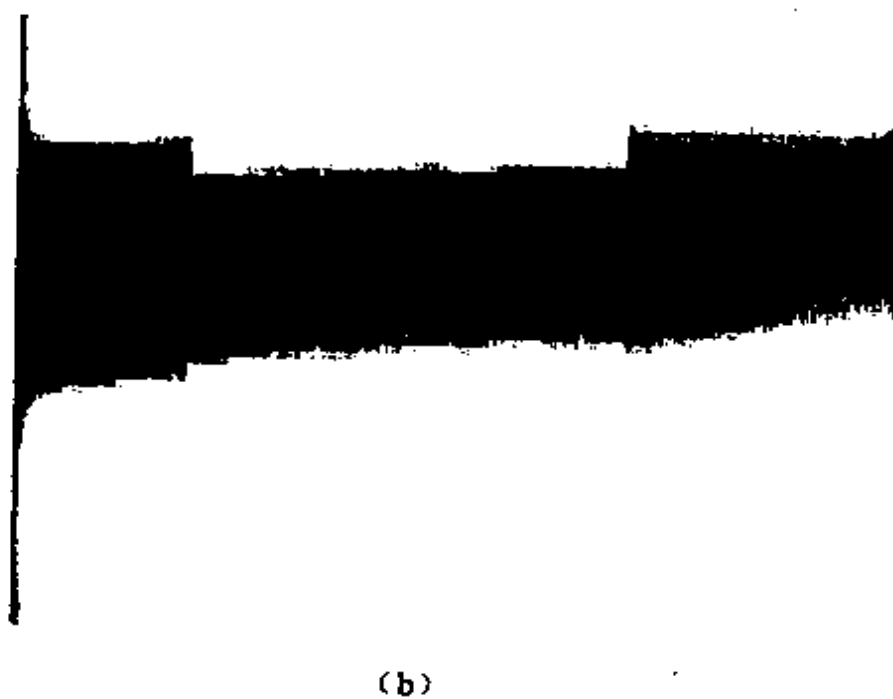
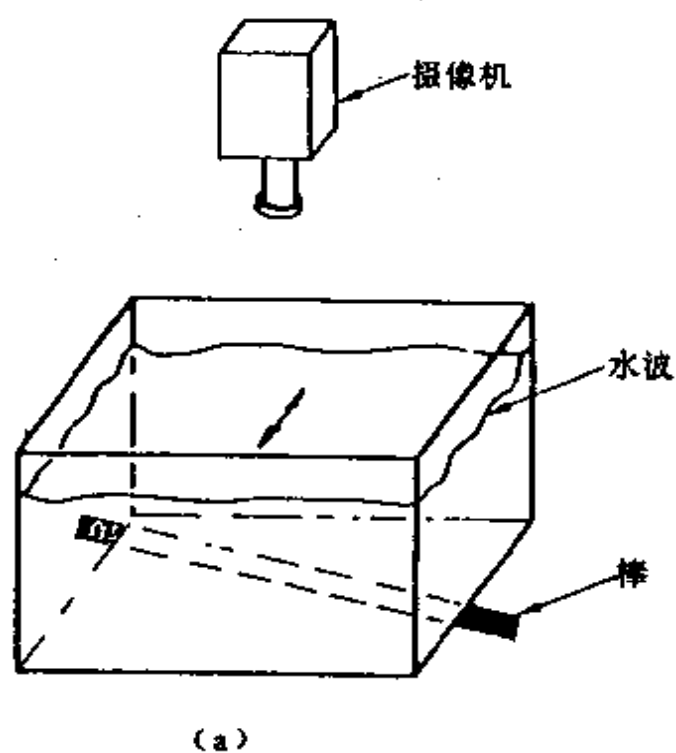


图5.1 时间不变性的模拟实验

G. Chaikin, P. Schenker 与 K. Wong 与 E. Cande, L. Pan, 以及本书作者和学生 R. Messner 等人在这方面进行了研究工作.

图5.2说明人眼的结构如何导致空间不变性. 图的最左边一层

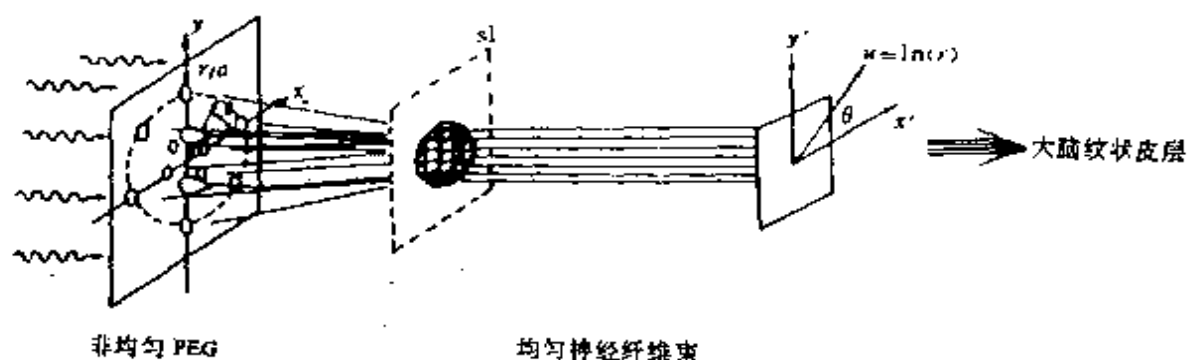


图5.2 人眼的空间不变性

是接收外界光线的视网膜, 视锥细胞的分布是不均匀的, 自中心到边缘密度按指数规律下降(即 PEG). 换句话说, 感光细胞间的距离自中心向边缘是按指数规律增大的. 图中自左向右第二层是视网膜神经节细胞, 接收来自视网膜感光细胞的信号. 细胞之间是很民主的, 每一个感光细胞都有一条神经连接到神经节细胞, 因而这是一种非均匀的输入. 视网膜神经节细胞的输出则是均匀的, 它通过均匀的视神经纤维束把信号传输到外侧膝状体, 然后, 再传到大脑的纹状皮层. 不难看出, 视网膜上一个尺度为  $r$  的像输入到神经节细胞以后, 输出图像的尺度变为

$$u = \ln r. \quad (5.1)$$

这说明人眼对视网膜上的像的尺度进行预处理的方式是进行一次对数变换. 当观测对象与眼睛的距离发生变化时, 虽然在视网膜上的像的尺度发生明显的变化. 但经对数变换后, 输入到大脑中的图像其尺度变化就很小了, 这就是眼睛的空间不变性.

在这里,我们要强调的是,在眼睛中并没有进行任何运算过程,对数变换完全靠眼睛的特殊结构实现的.同时,数目达到百万量级的坐标同时进行对数变换,是一种大规模的并行处理方式.可以看出,这是一种极有效的数据处理方式.无论人的眼睛是否真的按这种方式工作,上述借助于特殊的结构进行数据处理的思想无疑是一种很高明的设计思想.例如,我们可以设计有类似结构的摄像系统,用它来进行观测时便具有空间不变性.

### 3. 亮度不变性

当你走近一个物体时,根据辐射定律,该物体射入眼睛的光能将增多,同时由于眼睛有空间不变性,物体的大小看起来并没有明显变化,这样一来,物体的亮度就会显示出显著的增强.假如果真如此,那将是很不方便的:当你与物体的距离改变时,眼睛必须不断调整以适应亮度的变化.但是,在长期的进化过程中自然界作出了合理的选择,使我们得以免遭这类麻烦.实际上,当一个物体与眼睛的距离发生变化时,人所感受到的物体亮度变化远小于辐射定律所预言的结果.人眼的这种特性,称为亮度不变性.

亮度不变性与空间不变性有密切的关系,都是人眼的特殊结构的结果,即视网膜上感光细胞的 PEG 分布及视神经节细胞的均匀输出所致.在 T. Consweet 以及 D. Marr 关于视觉的书中可以找到有关的论述. D. Granrath, C. Hall. 与 E. Hall 及 R. Messner 与本书作者均发表过这方面的论文.这里仅以简单的一维 PEG 分布为例说明亮度不变性的起因.令物体在视网膜上形成的像的亮度分布为  $G(r)$ ,  $r$  代表视网膜上的位置(一维模型),同时,假定  $r$  处感光细胞受到亮度为  $G(r)$  的光照射时,其输出信号的强度为  $F(u)$ ,  $u$  为  $r$  处感光细胞输出信号所相应的位置,  $u$  与  $r$  的关系由 (5.1) 式给出.视网膜上接受的总光能为

$$\int G(r) dr. \quad (5.2)$$

当物体移近眼睛时,即使视网膜上像的亮度  $G(r)$  保持不变,但因像的大小明显增大,故由(5.2)式计算所得的总光能将随着物体与眼睛距离的缩短而明显增多,总输出信号则为  $F(u)$  在与(5.2)的对应区域内作积分.如果眼睛的作用仅是简单的坐标变换,则在对  $u$  积分时将出现一个雅可比行列式(Jacobian):  $e^u$ ,故与(5.2)式相应的总输出为

$$\int F(u)e^u du. \quad (5.3)$$

在(5.3)式中,输出信号的等效强度分布为  $F(u)e^u$ ,在边缘区明显增强,当物体与眼睛的距离发生变化时,(5.3)式的积分区域没有明显变化(这相应于空间不变性),但因多了一个因子  $e^u$ , (5.3)给出的总输出也像(5.2)给出的总光能一样发生显著变化.这样得不到亮度不变性,是不符合人眼的实际情况的.

