

联想记忆神经网络的若干问题¹⁾

张承福

(北京大学物理系 100871)

赵刚

(北京大学非线性科学中心 100871)

摘要

对联想记忆神经网络中的几个主要问题: 评判网络性能的准则、网络收敛条件、对角项的作用、各种不同演化模式的比较以及确定网络联接矩阵的各种方案的比较等问题, 进行了分析探讨, 提出了一些明确的观点。

关键词: 神经网络, 联想记忆, 容错性。

1 引言

人工神经网络具有大规模并行处理, 良好的容错性以及可学习性(自组织)等优点。它与传统的以符号推理为基础的人工智能途径恰成对比与互补, 成为实现人工智能的不可缺少的手段之一。

联想记忆神经网络是体现网络优势、具有广泛应用前景的一类网络模型。它主要利用神经网络的良好容错性, 能使不完整的、污损的、畸变的输入样本恢复成完整的原型, 适于识别、分类等用途。由于模式的缺损、畸变等形态复杂, 可能性多(指数型增长), 所以, 用传统方法是很难处理的。用神经网络实现联想记忆, 原则上可有两类方式。一是前传式网络, 即(一步)映射网络, 由输入模式 S_{in} 到输出模式 S_{out} 按下式映射:

$$S_{out} = \sigma(W \cdot S_{in} - \theta), \quad (1)$$

其中 W, θ 是网络的联接矩阵与阈值函数, 满足

$$\sigma(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases}$$

S 各分量取值 +1 或 -1。选择合适的 W 与 θ , 可以实现联想记忆。但是, 这种网络的容错性很差, 只是下述演化网络容错性的很小一部分, 因此实用价值不大。另一类是演化式(递归式)网络, 输入模式是初态, 按(1)式演化, 直至不变的终态输出。著名的 Hopfield 联想记忆网络模型^[1,2]即属此类。演化网络是一种非线性动力系统, 从动力学观点看, 贮存样本集应构成该系统的不动点吸引子, 而每个样本吸引子应有足够大的吸引域。吸引

1) 本文得到国家自然科学基金, 教委博士点及非线性重大项目资助。
本文于 1993 年 3 月 31 日收到

域的大小与形态反映了容错性的好坏。演化网络的容错性比前传网络好得多 (见图 1, 2),但其动力学行为也复杂得多。作者认为,只有对演化网络的性能有深入的了解,对下述问题有清晰的认识,才能充分发挥网络的优势。这些问题是:

1) 评判联想记忆网络性能的准则是什么?离开了标准去提出这种或那种模型,是没有意义的。贮存容量、容错性及对资源的利用效率是衡量联想记忆网络优劣的主要标准,三者不可或缺。

2) 演化网络的收敛条件是什么?即 W, θ 应满足什么条件,才能保证网络演化时的收敛性(即只有不动点吸引子,才不会进入周期解甚或无规的混沌解)?这里涉及到网络的对称性问题与对角项的作用问题。

3) 网络的演化模式问题。这里可有同步演化网络 (SN) (所有神经元按规则同步改变状态)与异步演化网络 (AN) (每次仅有一个神经元改变状态)之分。而在异步网络中,如何选择改变状态的哪个神经元?按其选择方式还可分为随机异步网络 (RAN) (次序随机选择)、顺序异步网络 (SQAN) (次序顺序选择)、和最降下降异步网络 (SDAN) (选择使“能量”下降最快的神经元)^[3]等不同模式。重要的是,这些演化模式各有何特点?从联想记忆要求来看,哪种模式较好?

4) 如何确定 W, θ ? 现有的几种方案,如 Hopfield 的外积法、Gram-Schmidt(G-S) 正交化外积法^[4]以及赝逆法^[5,6]等各有何优点与不足?这些方案是否充分发挥了网络的潜在性能?作者认为,在一定的技术条件下,能否及如何充分发挥神经网络的潜在性能,是影响神经网络研究能否从简单重复、“玩具式”演示走向实用的关键问题之一。

