

神经元传输函数对人工神经网络 动力学特性的影响^{*}

王瑞敏¹⁾ 赵 鸿^{2)†}

1) 兰州大学物理系, 兰州 730000)

2) 厦门大学物理系, 厦门 361005)

(2006 年 4 月 29 日收到, 2006 年 6 月 3 日收到修改稿)

以神经元局域场分布为基础, 重新研究了连续神经元传输函数对具有联想记忆的人工神经网络功能的影响. 与以往的认识不同的是, 研究发现连续传输函数与硬极限传输函数相比并不存在明显的优越性. 相反, 连续传输函数对网络的某些功能, 如最大存储率具有负面影响. 研究表明神经网络的特性主要决定于网络的动力学结构(具体体现为网络吸引子对应的神经元局域场分布). 网络的动力学结构可以通过选择合适的设计规则进行有效控制, 不同的传输函数虽然也能影响到网络的动力学结构, 但是它所带来的影响是被动的, 可控性很差.

关键词: 联想记忆, 神经网络, 吸引子, 局域场分布

PACC: 0545, 8710, 8730G

1. 引 言

神经元传输函数是设计人工神经网络的最基本的要素. 一般分为两种典型类型: 硬极限传输函数和连续传输函数. 硬极限传输函数是阶跃式输出, 如 +1 或 -1, 而连续传输函数输出连续的值. 相对而言, 连续传输函数更加符合生物神经元的真实特性^[1,2], 它能够被各种电子器件实现, 同时由于其连续可微, 被广泛应用在目前应用最广的多层反向传播(BP)网络中^[2-8].

连续传输函数被神经网络研究广泛采用的另一个原因在于前人的研究表明这种传输函数还具有动力学和计算性能方面的优越性: 可以消除反馈网络的周期性运动而使网络输出趋于稳态, 用来设计可应用于并行处理和超大规模集成电路的神经网络并保证其传输稳定性, 极大地消除反馈网络的伪吸引子或亚稳态而增强网络的联想性能, 改善货郎担问题神经网络求解的能力等等^[7-12].

由于神经元传输函数在神经网络中的重要地位, 它在网络动力学和功能方面的作用值得进一步的研究. 本文以具有联想记忆功能的反馈网络为例,

提出以网络的吸引子对应的神经元局域场分布来刻画网络的动力学结构, 重新分析传输函数对网络性能的影响. 具体来讲, 网络的任意初态最终会演化到网络系统的吸引子上, 在这些吸引子中, 与记忆模式一致或相似度很高的吸引子定义为记忆模式吸引子, 其他的叫做伪吸引子. 吸引子对应的神经元局域场特征决定了吸引子的稳定性, 也决定了网络的动力学特性. 而改变吸引子对应的局域场的方式有两种, 首先, 可以通过学习规则或设计规则控制网络的权值矩阵从而控制网络吸引子对应的局域场分布, 其次, 在权值矩阵确定的情况下局域场的分布也可以通过传输函数改变. 我们的研究表明, 以连续传输函数取代硬极限传输函数而导致的网络局域场结构的改变是被动的, 不能根据特定的目的而有效控制, 它对于网络特性的影响并不必然导致明显的正面效果. 与此不同, 合适的设计规则可以实现网络动力学结构的定向控制, 从而有效地根据特定的需要最大化地优化网络的性能. 值得指出的是, 局域场分布分析方法可以给出传统的平均场分析方法所不能取代的信息, 作者尚没有见到系统使用这种方法的文献.

本文第 2 节研究连续传输函数对 Hopfield 反馈网络^[10-14]性能的影响. 这种网络由著名的 Hebb 规

^{*} 国家自然科学基金(批注号: 10475067)和教育部博士点基金资助的课题.

[†] 通讯联系人. E-mail: zhaoh@xmu.edu.cn

则设计,具有对称的权值矩阵,其记忆模式吸引子对应的局域场分布由所存储的模式本身决定,没有额外的参数控制局域场分布的形态.这种情况下,连续传输函数就成为在给定存储率的情况下调节局域场分布的唯一手段.我们的研究表明,虽然合适的连续传输函数可以消除伪吸引子的数目,增大记忆模式吸引子的吸引域,但是并没有增大记忆模式的有效吸引域,同时还降低了网络的最大存储率.文献 [15] 最近提出一种新的神经网络设计规则,称为蒙特卡罗变异-优化选择规则,这种设计规则可以很好地控制记忆模式所对应的局域场分布,也就是说在给定存储率的情况下记忆模式吸引子对应的局域场分布可以通过设计规则进一步控制.为了方便讨论,这种规则设计的网络在本文中简称为整体优化网络.本文第 3 节将表明通过对记忆模式吸引子对应的局域场分布的控制,即使在硬极限传输函数的情况下也可以实现伪吸引子的完全消除而使记忆模式的吸引域最大化,同时可以增大记忆模式吸引子的有效吸引域.在整体优化网络中引入连续传输函数后,网络的性能不仅没有明显的改善,还会造成某些方面的负面影响.