亮度不变性是与感光细胞在视网膜上的指数衰减式分布(PEG)有密切关系的.在一维模型中,PEG分布是指感光细胞在视网膜上的分布密度为

$$\rho(r) \propto e^{-u} = 1/r. \quad (5.4)$$

感光细胞所吸收的总光能并不由(5.2)式表示,它可表达为

$$\int G(r)\rho(r)dr \propto \int \frac{G(r)}{r} dr. \quad (5.5)$$

与此相应,视神经总输出信号即为

$$\int F(u)du. \quad (5.6)$$

这相当于(5.3)式中不出现坐标变换的雅可比行列式  $e^u$ .由(5.6)式可以看出,输出信号的强度分布  $F(u)$  是与视网膜上像的亮度分布  $G(r)$  相对应的,而由于对数变换使(5.6)式的积分区域在物体移动时变化很小,故总光能(5.2)式的变化时总输出信号改变甚微,几乎保持不变.这就解释了亮度不变性.从物理上可以这样理解:当物体移近时,视网膜上像的边界扩大,增加的光能用于像的

大小的扩大,而边缘区感光细胞的分布稀疏,因而被感光细胞吸收的总光能并没有明显增加. 图5. 3用几何图形说明积分对应关系

$$\int G(r)dr \rightarrow \int F(u)du,$$

可保持亮度不变. 图5. 3(a)是  $G(r)$  曲线及相应的  $F(u)$  曲线, 曲线下的面积即代表相应的积分值. 图5. 3(b)相当于视网膜上的像比图5. 3(a)扩大一倍的情形, 此时  $G(r)$  曲线下的面积也增大一倍, 但相应的  $F(u)$  曲线下的面积并没有明显增大.

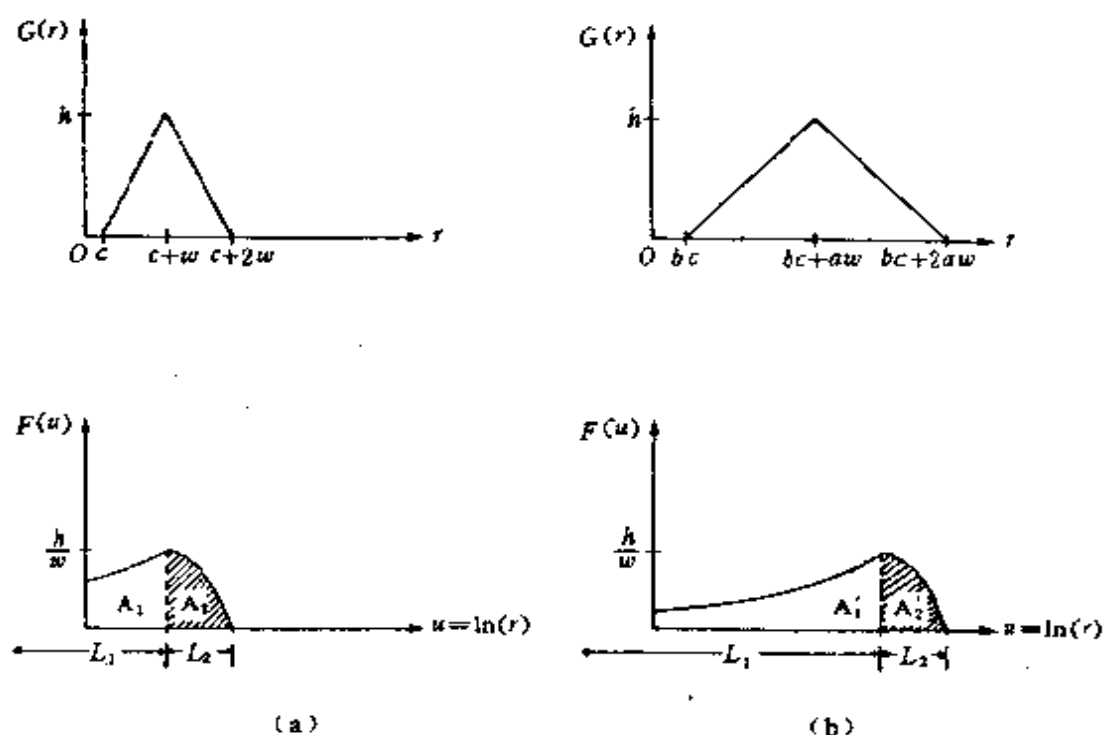


图5. 3 亮度不变性示意图

对于以上所讨论的 PEG 预处理方式,曾经有人提出批评,怀疑这种对数变换对中心的选择太敏感,不能得到比较稳定的预处理结果. 我的回答是:不会出现这样的毛病,事实上对中心的选择并不敏感. 下面用图5. 4和图5. 5说明这个问题.

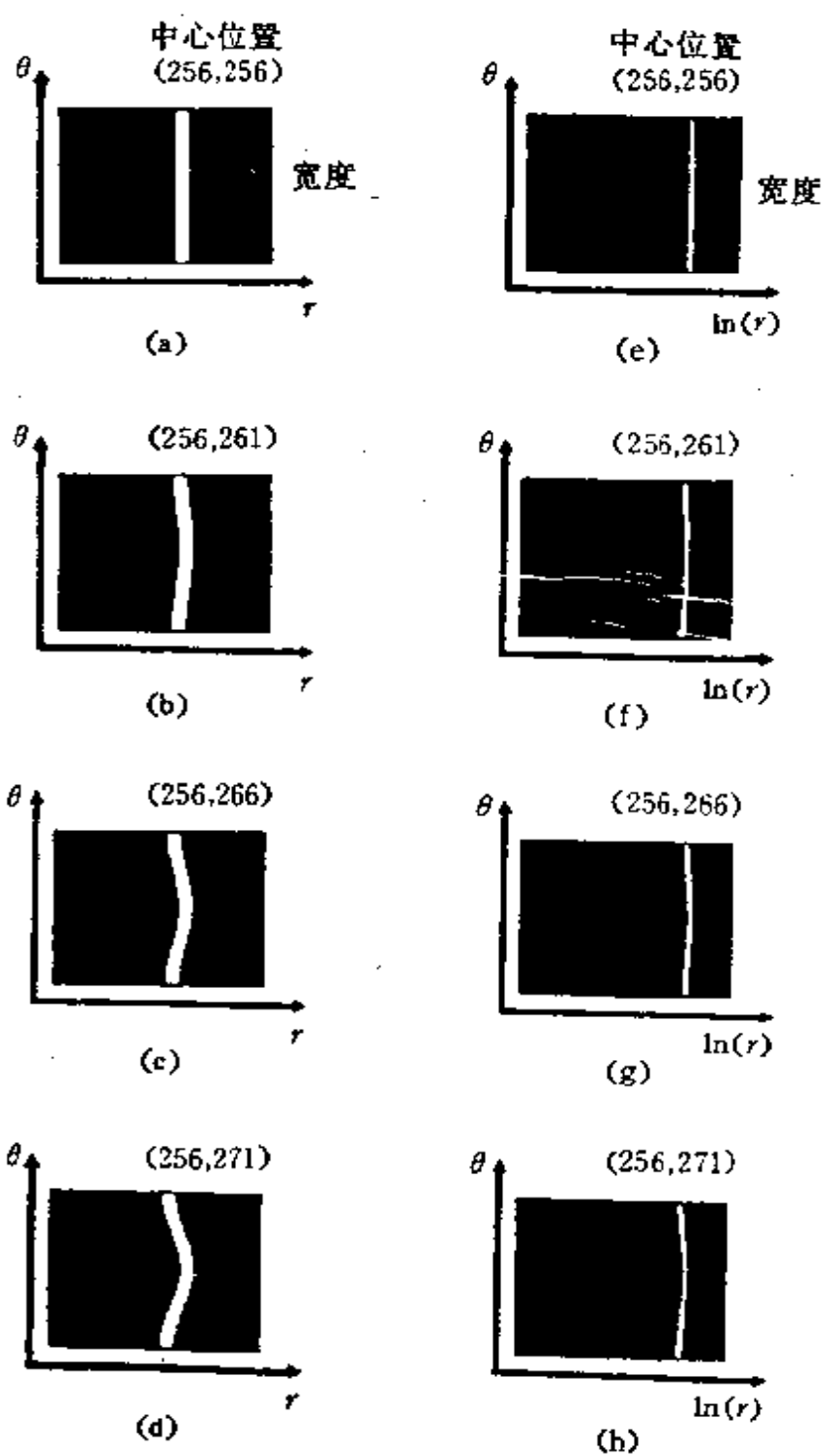


图5.4 中心移动所引起的图形畸变

频率是描写光信号随着视角变化快慢的物理量. 例如, 对于光信号作周期性变化的情形, 空间频率可定义为单位角度的周期数. 一个比较复杂的图像通常包含多个空间频率的分量, 将它作空间角度的傅里叶分析便可确定不同空间频率相应的强度, 这就是空间频率谱. 对光信号的分布除用空间频率外, 还要用对比度来描写. 对比度是描写光的明暗变化程度的. 对于亮度呈周期性变化的图像, 对比度可定义为最强与最弱的亮度之比. 人眼对于图像的分辨能力与图像的空间频率及对比度有密切关系. 一般说来, 对比度越高越容易分辨. 人眼对空间频率的响应能力则复杂一些, 图5.6是对人眼进行实验测量所得的结果.