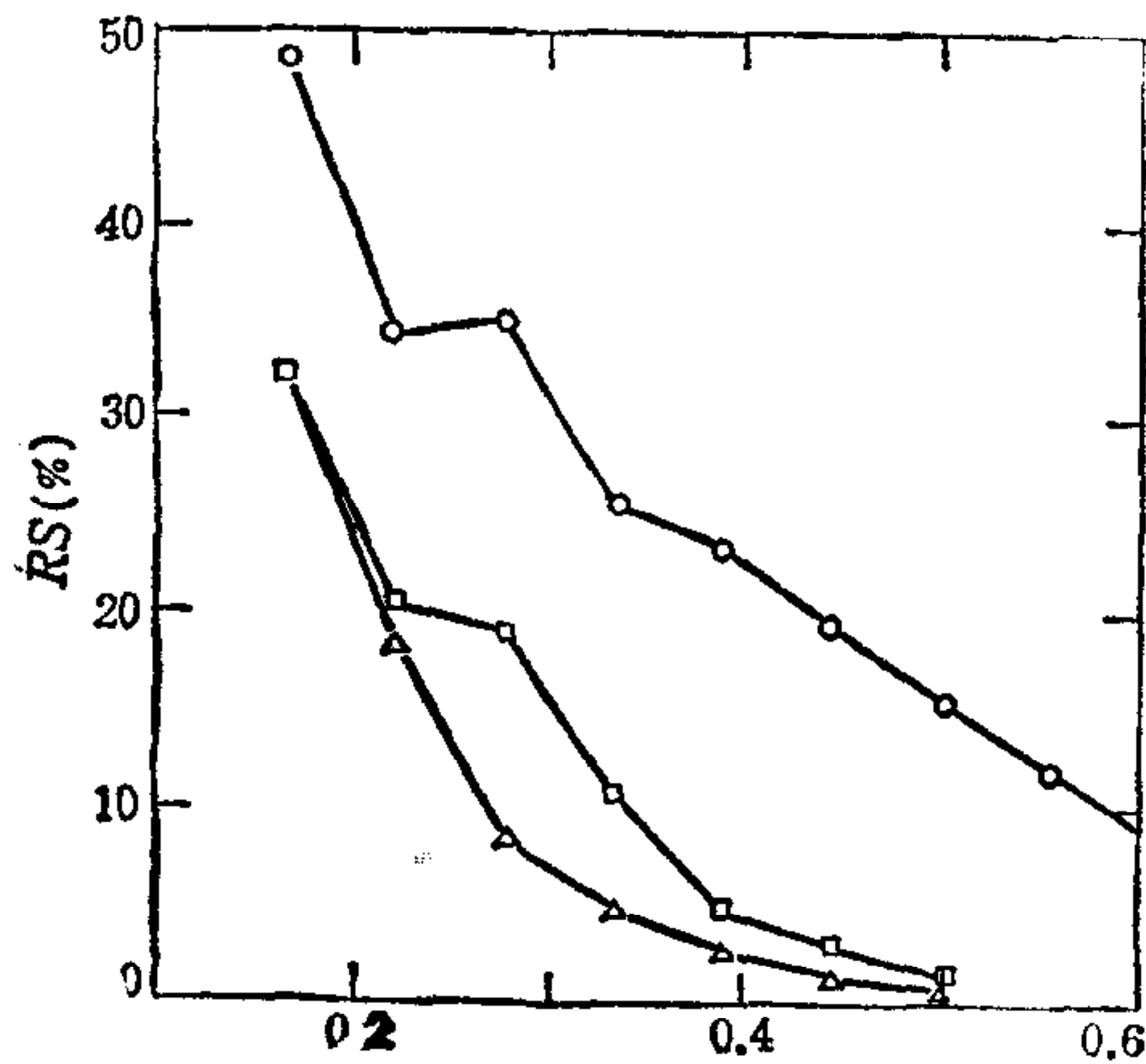


图 1 演化与映射的比较之一——总吸引域
○: 异步演化 □: 同步演化
△: 一步映射

注. 对异步网络,联接矩阵取零对角,其余则为非零对角。
该图对 $N = 18$ 的 G-S 模型作出。

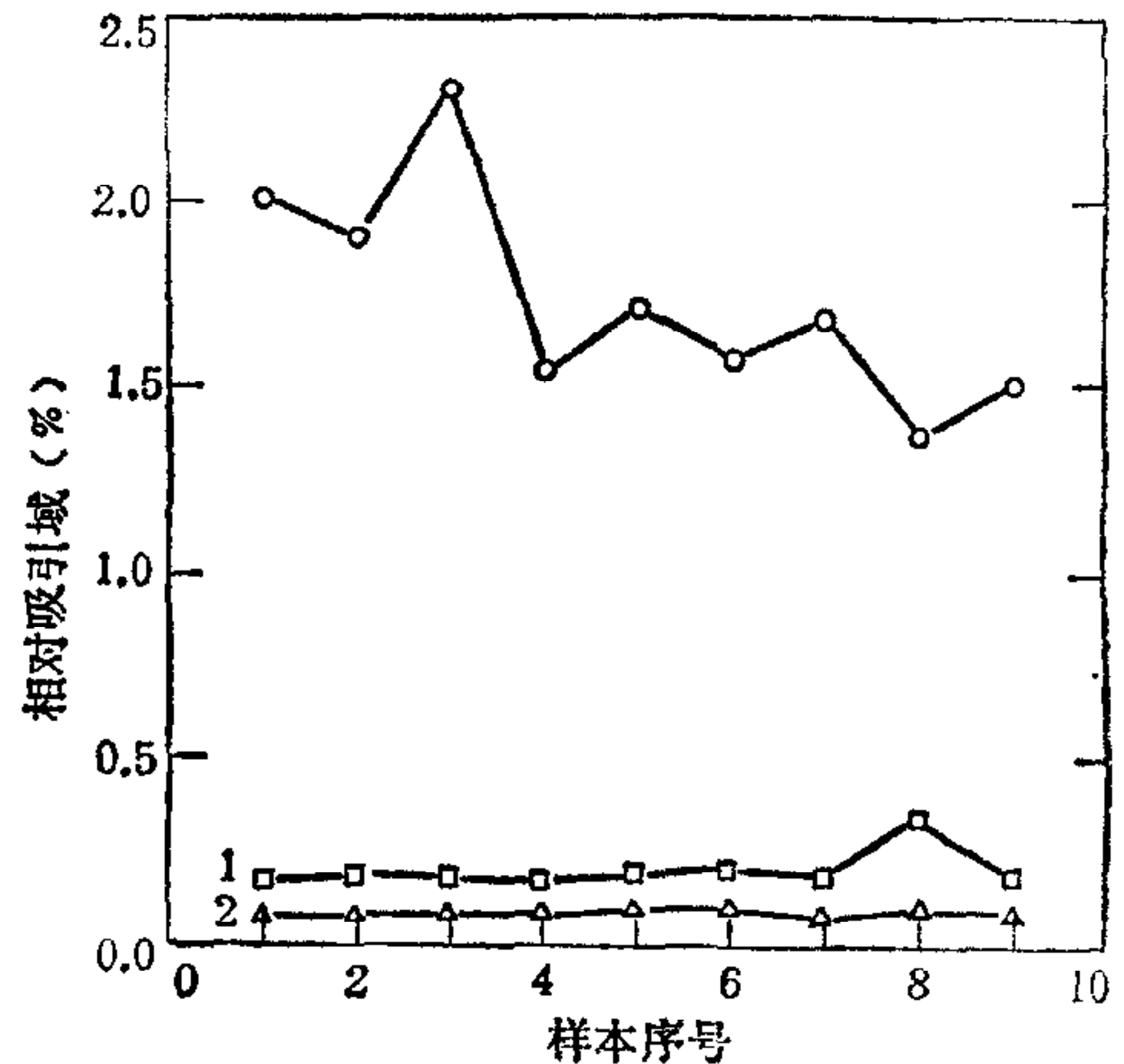


图 2 演化与映射的比较之二——各个样本吸引域

○: 异步演化 □: 同步演化
△: 一步映射

注. 样本集为 9 个随机样本,其余同图 1 注。

下面将分别讨论这些问题。对已有明确结论的问题仅作简单阐述,对尚无统一认识的问题提出作者的明确观点。

2 对联想记忆神经网络的基本要求

离开了基本要求去设计各种网络模型,讨论各种模型的优劣,是没有意义的。贮存容量与容错性是对联想记忆网络的两个不可或缺的基本要求,也是评判网络模型优劣的主要标准。此外,从实用目的及技术要求看,对网络资源的利用效率也是应该考虑的重要判据。贮存容量通常用贮存样本数 M 与神经元数目 N 的比值 $\alpha = M/N$ 来描述。提高容量的重要性是明显的,例如贮存 M 个样本所需的神经元数 $N = M/\alpha$, 联接权重数目 $\sim (M/\alpha)^2$, 若 α 增加一倍, 权重数可降为 $1/4$, 这在技术上是重要的。容量的重要性已得到充分重视, 如已有大量理论文章论证 Hopfield 模型的容量 $\alpha \simeq 0.15$, 很多工作在力求提高 α , 有的模型(如赝逆法^[5,6])声称可达 $\alpha = 1$, 等等。但是, 在追求提高容量的同时, 容错性的要求往往被忽视了。