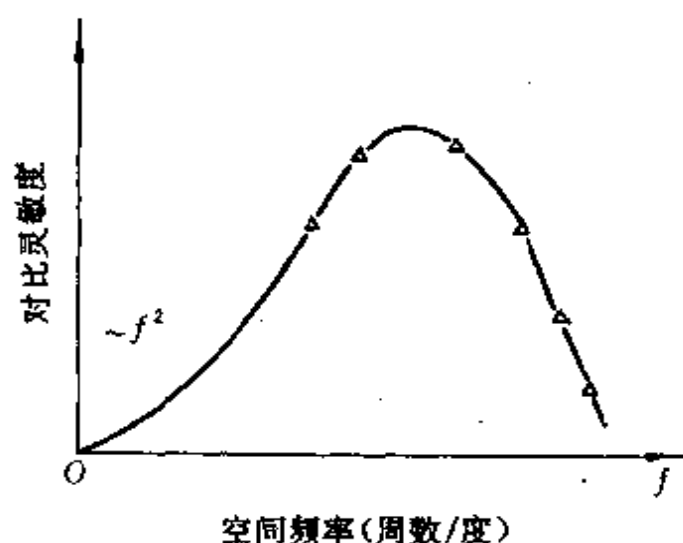


图5.6 人眼的空间频率响应

在图5.6中, 横坐标为空间频率  $f$ , 以视角间隔为  $1^\circ$  时所包含的周数表示; 纵坐标称为对比灵敏度 (contrast sensitivity), 定义为眼睛所能觉察的最低对比度的倒数. 图中曲线是大量实验的统计平均结果. 当空间频率很高时, 灵敏度随频率  $f$  的增大呈指数下降趋势, 这表幕人眼无法分

辨过于精细的空间结构. 在低频部分, 灵敏度随频率的降低按频率的平方形式下降, 这意味着过于平缓的空间变化人眼也难以辨别出来. 人的眼睛只对一定范围内的空间频率比较敏感, 这相应于曲线峰值附近的区域.

在人的视觉系统的初级部分中至少包含7个不同频带的空间

频率通道,每个通道相当于一种 Wiener 滤波器,只允许空间频率在一定范围之内的信号通过.为了说明人眼的多通道结构在视觉中的作用,下面我们考虑一个简化的复通道模型,只包含低频、中频和高频三个通道.图5.7是人眼三通道模型示意图,这是一个与图5.6相应的图,横坐标与纵坐标仍分别表示空间频率和对比灵敏度.图中的虚线即为图5.6的曲线,表示人眼对空间频率响应的总灵敏度,它由下面三

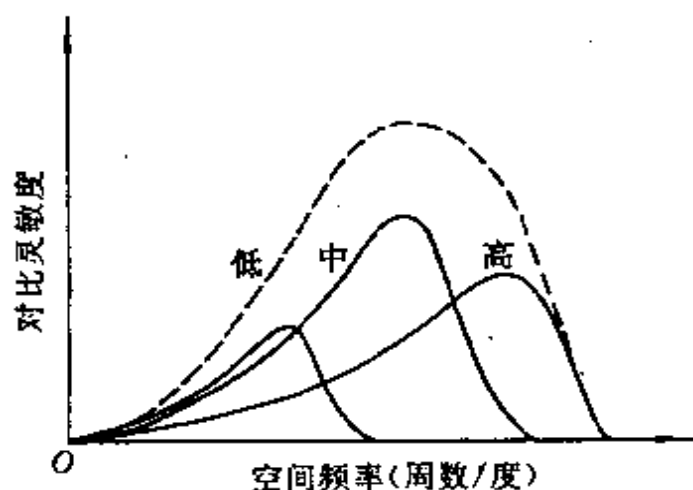


图5.7 人眼的多通道模型

条分别描写三个通道灵敏度的曲线叠加而成.低通、中通和高通灵敏度曲线的峰值分别落在总灵敏度曲线的低频、中频和高频部分.

图5.8给出了一个图像(飞机)分别经过低频、中频和高频三个带通滤波器处理以后所得到的图像.图5.8(a)是经过低通滤波器以后所得到的图像,这是一个相当模糊的形状,其作用是对被观察对象进行粗的分类,例如区分究竟是飞机还是汽车等.图5.8(c)则是经高通滤波器处理后训练的图像,在这个图中主要显示出物体的边缘,信号比较细小、微弱.高频通道的作用一般是在一个较大的类别内进行小类别的区分,例如判别是螺旋桨教练机还是喷气式飞机等.图5.8(b)则是中频通道处理的结果,其特征介于低频图和高频图之间,能显出一定的细节又比较明亮.中通滤波器的作用一般是进行个体识别,例如可看出飞机上的国旗从而判别出是哪个国家的飞机等.人的眼睛比上述模型复杂得多,包含更多的通



道,这些通道的联合作用可以完成很复杂的图像处理和识别工作.

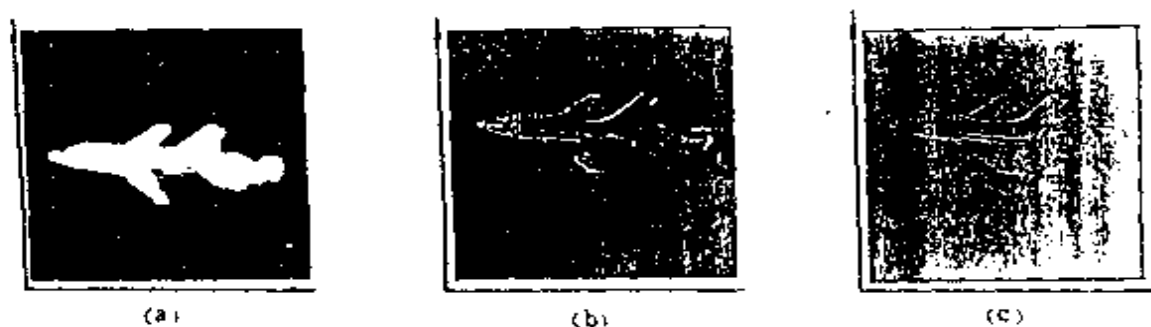


图 5.8

从数学上讲,将一个图像按不同的频带进行分解处理,这可通过傅里叶变换来实现,但即使采用快速傅里叶变换(FFT)也是很慢的.实际上,人眼并不是计算机,并没有进行傅里叶变换,那么人眼是怎样完成多通道处理的呢?后来人们发现人的眼睛是采用区域平均和信号组合方式实现多通道信息处理的.图5.9以三通道模型为例说明这个问题.图的最左方是原始图像,即各传感器所接收到的信号,将 $3 \times 3$ 个传感器进行平均得到一个低频图像,将 $6 \times 6$ 个传感器作平均得到另一个低频图像,将这两个低频图像相减,便得到中频图像,而将原始图像减去中频和低频图像便得到高频图像.

### 三、边缘探测和角度探测

边缘是描述一个图像的重要基元,边缘探测在人的视觉中有重要意义.我们都有这样的经验:物体的边缘部分看得特别清楚.又如,一个图像中白色区域与黑色区域交界处显得特别明亮.人眼的边缘探测是怎样实现的呢?

在一个图像中,强度的突变相应于一条边缘,这对应于图像的一阶导数的极值和二阶导数过零.因此,有一种观点认为边缘探测

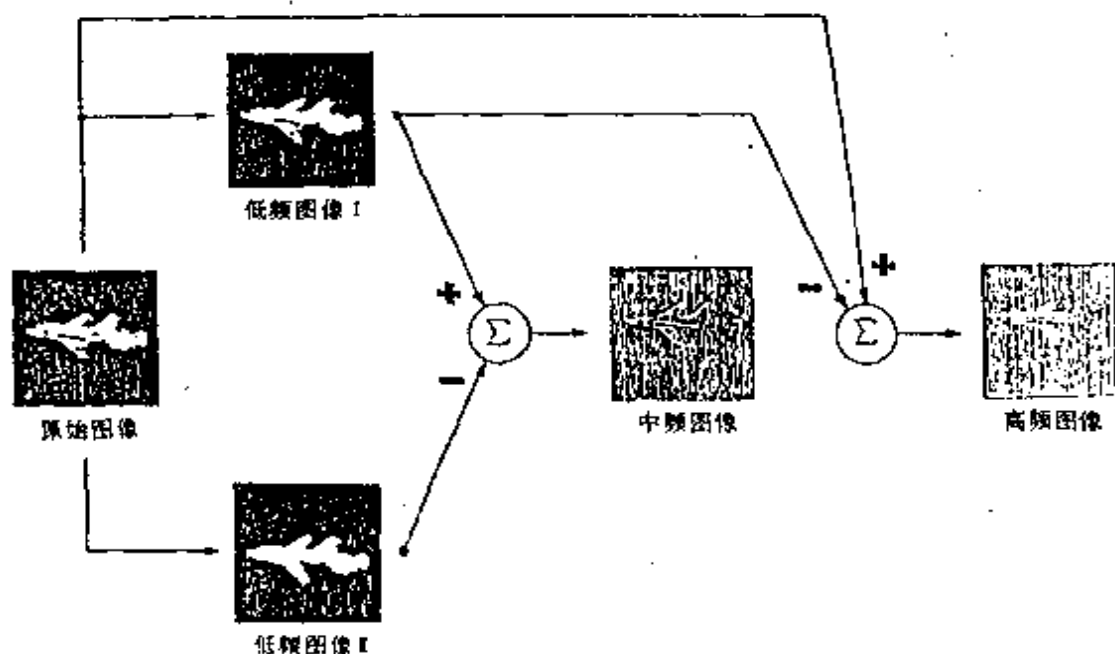


图5.9 用区域平均和信号组合的方法实现多通道信号处理

是一个计算图像二阶导数零点的过程,这可以由一个 $\nabla^2 G$ 型滤波器来实现.这里 $\nabla^2$ 代表拉普拉斯算符, $G$ 是二维高斯分布. $\nabla^2 G$ 函数形如墨西哥帽,有一个强的中心区和一个符号相反的周围区,它可以用两个宽度不同的高斯分布之差(difference of Gaussians,简称DOG)来模拟.

但是我并不同意这种观点,因为人的眼睛并没有进行微分运算.实际上,边缘探测是一个资源共享(resource sharing)的问题.当感光细胞接受到光信号时,要通过离子交换,使信号转变为电脉冲输出.我们已经指出神经细胞是很民主的,它们在获取离子时也是地位平等的.当相邻两个感光细胞接受到强度相近的光信号时,它们将近于同等地获取离子.但在边缘处,相邻感光细胞的光强度差别很大,被光照射的细胞几乎可获得全部离子.因此,在同样光

强的情况下,处于边缘的感光细胞将输出更强的信号,这就是边缘显得特别明亮的原因.这种资源共享的原则为我们提供了一种很好的设计思想,可利用来设计红外探测器等.

最后,介绍 Fukushima 关于图像识别的一项工作,其基本想法是依据有关人眼的知识提出来的.在人眼的600万个视锥细胞中,有一部分是用来区分三种颜色的,还有一些是对角度敏感的.某些图像的识别部分地可归结为角度探测.图5.10是 Fukushima 用以进行字母识别的多层系统的原理图.一个字母中的一笔可以看作是一个角度,可由角度探测来确定.每一层的处理过程均包含特征提取和粗粒平均两个步骤.这样的识别系统具有平移不变性和畸变不变性,并且是很稳固的(very robust).

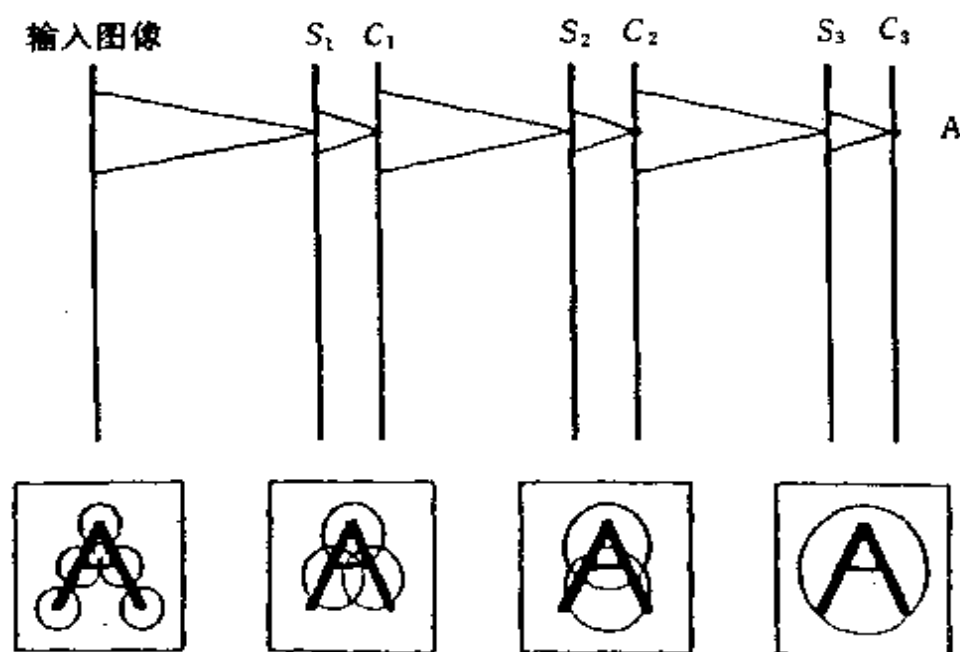


图5.10 Fukushima 的字母识别系统.

## 第六章 动态与展望

### 一、神经网络研究的动态

计算机的发展经历了几个阶段:第一代计算机是机械计算机,是很原始的;第二代计算机是真空管计算机,体积庞大,要用水来冷却,是很不方便的;第三代计算机是用半导体晶体管制成的计算机,由于包含大量焊点,焊头可能会脱落,使其可靠性受到限制;目前的计算机属于第四代,是由大规模集成电路制成的;日本人正在鼓吹第五代计算机,不少国家投入了很大的力量研制第五代计算机,但严格讲,这还不能算作新一代计算机.前几代计算机均有技术上的飞跃而称新于世,而第五代计算机只是第四代计算机再加上人工智能软件而已.至于神经网络则被称为第六代计算机,目前有电子学神经网络和光学神经网络两类,电子学神经网络中包含大量连接线,这是制造时所遇到的巨大困难.光学神经网络的单元体积很大,也限制了它的规模.到目前为止,已经制造成功的神经网络规模还不超过一百个单元,这些困难只不过是受囿于当今工业技术水平而致,神经网络的发展前途是极大的.看来,将来会出现传统计算机、人工智能系统及神经网络电脑并存的局面,因为它们各自具有不同的优点和长处.应当发挥它们各自的长处,并且互相配合起来使用,以期取得更优秀的效果.

计算机的运算速度是描写其功能的重要指标之一.对于串行式运算的计算机,由于CPU与存储器只通过一条总线连接,运算速度受到总线传输速度的限制.如果计算机的尺度为30cm,传输速度以光速计算,则运算速度为每秒 $10^9$ 次.这可以看作是超级Cray计算机的速度上限.近年来发展了一种联接机,是将多个计

计算机联在一起,以并行方式运行.这种计算机要进行 CPU 之间的通信,并且要由统一的时钟信号控制,估计其速度上限为每秒 $10^{11}$ 次.尽管这样的运算速度已经高得惊人,但仍然是不够的.例如,人工视觉系统远达不到人眼的视觉水平.人的大脑约有 $10^{10}$ 个神经元,各神经元之间非同步动作,采取协同方式.每个神经元平均与其它 $10^4$ 个神经元相连接,故大脑中有 $10^{14}$ 个突触.若以神经脉冲发放速度为每秒10个计算,大脑的运算速度就达到每秒 $10^{15}$ 次,这相当于100万台超级 Cray 计算机的总计算能力.在提高运算速度方面,最有竞争能力的还属神经网络电脑,目前计划在近期内运算速度达到每秒 $10^9$ — $10^{11}$ 次,若干年内达到 $10^{12}$ 次.为此,在理论上和技术上均必须有突破性进展.

除了运行速度外,计算机还有许多其他方面的性能和特点,例如,运行模式可分为模拟式、数字式和混合式;处理单元可分为线性、准线性和非线性;功能可分为数值运算型、符号运算型和自学习型;单元的规模可分为小型、中型和大型等等.在了解现状和确定发展方向时,这些方面都应注意到.

神经网络的研究工作经过多年的准备与探索,现在已进入决定性的阶段.日本、美国及西欧各国均制订了有关的研究规划.

日本制订了一个“人类前沿科学计划”.这项为期15—20年的计划投资规模超过1万亿日元,这还仅是“种子投资”,还希望借此吸引各方面投入更大的力量.在日本选定了二三十个小城市作为基地,包括教育、科学与工业等诸方面,准备从小开始进行人才培养,以期新一代能成为未来科学研究的中坚力量.在“人类前沿科学计划”中,神经网络和脑功能研究占有重要地位.在日本,已经投入很大力量研究人工智能系统,称为第五代计算机.他们对神经网络不采用“第六代计算机”的叫法,以免造成第五代计算机已经失败的印象.所谓“人类前沿科学”首先指的就是有关人类大脑以及借鉴大脑而研制新一代电脑.日本的此项研究计划已得到一些

国家的支持。

美国自80年代以来,经济方面面临严峻的局面.80年代初期,美国是高科技贸易出超国,而到1986年以后,变为贸易入超国,其中半数以上来自日本.美国官方和民间的主流意见不赞成采取贸易保护措施,而主张以加强发展高科技作为对策.所谓高科技,现阶段主要包括计算机、集成电路、电子设备、无线电通信及航天技术等,其中,神经网络研究是一个重要方向.美国国防部投资4亿美元,由国防部高级研究计划局(DARPA)制订了一个8年研究计划,并成立了相应的组织和指导委员会.