注意到按通常的编码方式, N 个神经元可贮存 $M = 2^N$ 个不同模式, 而神经网络最多只达到 $M \sim N$ (高阶关联网络除外, 可更高, 但那是以增加联接键数目, 即增加技术难度为代价的), 两者相差悬殊。所以应强调指出, 丧失容错性的神经网络是毫无意义的。容错性往往被忽视的根本原因在于不明确联想记忆网络的基本要求; 同时也有其技术上的原因: (1) 容错性如何描述? 缺乏一个明确的标准; (2) 目前还没有方法能对容错性进行解析计算; (3) 容错性的数值计算(“探测”)工作量很大, 要有一定统计意义, 决不是靠“举几个例子”就能说明的。尽管有种种困难, 为了深入了解各种网络模型的性能, 比较其优劣, 容错性的探测仍是必不可少的。

容错性由吸引域决定。要完全描述 N 维空间的吸引域形态是极困难的, 实际上也是不可能的。通常可用如下指标进行粗略刻划:

1) 样本吸引域体积, 即属于该吸引域的状态总数 $B_{i\alpha}$, $\alpha = 1, 2, \dots, M$ 。对于 N 很大的网络, $B_{i\alpha}$ 一般都很大, 但这里往往有假像。例如, 增加 n 个冗余的神经元(即它们对所有样本都取相同值, 在样本识别中没有信息量), 选择最佳联接(即它们到各神经元的联接权值为零), 可使样本吸引域增加 2^n 倍。但实际上, 网络的效率是下降的。因此, 用无量纲的比值 $M \cdot B_{i\alpha}/Q_0$ 来刻划更为确切, 其中 $Q_0 = 2^N$ 是状态空间总体积, Q_0/M 是

每个样本可能占有的最大平均体积。比值 $R_i = \sum_{\alpha=1}^M B_{i\alpha}/Q_0$ 反映所有样本吸引域的总体积, 也反映网络资源利用的效率。 $0 \leq R_i \leq 1$, 若 $R_i \ll 1$, 则表示利用率很低, 绝大部分状态空间属于无用的伪吸引子的吸引域。