西欧共同体也制订了相应的研究计划,称为适应性智能的基础研究和神经计算机计划.此外,西欧一些国家还有本国的研究计划,如联邦德国从1988年开始执行叫作“神经信息论”的研究计划.

目前,在美国、日本及西欧共有数千名科学家从事神经网络的研究,这个队伍正在日益扩大.为了加强国际交流与合作,还成立了国际神经网络协会(International Neural Network Society,简称 INNS),成立后会员人数增加很快,来自许多不同的国家. INNS 的会员分布在不同的学科,体现了神经网络研究的综合性特点,其中,工程技术人员占27%,计算机与信息科学家占25%,生命科学家占20%,物理学家和数学家占19%,企业家占2%,还有7%来自其它行业.

## 二、光学-超导神经网络电脑

近年来,高温超导体的研究取得突破性的进展.这种新型超导体的出现也为神经网络的研究开辟了新的途径.在这一节中将介绍作者于1987年6月取得专利的一项方案:光学-超导神经网络电脑.

1986年1月,缪勒和柏诺兹观测到35K 的高温超导体,1987年3月,朱经武发现了95K 的超导体  $\text{YBa}_2\text{Cu}_3\text{O}_7$ ,1987年7月,白宫开始

了一项计划以推动高温超导的应用研究,高温超导的机制仍可归结为库珀电子对,从经典物理学的观点看,电子对所占据的相空间比单电子小得多,因而电子对的熵值很低;从量子力学的观点看,电子对为玻色子,不受泡利原理的限制而凝聚在基态,使其流动时没有电阻,低温超导体电子对的形成机制早已查明是电声子作用,由于声子的最高频率为德拜频率,这使得超导临界温度限制在40K 以下.高温超导体临界温度可达120K,至少在90K 是稳定的.因此,高温超导体中形成电子对的机制不能简单归结为电声子作用,其机制尚不完全清楚.

高温超导体的重要特点是它具有多晶粉末结构,并且是各向异性的,超导电性只出现在特定方向上,这使高温超导体具有新的特性.例如低温超导体可出现磁浮现象,而高温超导体可出现磁悬现象,即磁体可将其下方的超导体悬挂起来,这是由于磁场可经由多晶体超导区的边界穿过超导体,如橡皮筋一样把超导体吊起来.

当外加磁场或通过超导体的电流超过一定强度时,将使超导性被破坏.高温超导体所能忍受的磁场强度和电流密度目前还比较低,这限制了它的应用范围.高温超导体的应用包括三种类型.一是大尺度超导体的应用,如制造超导体磁体用于悬浮列车或造加速器,以及用于功率传输和通信等.这类应用一般需要较强的磁场和电流强度,目前差距较大.再一种是小尺度超导体的应用,如用于辐射传感器(探测器)或制造玩具等.这类应用只要求很低的磁场强度和电流强度,目前高温超导体的水平已达到这类应用的要求,还有一种则是中等尺度超导体的应用,用于计算机及图像处理.以往的工作主要考虑用约瑟夫森结作为二极管制造计算机,这是无增益元件,用于计算机的前景不很大.我们将考虑利用高温超导体制造三极管,并利用它对光比较敏感的特点用光进行控制.制造这类器件对超导材料的要求在不远的将来便能达到.

我们曾经指出,制造光学神经网络时,神经元体积很大,这是

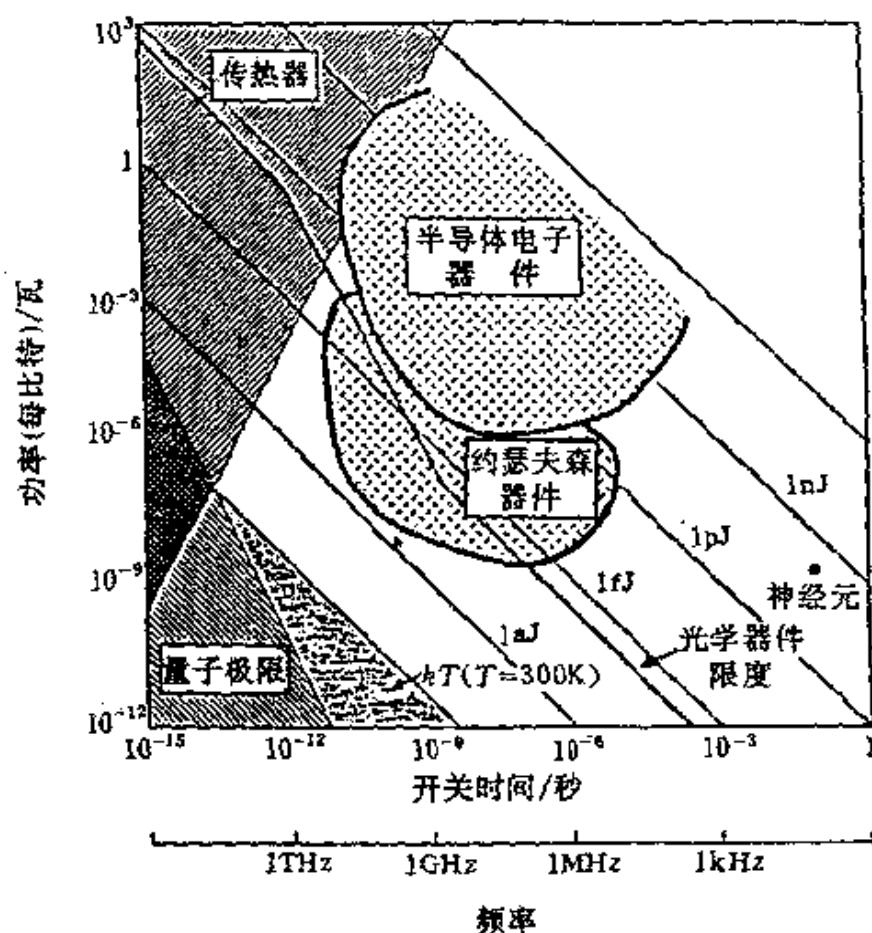


图6.1 各种单元的操作频率和耗散功率

成棋盘状二维阵列,每一个交叉点是一个超导三极管,它们是连续的模拟式开关,用作处理单元.在图中,输入电流沿焦面横向流动,输出电流沿焦面纵向流动.钇钡铜氧超导的多晶结构是各向异性的,在制作中,令具有超导性的平面与棋盘面重合,当来自上方的红外辐射沿棋盘垂直轴方向照射到超导线交叉点上,将改变作为阴极和阳极的超导电极的超导材料的比例,从而影响输出电流,输入与输出电流之比取决于所吸收的光子数,故能反映辐射强度,辐