2) 用 $R_c(d)-d$ 曲线可对容错性作进一步描述。 R_c 定义为 $R_c(d) = B(d)/B_0(d)$, d 是哈明 (Hamming) 距离, $B(d)$ 是距离为 d 的吸引域状态数, $B_0(d)$ 则是该层状态总数

$$B_0(d) = C_N^d = N(N-1)\cdots(N-d+1)/d!$$

$R_c(d)$ 取值 $[0, 1]$, 它有明确的物理意义: 输入状态与样本相差 d 时能正确识别的几率。如果吸引域是逐层填满的、各向同性的“球”, 则应有

$$R_c(d) = \begin{cases} 1, & d \leq d_c, \\ 0, & d > d_c, \end{cases}$$

即 $R_c(d) - d$ 呈矩形。对此的偏离反映吸引域形态的非各向同性与非均匀性。上述 R_c 或 $R_c(d)$ 可用全搜索方法(计算所有状态的演化及归宿)算出,信息全,但工作量很大。对 $14 \leq N \leq 18$ 的小网络用全搜索计算,此时 $Q_0 = 16384 - 262144$, 所得结果具有统计意义。对于更大的网络,全搜索已不现实,可采用随机抽样法进行统计(如对每个 d , 取 500—1000 个状态抽样统计算出 $R_c(d)$)。

对于联想记忆网络,单纯提“增加容量”是远远不够的,应该用 α, R_c 及 $R_c(d)$ 等多种指标才能较好地评判各种模型的性能与优劣。

3 网络的收敛条件

现有理论表明,网络演化时收敛的充分条件为:(1)对称性,(2)关于对角项的限制。

无论同步网络^[7]或异步网络^[4],都要求网络的联接矩阵对称, $W_{ij} = W_{ji}$ 。原因在于对称条件下能够找到“能量”函数,证明它在演化中单调下降,因而收敛。应该指出,对称性是一个很强的条件,它代表 $N(N-1)/2$ 个条件。它只是充分条件之一,而并非必要条件。一个明显的反例是:若 W 对称,网络收敛;现构造另一网络 $W' = L \cdot W$, 其中 L 是任意正对角矩阵(即 $L_{ii} > 0; L_{ij} = 0$, 对 $i \neq j$)。 W' 一般不对称,但由于神经元的非线性性质,即 $\sigma(x) = \sigma(ax)$, $a > 0$, 易见 W' 网络与 W 网络具有完全相同的动力学行为,当然也是收敛的。因此,可有无穷多个对称性各不相同的网络 W' 与 W 网络完全等价,均收敛。当然,此情况可称 W' 是“本质上对称”。实际上,即使是“本质上不对称”(即无法通过 L 变换使之对称化)的网络,只要偏离不很大,网络实际上也是收敛的。此问题的进一步讨论见文[3]。这里只强调一点,对称性与收敛性有重要关系,但仅依表观的对称性来区分网络收敛与否,或认为对称条件不可动摇,则是不妥的。

网络收敛性还与矩阵对角项 W_{ii} 有重要关系。同步网络与异步网络对 W_{ii} 的要求不同,正是这种不同,导致两种网络在容错性上的巨大差别。同步网络收敛条件要求 W 正定,即 W 所有本征值 $\{\lambda_i\} i = 1, 2, \dots, N$, 均大于零,否则可能形成大量二周期解。从矩阵迹的不变性等分析可知,这要求所有 $W_{ii} > 0$ 而且足够大。即网络必需有足够大的正对角项。而异步网络收敛条件只要求 $W_{ii} \geq 0$ 即可,下面分析可知,最佳选择是 $W_{ii} = 0$ 。

神经网络的一个特点是“集体决策”,这也是其容错性存在的基本原因。这反映在联接矩阵中则要求各矩阵元的均衡性,即没有哪个矩阵元占压倒优势。对于非对角元素,例如 $|W_{ij}|$ 特别大,(即 $|W_{ij}| \gg |W_{ik}|$, $k \neq j$) 则 i 神经元的状态主要由 j 神经元决定之,即 i 与 j 两种神经元是强相关的,独立性差,容错性也差,这不是好的网络。至于对角元素,正的对角元素总是有利于所有状态稳定的。靠增加 W_{ii} 能够使样本成为吸引子,但同时也导致大量伪吸引子(非样本吸引子)的增加,使样本有效吸引域明显下降。一个极端的例子是,假如 W_{ii} 足够大,即 $W_{ii} \gtrsim \sum_{j \neq i} |W_{ij}|$, 则所有 2^N 状态都是吸引子,没有

容错性,这样的网络没有任何意义。计算表明,用赝逆法或 G-S 正交化外积法(两者等价,见附录)等构造的矩阵,对角元素明显大于非对角元素(见图 3),而且随 α 的增大而增

表 1 $N = 14$ G-S 网络吸引域随 M 的变化

M	$1/\langle A \rangle$	\bar{B}_s	\bar{R}_s	N_{spur}	\bar{B}_{spur}
3	2.38	1738	31.8%	11	1015
4	2.87	810	19.8%	60	219
5	3.34	603	18.4%	331	40
6	3.79	294	10.8%	1048	14
7	4.43	36	1.6%	2429	7
8	5.28	13	0.6%	3714	4
9	6.14				

大, $\alpha = 1$ 时只有正对角项。与此同时, 伪吸引子数目及其所占的空间急剧增加, 样本吸引引域很快衰减为零, 如图 1 及表 1 所示。总之, 依靠增加对角项来提高贮存容量的做法是得不偿失的, 是不符合联想记忆网络要求的。