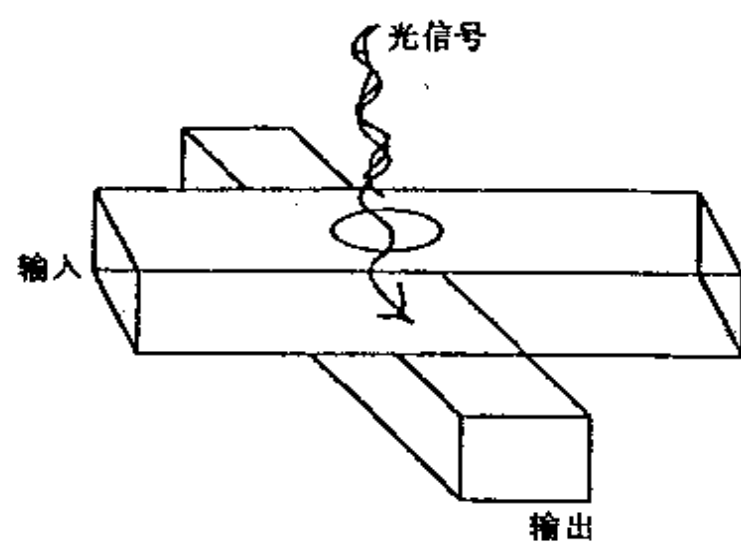


图6.2 高温超导体三极管

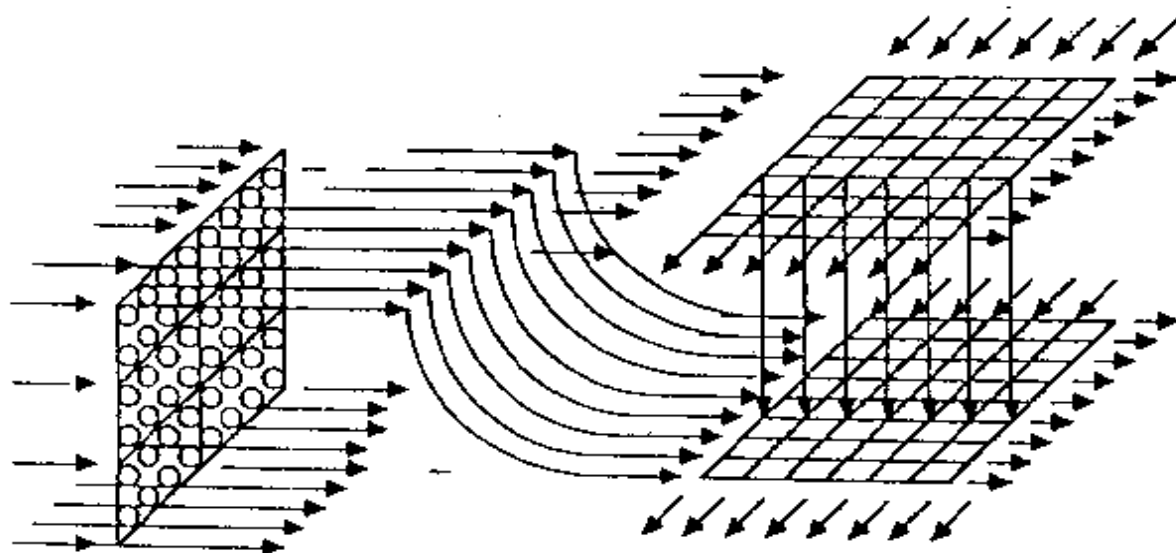


图6.3 光学-超导神经网络示意图

射的作用相当于使电流的方向转 $90^\circ$ 角,超导阵列的上方是控制辐射的强度和频率的调制系统,是由一块晶体和一片薄膜镜头组成的,可由外界控制,当改变外加电流时,会影响晶体的介电常数,它向超导阵列发射的光辐射也随之变化,将超导阵列的输出电流作为上层光学系统的输入,形成上下层之间的迭代过程,使组成一个神经网络计算机,上下层之间是负反馈关系,使系统能自动调节.当输入为均匀时,由于超导材料有缺陷使输出不再均匀,这时可以设计一薄膜补偿材料的缺陷,这样可以正确反映外界的不均匀信号.

用超导材料制作神经网络有如下几点好处:

(1) 它基于库珀电子对的性质,兼备了光学和电子学系统的优点.

(2) 它的操作单元很小,尺度只有光波波长的量级,由于耗散功率很低,故单元的排列可达到很高的密度(间隔约 $300\mu\text{m}$ ),同时,由于开关的速度很高,约1皮秒( $10^{-12}\text{s}$ ),这使得整个神经网络的运算速率可达每秒 $10^{13}$ 次.

(3) 这是一种无其他导线的立体结构,用光进行通信,同时,不需要外加的光探测器,因这种高温超导体是粉末状多晶体,光可以从超导区的边界部分透入而被吸收.

(4) 这种系统具有硬件容错性,可以消除超导材料的缺陷造成的影响.

(5) 这种系统可采用迭代反馈的学习方法.

(6) 可通过改换上层的模板去解决不同的问题,模板可采用由计算机制作的全息图.

(7) 可不作计算机而作为电话交换机电路芯片,可以完成100万人对100万人的通信.

低温超导体的光学及电学性质已有许多实验研究工作,高温超导体发现以后,可以用这种新材料重复低温超导体所做过的实

验. 缺点是用它制作大规模集成电路还有技术上的困难, 并且高温超导体的临界磁场太低. 还需要特别强调, 我们应当基于高温超导体的新的性质去做一些过去未曾做过的实验. 高温超导体有两个重要的性质: 一个是它具有整齐排列的粉末状晶体结构, 另一个是各向异性. 前一特性已经用来直接作光的探测器, 后一特性可用来作超导体三极管. 除了以上两种研究途径以外, 还可以将新旧材料结合起来, 其优点是使成本和重量降低, 缺点是温度升高后噪声增大. 从目前看来, 液氮温度下的高温超导体是最理想的超导材料. 高温超导体的研究应进一步揭示形成电子对的耦合机制, 探索如何提高它的性能(机械性能, 稳定性, 临界磁场及临界电流等), 才能拓宽它的应用范围.

高温超导辐射探测器与传统探测器的不同之处在于: 传统探测器的探测机制是光产生电子-空穴对. 通常是数字式(digital)的, 其尺度一般为毫米、亚毫米量级, 排列密度为 $120 \times 60$ 或 $90 \times 90$ , 受热噪声限制较大, 灵敏度由电子迁移率决定; 高温超导体辐射探测器基于辐射与库珀电子对的作用, 若利用电子对被辐射破坏的效应, 则与传统探测器相似. 我们这里用光子与电子对的间接作用, 当电子对穿过超导区边界时, 电子对拆散是一种统计效应, 这是一种共振吸收机制, 辐射使电子对数目连续地减少, 因而可做成模拟式(analog)开关. 高温超导体探测器可做到微米量级, 密度可达到 $10^4 \times 10^4$ . 高温超导体辐射探测器没有单个电子的离散效应引起的噪声(shot 散粒噪声), 这是由于散粒噪声是电子的空间局域性效应, 电子对在动量空间占据范围很窄, 故坐标空间定域很宽, 不出现散粒噪声, 高温超导辐射探测器的灵敏度是由量子效应决定的.

总的来说, 高温超导辐射探测器的优点可归纳如下:

- (1) 可做成小尺度探测单元, 以高密度排列;
- (2) 有极高的灵敏度, 可探测到单个光子;

- (3) 不需要 A/D 转换,探测与信息加工均为模拟式的;
- (4) 采取协同方式(concurrent)工作,不需要统一标准时钟;
- (5) 可进行相干波前探测.可由阵列中各点响应时间的先后判断入射平面波的入射角度,而传统探测器没有这一特性.利用这个性质可消除空气中湍流的影响.

## 第七章 第六代计算机的体系结构： 神经网络、超级计算机和光学数据库机<sup>①</sup>

### 一、引言

第六代计算机是一种灵巧的、具有高效输入输出的、并至少比前代计算机快几个数量级的计算机。它可能是一种所谓“三合一”系统，正如一个三极管，它由三部分技术组成：(1) 神经网络计算机，是智能控制栅极。(2) 光学数据库计算机，是信息源、接收器、阳极。(3) 超级计算机，是数字捣弄机(数字快速处理)、阴极。

历史上第一代计算机是由每秒操作一次(即 1 ops)的继电器组成的，见图 7. 1。真空管技术的发明导致第二代计算机，如 Deuce Valve 400 和 ENIAC(电子数字积分器与计算机)，ENIAC 用了 1800 个真空管，耗电 140 千瓦(足以供 40 个家庭之用)，装在一个  $10 \times 17 \text{m}^3$  的箱子内，总重 27t。它完成于 1946 年，每秒操作一千次，即  $10^3$  ops。Bardeen, Brattain 和 Shockley 于 1947 年发明了晶体管，这导致基于半导体技术的第三代计算机的诞生，以 IBM 7090( $10^4$  ops)与 CDC 1604( $10^5$  ops)为代表。1958 年，Texax Instrument 的 Jack Kilby 在解决一个由五组分组成的军事装置的固化联接问题研究时，首创了集成电路(IC)，这是第四代技术。在

---

① 本文最初发表于 3rd International Conference on Supercomputing, Boston, MA. May 15, 1988. 本文译者“Learning and Recognition: A Modern Approach”. Edited by K. H. Zhao, C. F. Zhang, Z. X. Zhu. World Scientific, Singapore, 1989. — 张承福

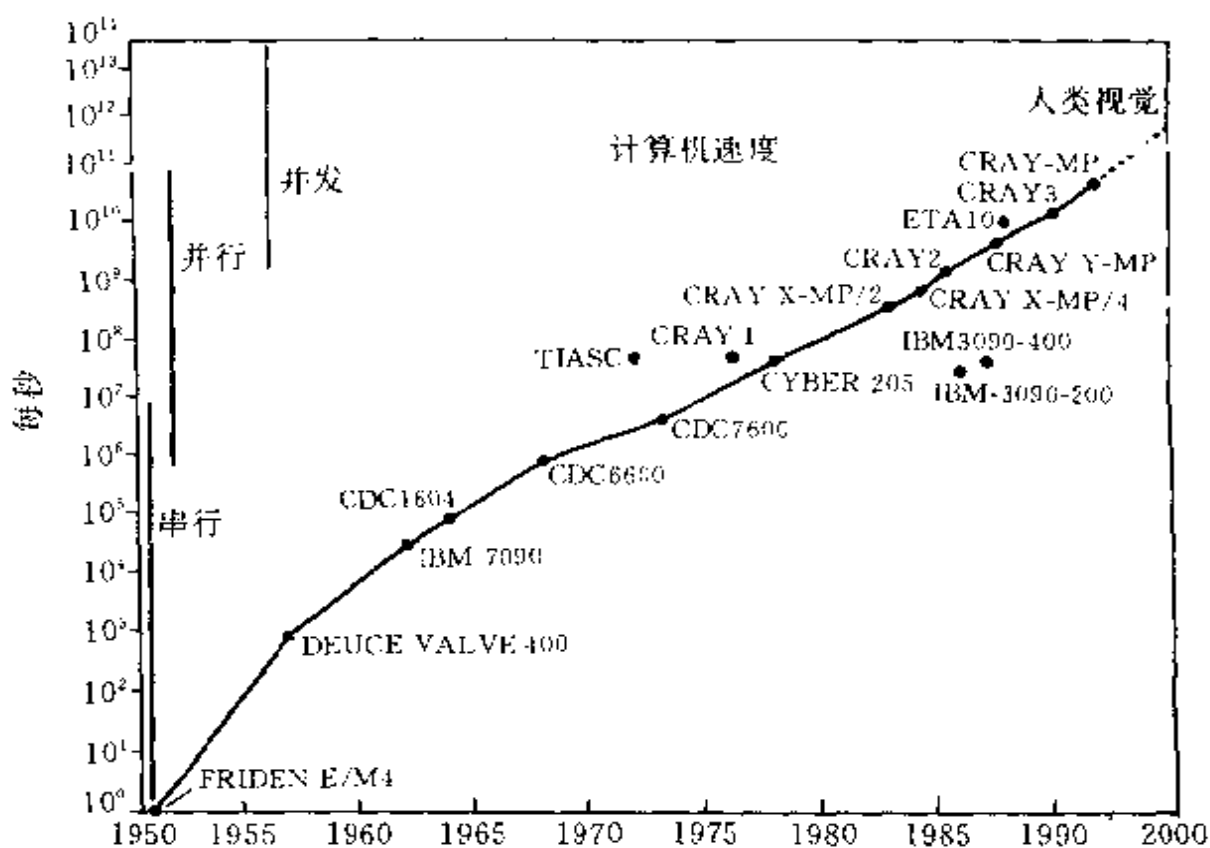


图7.1

70年代初,这一技术改进为包含 $10^4$ 个组分的微芯片(大规模集成,LSI),而于80年代进而发展为包含 $2 \times 10^6$ 组分的芯片(超大规模集成,VSLI). 大的第四代计算机以 TI-ASC, Cray 1, Cyber 205 以及 IBM 3090为代表,其操作速度均为 $10^8$  ops 量级. 第五代计算机通常认为是第四代的硬件与人工智能(AI)软件的结合体. 1986年日本国际贸易与工业部(MITI)开始了“人类前沿科学计划”(HFSP),东京新闻界将之称为第六代计算机计划. HFSP 的主要目的是企图将由神经网络、生物学、机器人和仿生材料等方面国际合作所发展出的人类智能功能体现到新的计算机中去. 预计到2000年全世界由第六代计算机所得的年收益将达600亿美元.

所有计算机基本上可按四根轴加以分类(见图7.2), (1) 操

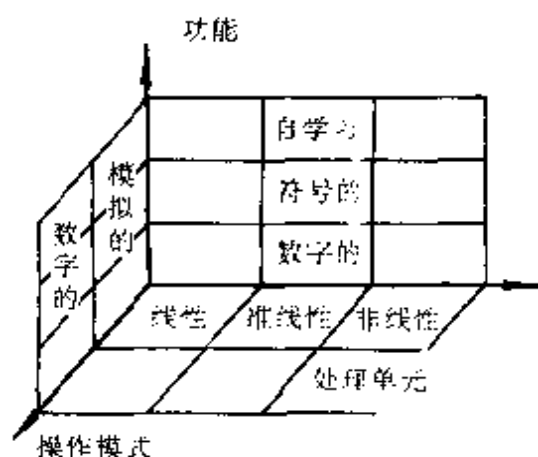


图7.2

作模式；(2) 处理单元；(3) 历史功能；(4) 粒度(未在图2中表示出来)。由这种分类,从图中空缺的元胞中也许能导致新的发展。

由发明所引起的进程通常是难以预见的。但是,不同领域科学家之间的交流,例如通过由国际交叉学科组织国际神经网络协会(INNS)主办的综述性国际杂志“神经网络”的交流,将增强发展的机会。INNS是一个非盈利的、教育性的论坛,它由来自35个以上国家的数千名成员组成,预期它将促成国际性的人类前沿科学计划(IHSFP)的实现。为实现IHSFP这一总体目标,INNS既需要得到来自超级计算公司如Fujitsu, Cray, IBM, CDC, DEC等的支持,也需要得到如(日本的)Inamori基金会和诺贝尔基金会的支持。实际上,还需要有联邦政府的财政资助以促使第三世界国家能均衡地发展。

本文描述了一种基于神经网络计算机与光学数据库计算机协同作用的合理的体系结构设计,它可增强超级计算能力,并促进用于人类福利的第六代技术的发展。

芯片制作的限制在于:(1) 芯片上线路图光刻的宽度为0.5

$\mu\text{m}$ ; (2) 芯片的发热; (3) 防止电子从一回路向另一回路漏电的绝缘层太薄; (4) 在小型化时回路的 RC 时间常数是不变量. 已采用了如下方法以克服这些限制: 用 X 射线和电子束刻图以提高空间分辨力; 用光导纤维的光子回路; 用有机分子开关的生物芯片等. 用高温超导体可使芯片在液氮或干冰条件下运行, 解决制作相容性问题. 由 Cray 机前首席科学家 Chen 领导的一个美国公司, 在 IBM 公司的支持下, 将于 1990 年制造出比 Cray 机快两个数量级的超级计算机, 这些新的超级计算机可以认为是属于第六代的, 这不仅是由于它们的速度, 也因为它们具有某些超出通常以规则为基础的人工智能的人类智能, 并且利用了光学数据库技术.

## 二、体系结构

串行式冯·诺依曼计算机的理论极限是由其通信速度决定的. 例如, 若它以光速进行通信, 即  $3 \times 10^{10} \text{cm/s}$ , 若距离为 30cm, 则量多达  $10^9$  ops. 并行机可使上限提高几个数量级, 但是为了同步化, 封锁步需要细緻的编程技术. 必须在通信与执行之间进行折中, 这将成为新的、非冯·诺依曼瓶颈. 神经网络式的并发机可以模拟人脑的异步计算, 其异步性可以自组织而无须通常的编程. 网络的优化是由有导师训练和学习完成的. 例如, 人类视觉系统可容易地实现图像放大、特征提取和模式识别. 其等价的并发处理能力可作如下估计:  $10^{10}$  神经元 (细粒处理单元 (processing elements, PE)), 每个神经元有  $10^4$  个互联键, 每秒可作约 10 次操作, 则总计可达  $10^{15}$  ops. 这一计算能力约等价于一百万台 Cray 机的同时操作, 可称为 Mega Cray (兆 Cray). (当今计算机技术中的一个有趣的课题是, 如何用慢速的开关实际高速的计算? 即除了用大规模处理单元外, 应该用什么体系结构?)

神经网络体系结构的顶层是一个星形 (见图 7.3), 星形表示处于具顶端的每一个处理器与其它处理器间有相等的联接性. 此层起



行联接的,这相当于拓扑上的层状结构.例如,输入层作特征提取,接着是用于推广的内部知识表示层,然后通向用以分类的输出层.

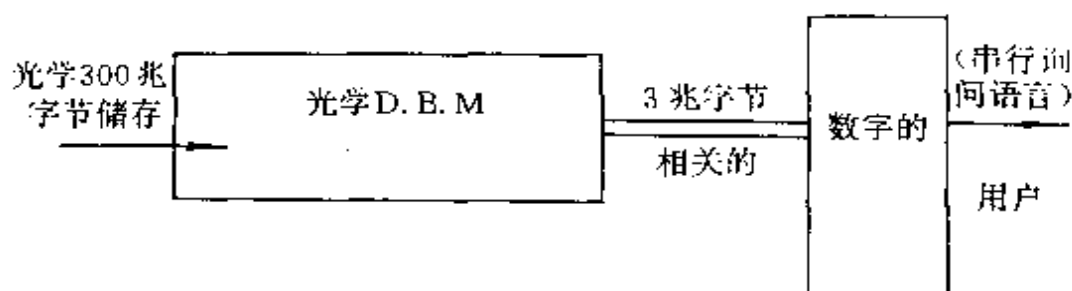


图7.4 数据库机

图7.4显示一种光-电子学的版本,状况为3兆字节帧传输率和300兆字节帧传输率;途径为多入口光盘和负全息.这里作为联想实例的输入-输出矢量对是直接输入神经网络的,无须通常的编码与解码;神经元之间的互联权重值由一个负反馈迭代系统适应性地确定.特别是,处于输入层与输出层之间的中间层,可通过所有联想对之间的联接权重提炼出一种表示.这种可变的、自组织的知识表示可获得类似于人类智能的推广与抽象.