表中各项定义为

$\langle A \rangle$: 若用 $\langle \cdot \rangle$ 表示对量的平均, 则

$$\langle A \rangle \equiv \langle \langle \text{非对角项} \rangle / \langle |\text{对角项}| \rangle \rangle;$$

\bar{B}_s : 样本平均吸引域;

\bar{R}_s : 样本吸引域占总体积 ($2^{14} = 16384$) 的百分比;

N_{spur} : 伪吸引子数目;

\bar{B}_{spur} : 伪吸引子平均吸引域。

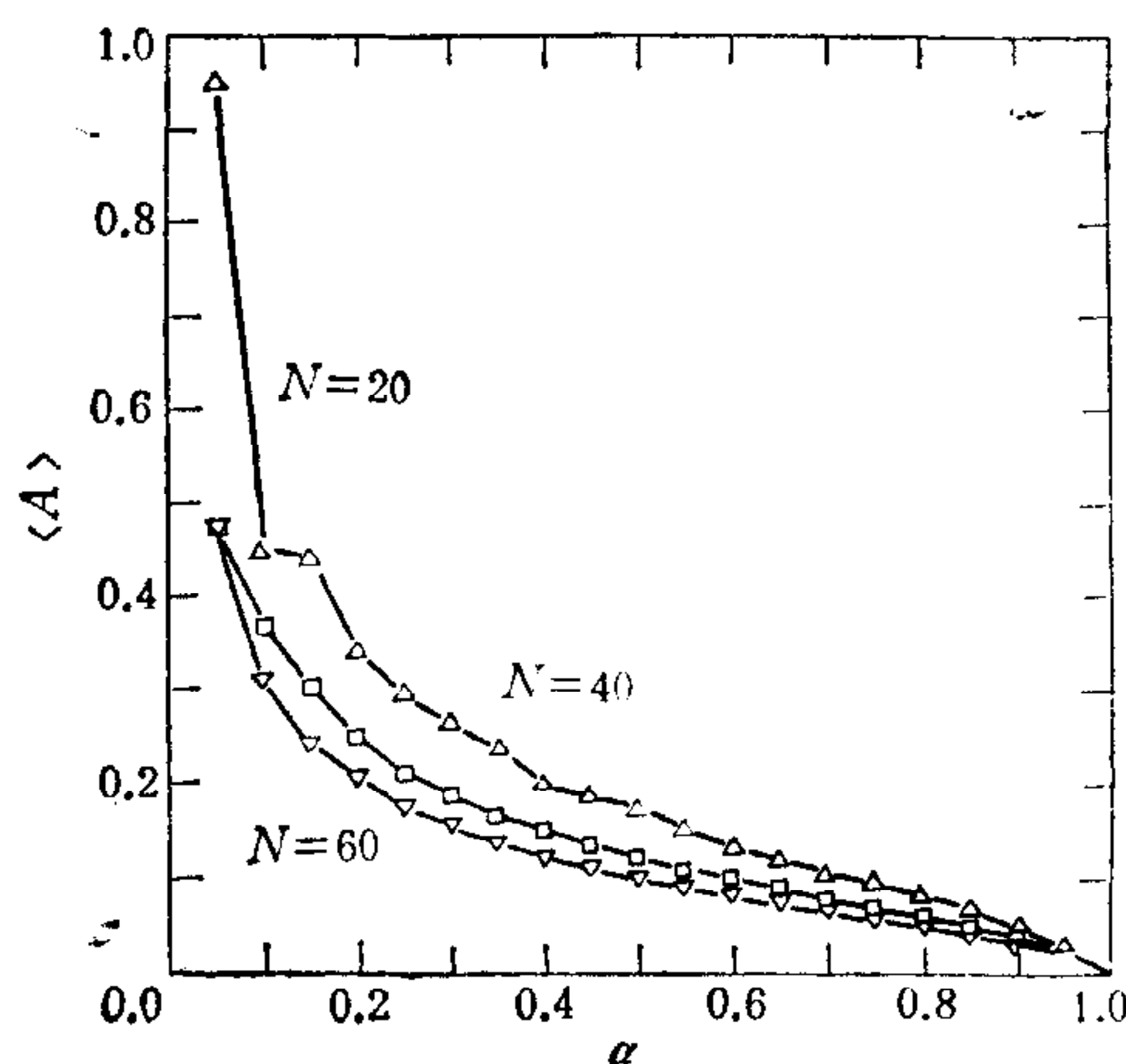


图 3 对角项随 α 的增加

注. 对三个不同大小的网络用 G-S 模式计算结果. $\langle A \rangle$ 的定义同表 1.

Δ : $N = 20$ \square : $N = 40$ ∇ : $N = 60$

4 不同演化模式的比较

无论哪种方案, 同步网络的吸引域都明显小于相应的、零对角的异步网络的吸引域, 如图 1, 2 所示。原因是清楚的: 对于任意给定的网络, 无论同步还是异步演化, 不动点 S^* 的条件都是一样的, 即

$$b_i^* \equiv S_i^* \left(\sum_{j=1}^N W_{ij} S_j^* - \theta_i \right) > 0, \quad i = 1, 2, \dots, N,$$

它们有相同的不动点集, 但其吸引域并不相同。若 $W_{ii} = 0$, 异步网络没有周期解; 而同步网络还有大量二周期解, 因此样本吸引域必然缩小。若增大 W_{ii} 以抑制二周期解, 则又会产生出大量新的不动点, 样本吸引域依然很小, 无补于事。因此, 通常取 $W_{ii} = 0$, 保证异步网络收敛, 同时使吸引域尽可能大。

异步网络又可分为随机异步网络 (RAN)、顺序异步网络 (SQAN) 及最陡下降异步网络 (SDAN) 等不同演化模式。

Hopfield 最初提出的异步网络是随机模型, 此模型具有生物学背景。问题在于, 随

机模型中每一状态的演化归宿是不确定的,与过程有关,是一种几率分布.这种不确定性对人工智能的目的往往是不方便的:同一输入模式,几次输出都各不相同(而且往往并不相近).这种不确定性有多大?一种直观的理解是:处于吸引域内部的状态,其归宿应是唯一确定的,而处于吸引域边缘区的状态,则可能因路径的不同而进入不同的吸引子.对 $N = 14$ 的网络,用等几率原则(每条可能路径几率相等)进行了统计^[3],结果表明,几乎绝大部分状态均可进入若干不同的吸引子,只有极少数很靠近吸引子的状态的归宿才是唯一的.如果只利用这一确定性的小区域,等于大大降低了网络的容错性与利用率.导致这一结果的根本原因在于神经网络的状态空间是 N 维的高维空间,空间维数越高,边缘区比重越大.对于欧氏空间,半径为 r 、厚为 Δr 的球壳与球体积之比值为

$$\frac{\Delta V}{V} = N \frac{\Delta r}{r},$$

即球体积几乎集中在表面 $\Delta r = r/N$ 一层中,当 N 很大时这是很薄的一层.对于离散空间,表达式复杂些,但道理是相似的^[3].

因此,随机模型虽然更接近大脑模型,但用于人工目的则并不一定合适.(注.大脑神经元数目极大, $N \sim 10^{11}$,具有非常大的冗余量,矛盾不明显;但对人工网络,出于技术需要,物尽其能的原则是必须考虑的).

为使异步网络成为确定论的,必须对如何选择改变状态的那个神经元作出规定,两种自然的规定是:顺序选择(SQAN)与“最陡下降”选择(SDAN).分析与计算表明,两种模型下吸引域的大小相差不明显,但SDAN具有如下优点:(1)网络收敛较快,这是由于每一步能量都有最大下降之故.(2)对于接近对称但不完全对称的网络,SDAN具有更好的收敛性,不易出现周期解,而SQAN则相对较差^[3].

总之,作者认为,在众多演化模式中,最陡下降异步网络具有相对最好的性能.

5 网络联接矩阵的确定

网络的贮存容量和容错性既与样本集在状态空间的分布有关,也与联接矩阵的确定方案有密切关系.Hopfield模型是以外积法构成 W 的,即

$$W = \sum_{\alpha=1}^M S^{\alpha} S^{\alpha} - MI,$$

其中第二项保证对角项为零, I 是单位矩阵.若样本集随机分布,即样本准均匀地分布于整个状态空间时,可证其最大容量 $\alpha_c \simeq 0.15$. $\alpha \lesssim \alpha_c$ 时它具有良好的容错性.贮存容量小是该模型难以从“玩具问题”的演示走向实用的重要障碍.考虑到任何实际问题的样本集不会是随机的,因而有很强的相关性.由于相关性,样本集在状态空间不是准均匀分布的,而是集中于某些区域,这将降低贮存容量.例如,想用该方案贮存 52 个大小写英文字母就很困难,更何况几千个汉字的识别了.也许有人会提出,可增加神经元的数目 N 以提高容量.实际上这也是行不通的.合适的神经元数目应与样本集包含的信息量有关,靠图形的细化增加的大量神经元并不能增加识别的信息量,而只是增加了大量冗余神经元.这既无助于问题的解决,反使网络的效率降低.

采用 Gram-Schmidt 正交化方案构成 W 不失为一种较好的方法^[4]。它可适用于相关样本集, 也可使容量明显高于 Hopfield 模型。但有人认为它可达 $\alpha \simeq 1$, 则是不对的。如上所述, 它所构成的 W 也是对角为主的, 当 $\alpha = 1$ 时将变成单倍阵(归一化后), 无意义。为获得较好的容错性, 应该令 $W_{ii} = 0$ (见图 1, 2 中同步演化与异步演化的比较)。计算表明, 当 $\alpha \simeq 0.5$ 时, 零对角 G-S 模型有较好的特性, α 再增加时, 或者样本不再是吸引子, 或者吸引域变得很小, 失去实用价值。