(2) 光学计算数据库机能够实时地进行模/数及数/模转换,以与储存的程序、相当的库数据等接口.光学数据库机的基本功能是:(a) 由神经网络监督的、自上而下的模拟值的读写,适应性的、立体的或紧凑的光盘储存.(b) 某些内部的联想记忆,用于可容错的查找.(c) 数据汇合.对于图7.5所示的情形,需要有300兆字节的帧输率,这可由光学相关数据库计算达到,它可减少非相关数据并以每秒一帧(3兆字节)速率提取相关信息,这与用户终端数字串行式询问语言是相容的.此外,亦可将体全息和页全息用于储存.

(3) 此系统包括一个通用的快速数字捣弄超级计算机,它具

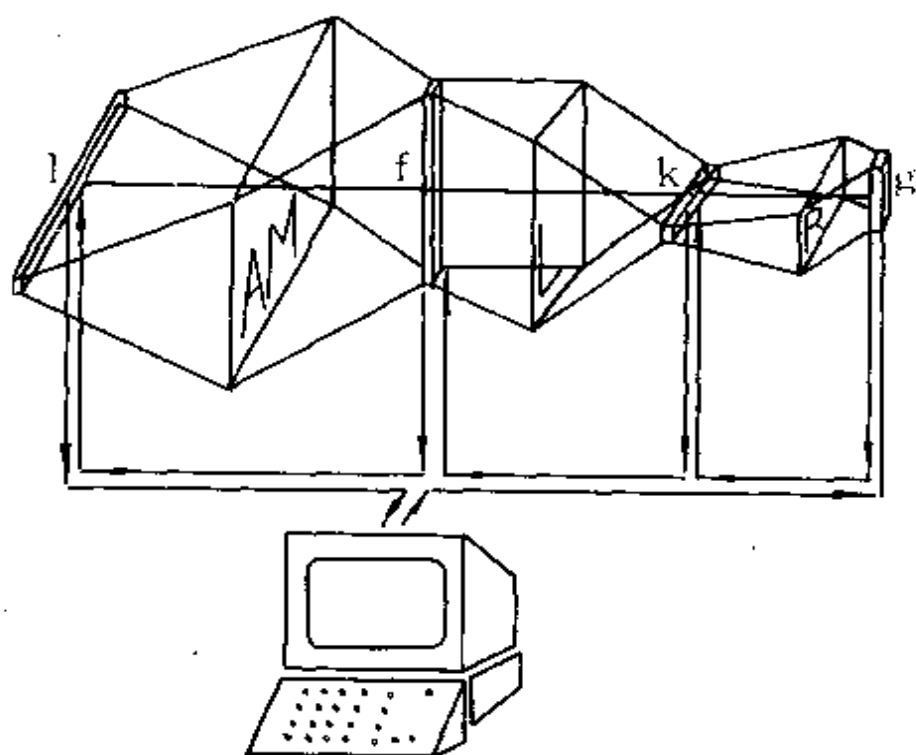


图7.5 用于计算的一种体系结构

有：(a) 一个矢量程序编辑语言(如 Gibbs 表示法)；(b) 某些精确的、内部固化的程序；(c) 若干可编程的、并行的、适应性的子程序。只有恒定的信息被数字代并准备作超级计算。操作员的命令可由声音输入；串行式询问机输入；或由分别位于(1—3)三部分的终端工作站输入，但是最终命令总是由神经网络进行检查并加以联接。

为使图7.1中的操作速率曲线外推到 Mega Cray，实施日本的 HFSP 计划的合适的估计时间似是20年。然而，本研究表明，若用上述基于协同处理技术的“三合一”体系结构，下一代 Kilo Cray 的发展可在远小于上述20年时间内实现，这是因为协同处理的每一部分都是成熟技术。这一基本思想依赖于下面的新技术。

### 三、神经网络计算

关于从部分图像联想式回忆以及非线性信号处理,文献中已有大量成功的例子.一种普遍的方法可概括如下.我们将定义三种神经元:细粒、中粒与大粒的处理单元.细粒处理单元用小写 neuron 表示,它没有内部储存,类似于大脑海马部的神经元,用于容错的联想式回忆.中粒处理单元用 Neuron 表示,它有内部储存,类似于生物上的感受器及运动控制神经元,它对于趋近的危险会作出反应.大粒处理单元用全大写的 NEURON 表示,它有内部储存,控制逻辑以及通信,相当于一台个人计算机.自然界的 NEURON 表现为祖母细胞或 pacer/conductor 细胞.这三种神经元以及它们的相应回路有四种相互作用:(1)激发的;(2)抑制的;(3)突发的;(4)渐变的和延迟传输的.一般而言,它们遵循所谓中间响应规律,S 形函数(双曲正切函数或逻辑函数),即对于弱信号以非线性快增函数予以放大,对于强信号则以非线性饱和函数予以抑制.神经网络的一般定义是,它是满足如下条件的系统:

1. 非线性 $\approx$ S 形函数 $\approx$ 分段线性,即某些点产生非线性,如下图所示:



2. 非局域性 $\approx$ 加权外积 $\approx$ 外积(自谱,即等权外积),如下图所示:

$$\left[ \begin{pmatrix} \vdots \\ \vdots \\ \vdots \end{pmatrix} (\dots) \right]$$

3. 非恒定性 $\approx$ (时间上)分段恒定 $\approx$ 迭代算法,如下图所示:



4. 非凸性 $\approx$ 制约全局优化 $\approx$ 模拟退火,如下图所示:



5. 有待发现的其它性质?

由于分段线性、分区局域、暂时恒定以及局部凸性近似,这一理论是可以数字模拟的.三十年以前 Rosenblatt 及其合作者建立的神经网络仅基于第一条非线性以及随机执行.因此,事后看来,Minsky 和 Papert 指出了这种感知器的应用局限性并提出了有用的、基于规则的另一人工智能系统(AI),是毫不奇怪的. AI 工作于封闭系统,规则统治一切;而神经网络(NN)工作于开放系统,那里规则是有待发现的.在神经网络中,这些努力的各种开拓见图7.6.

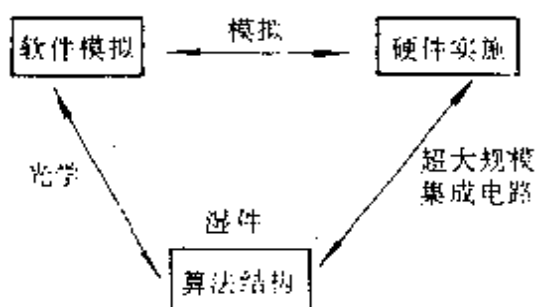


图 7.6

这是湿件(wet-fare).这一术语是由 Carver Mead 提出的,它既非软件亦非硬件,而更像是 Hecht-Nielson 提出的基于非可编程但可训练的神经网络的“网件”(net-ware).一类特定的层状网络,Rumelhart,Sejnovski 联接主义网络已表明能够作语音插入,如 Net-Talk, Boltzmann 机以及误差反传播网络等.

#### 四、结论

Hopfield 将平均场格点模型的最小能量原理应用于神经网络,并由此拓宽了领域,将物理学家和工程师们包括了进来.形象化地说,Hopfield 开创了一条美妙的过道,它有一个简单的门把手,我们可以扭动此把手进入由 Grossberg 设计的广义的神经建筑圣殿,为它数学形式的美妙所惊异,但缺乏可演示的效用.依赖于在圣殿中不同的观点,人们可以握住固定的门把手,也可能批评说这似乎是空的圣殿.我们希望,不同学科间的合作将促进神经网络这一最具挑战性的科学前沿研究的发展.本文描述了一种基于神经网络计算机与光学数据库计算机协同作用的体系结构,这样一种结构可以增强超级计算能力,并将之推向第六代技术.

志谢:作者感谢 Frank Polkinghorn 和 Lee Giles 对本文的帮助.

#### 参 考 文 献

1. T. Kohonen, "Self-organization and Associative Memory", Springer-Verlag, New York 1984
2. Stephen Grossberg, "Studies of Mind and Brain", D. Reidel Publishing Co. Boston 1982
3. John Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons", Proc. Natl. Acad. Sci., Vol. 81, pp. 3088-3092, 1984

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987



4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts.* Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987



4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987

4. D. Psaltis and N. Farhat, "Optical Information Processing based on an Associative Memory Model of Neural Nets with Thresholding and Feedback", *Optics Letter*, Vol. 10, pp. 98-100, Feb 1985
5. Harold Szu, "Globally connected network models for computing using fine-grained processing elements", *Proc. International conf. on Laser '85*, pp. 92-97, 1986
6. H. Szu, and R. Messner, "Adaptive Invariant Novelty Filters", *Proc. IEEE*, Vol. 74, No. 3, pp. 518-519, Mar 1986
7. H. Szu, and R. Hartley, "Fast Simulated Annealing", *Phys. Letts*, Vol. 122, pp. 157-162, Jun 1987
8. R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", *Optics Letters*, Vol. 11, pp. 482-484, Jul 1986
9. H. Szu, and R. Hartley, "Nonconvex Optimization by Fast Simulated Annealing", *Proc. IEEE*, Vol. 75, pp. 1538-1540, Nov 1987