另一种较流行的方法是赝逆法^[5,6], 其基本精神是将问题化为矩阵方程的求解。将 M 个样本组成的样本集表达成 $N \times M$ 矩阵 S , 则样本集成为不动点吸引子的条件可表为

$$W \cdot S = B, \quad (2)$$

其中 W 是待求的 $N \times N$ 网络联接矩阵, B 是 $N \times M$ 矩阵, 其矩阵元 $B_{i\alpha} = S_{i\alpha} A_{i\alpha}$, $A_{i\alpha} (i = 1, \dots, N; \alpha = 1, \dots, M)$ 是任意正数。由(2)式解出的 W 可有无穷多种, 虽然都保证样本都是不动点, 但容错性, 伪吸引域等各不相同, 取决于 $A_{i\alpha}$ 的设定。如何设定 $A_{i\alpha}$ 以保证网络有最好的容错性? 这与确定 W 是同等困难的任务。如果假定 $A_{i\alpha} \equiv \text{常数}$ (不妨取为 1), 则 $B \equiv S$; 又若 S^* 是 S 的赝逆(广义逆), 即满足 $SS^*S = S$; 则可给出 $W = SS^*$, 它显然满足(2)式。需要指出的是: (1) 此 W 只保证样本集是不动点, 容错性并无保证。(2) 此 W 中的对角项是主要项, 且随 α 增大而增大(见图 3), 当 $\alpha = 1$ 时, $S^* = S^{-1}$, $W = I$, 即所有 2^N 状态都是不动点, 网络无意义。(3) 当 $M < N$ 时, S^* 仍是不唯一的。文[6]给出的表达式为

$$\begin{aligned} S^* &= (S^T S)^{-1} S^T, \\ W &= S (S^T S)^{-1} S^T. \end{aligned} \quad (3)$$

即

可以证明(见附录), 此 W 实际上与 G-S 正交化外积法是一致的, 并未给出新结果。因此本文在比较中仅列出 G-S 法的结果。

以上几种方案都是人为设定的, 它们在一定范围内是有效的, 性能也较好, 因而被广泛采用。但是从联想记忆的要求来看, 这些方案远非是优化的, 并未充分发挥网络应有的潜能。作者认为, 只有充分发挥网络的另一优势——可学习性, 才可能接近这一目的。用学习方法确定 W 并不是新想法, 已有不少训练算法。但很多方案只着眼于使样本成为不动点, 缺乏对容错性的考虑与保证, 达不到好效果。本文提出了一种学习方案(其基本思想同文[8]), 要点是在保证样本集是不动点的前提下, 具有尽可能大的吸引域。大量模拟计算表明, 此法是可行的。训练网络可以达到 $\alpha \simeq 1$ 而仍有较好的容错性, 明显优于上述各种方案。图 4 给出一例结果作为比较。关于训练算法, 作者将另文讨论。不能说已经找到了最佳方案, 但至少说明当前流行的几种方案并未充分发挥网络的潜能。如

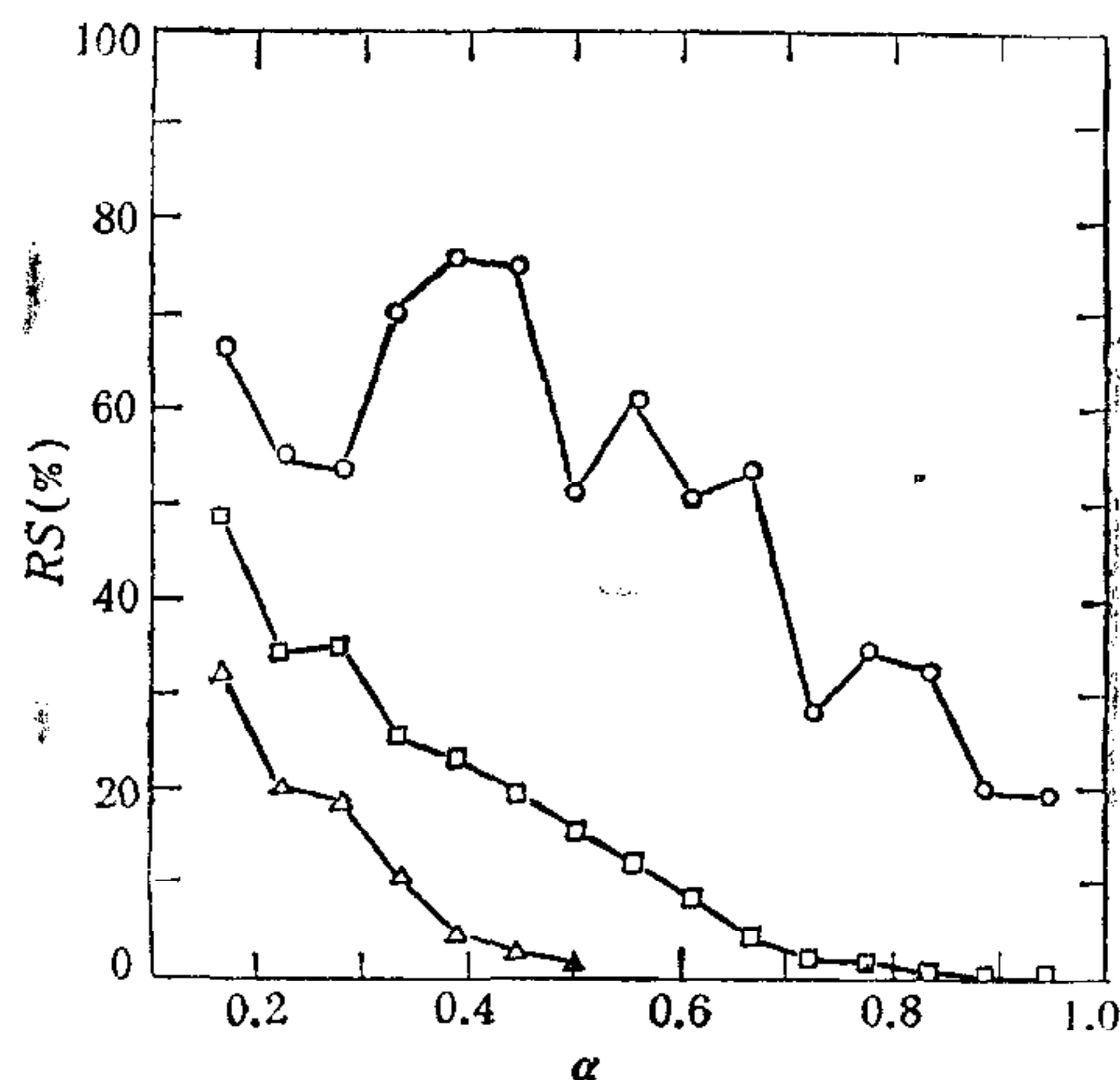


图 4 同步与异步网络的比较

- : 训练算法, 异步演化, W 零对角;
- : G-S 算法, 异步演化, W 零对角;
- △: G-S 算法, 同步演化, W 非零对角。

何提高网络效率,做到物尽其用,是联想记忆网络能否跳出玩具式演示的阶段走向实用的关键之一。为了提高网络的效率,还有一些问题,例如样本集的选取与优化,网络大小的确定等等,也是需要认真考虑的。限于篇幅,这里不再讨论。

6 结论

1) 联想记忆网络的性能取决于贮存容量、容错性及利用效率,缺一不可。

2) 靠增大网络矩阵对角元素的方法提高贮存容量的做法弊大于利,是不足取的。零对角项是最佳选择。

3) 异步网络性能明显优于同步网络。随机异步网络有很大不确定性。最陡下降异步网络是相对最好的演化模式。

4) Hopfield 外积法, G-S 正交化外积法等方案并未能充分发挥网络的潜能;用合适的训练方法可以达到 $\alpha \approx 1$ 而仍有良好的容错性。

参 考 文 献

- [1] Hopfield J J, Neural Networks and Physical Systems with Emergent Collective Computational Abilities. Proc. Natl. Acad. Sci. U.S.A., 1982, **79**: 2554,
- [2] Hopfield JJ, Neurons with Graded Response Have Colletive Computational Properties, *ibid.*, 1984, **81**: 3088.
- [3] Zhang C F, Wang X J and Zhao K H. On the Dynamical Properties of Evolutionary Neural Network, *Commun. Theor. Phys.*, 1992, **18**:233.
- [4] James T C, Applied Analytiical Math. for Phys. Sci., New York, John Wiley & Sons Inc., 1975, 19.
- [5] Personnoz L et al. J. Phys. (Paris) Lett., 1985, **46**: L359.
- [6] Kanter I and Sompolinsky H. *Phys. Rev.*, 1987, **A35**: 380.
- [7] Goles E et al. Proc. of AIP Conf. 151. New York, 1985: 165.
- [8] Kranth W et al. J. Phys. 1987, **A20**:L745.

SEVERAL POINTS ON THE NEURAL NETWORK FOR ASSOCIATIVE MEMORY

ZHANG CHENFU ZHAO GANG

(Department of Physics, Peking University 100871)

ABSTRACT

In this paper, several key problems on the neural network with associative memory are analysed and discussed. They are: the criterion to judge the performance of networks, the conditions for network convergence the role of diagonal terms of matrix, the comparison of different methods to determine the connection matrix of the network, and so on.

Key words: Neural network; dissociative memory; fault-tolerance.

附 录

本附录证明, Gram-Schmidt 正交化外积法所得的网络矩阵与文[6]给出的赝逆法矩阵实际上是一致的。

设样本矩阵 S 是 $N \times M$ 矩阵。G-S 正交化后的样本矩阵 S' 总可表成

$$S' = S' \cdot P,$$

其中 P 是 $M \times M$ 变换矩阵。 S' 的转置为

$$S'^T = P^T \cdot S^T.$$

从而按 G-S 法写出网络矩阵

$$W_{GS} = S' \cdot S'^T = S \cdot P \cdot P^T \cdot S^T.$$

正交化后 S' 满足 $S'^T \cdot S' = I$, I 是单位阵, 即

$$P^T \cdot S^T \cdot S \cdot P = I.$$

若 P 有逆 P^{-1} , 这只要求样本集线性无关, 则由上式有

$$S^T \cdot S = P^{T-1} \cdot P^{-1} = (P \cdot P^T)^{-1},$$

故有

$$W_{GS} = S \cdot (S^T \cdot S)^{-1} \cdot S^T,$$

而这正是文[6]中给出的赝逆矩阵。证毕。



张承福 1962年毕业于北京大学物理系, 长期在北京大学物理系任教, 从事理论物理课程教学及等离子体与热核聚变的理论研究。近几年来从事人工神经网络以及非线性动力系统方面的研究。



赵刚 1992年毕业于北京大学物理系, 获学士学位。后在张承福教授指导下从事人工神经网络的动力学性质研究。现在北京大学物理系攻读理论物理非线性物理专业硕士学位。