本文第 4 节分析说明决定网络功能和特性的主要因素是网络的动力学结构而不是传输函数本身,传输函数通过影响网络吸引子对应的局域场分布而影响了网络性能,从而使得研究者误解了传输函数的作用.事实上,合适的设计规则才是控制局域场分布的基本手段.

2. Hopfield 网络

具有联想记忆功能的反馈神经网络作为一类基本的人工神经网络自 20 世纪 80 年代至今,一直备受物理学家尤其是非线性动力学和统计物理领域学者的重视.一方面,它的实现简洁明了,有助于对人工神经网络基本原理的理解;另一方面,这种网络可归属于统计物理学中著名的 Ising 模型的框架下,同时又是一种典型的高维非线性动力系统.目前,这种网络已经被广泛地应用于解决各种实际问题,如信号和图像的传输和识别,方程优化,视觉研究等^[16-22].

我们研究由 N 个神经元组成,神经元权值矩阵为 \hat{J} 的反馈神经网络,其动力学演化方程由下式

$$\xi_i(t+1) = f(h_i(t)) \quad (1)$$

给出,其中

$$h_i(t) = \sum_{j=1, j \neq i}^N J_{ij} \xi_j(t) \quad (2)$$

称为 t 时刻作用在第 i 个神经元上的局域场, $\xi_i(t)$ 是第 i 个神经元在 t 时刻的状态, J_{ij} 代表第 j 个神经元对第 i 个神经元的作用, f 是神经元的传输函数.传输函数 f 可以有多种形式,本文将研究两种传输函数,一种是对称硬极限函数,或称符号函数

$$f(x) = \text{sign}(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (3)$$

另外一种是连续传输函数——双曲正切 S 型函数

$$f(x) = \tanh(\beta x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}}, \quad (4)$$

β 是传输函数 f 在坐标原点位置的最大斜率,也称为增益因子.设 $\{\xi_i^\mu, i = 1, \dots, N; \mu = 1, \dots, m\}$ 表示 m 个给定的记忆模式,这里 ξ_i^μ 代表第 μ 个给定记忆模式的第 i 个分量,它常常被限定取两个分立的值,即 $\xi_i^\mu = \pm 1$. 对于某个特定的模式来说,如果存在一些与它有一定的相似度的初始状态,它们随着网络的演化最终中止在这个模式上,我们就说该网络具有对这个模式的联想记忆功能,这个模式就被网络存储.一个网络的存储率定义为 $\alpha = m/N$. 从动力学的角度来看,一个被成功存储的模式就是系统的不动点吸引子,满足不动点条件

$$\bar{h}_i^\mu \xi_i^\mu = \sum_{j=1}^N J_{ij} \xi_j^\mu; \bar{h}_i^\mu \geq 0, \quad (5)$$

这个条件也可以写成

$$\bar{h}_i^\mu = \xi_i^\mu \sum_{j=1}^N J_{ij} \xi_j^\mu \geq 0, \quad (6)$$

Hopfield 发现如果以 Hebb 学习规则确定权值矩阵 \hat{J} , 即

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^m \xi_i^\mu \xi_j^\mu, \quad i, j = 1, \dots, N, \quad (7)$$

则网络就可以存储一定数量的记忆模式.上式中,一般设定 $J_{ii} = 0$, 这意味着神经元自身的状态对其下一时刻的状态没有贡献.

2.1. Hopfield 网络的存储率

将 (7) 式带入 (2) 式, 得到

$$h_i^\mu = \frac{N-1}{N} \xi_i^\mu + \frac{1}{N} \sum_{j=1}^N \sum_{v=1, v \neq \mu}^m \xi_i^\mu \xi_j^v \xi_j^\mu, \quad (8)$$

当 N 足够大的时候, 上式的第 1 项趋于 ξ_i^μ . 由于 ξ_i^μ , ξ_j^v 和 ξ_j^μ 之间没有关联 ($v = \mu$ 除外), 因此第 2 项可以看作是均值为 0 的高斯噪声, 并且相当于步长为

$1/N(N-1)(m-1) \approx Nm$ 步的一次随机行走,由此可以推导出局域场 h_i^m 的概率分布函数为

$$P_{\pm}(h_i^m) = \frac{1}{(2\pi\alpha)^{1/2}} \exp\left(-\frac{(h_i^m \mp 1)^2}{2\alpha}\right), \quad (9)$$

其中 P_+ 给出 ξ_i^m 取 +1 的情况, P_- 给出 ξ_i^m 取 -1 的情况, P_+ 和 P_- 关于坐标原点对称. 图 1(a)~(c) 分别给出 $\alpha = 0.03, 0.06, 0.12$ 情况下的局域场分布函数 $P_{\pm}(h)$. 可以看出, 当存储率较小时, P_+ 和 P_- 之间存在明显的间隙, 间隙的存在意味着不动点条件被满足. 间隙的宽度随着存储率的增大而减小. 当存储率增大到一定值的时候, P_+ 和 P_- 开始相交, 表明不动点条件对某些模式不再被满足. 图 1(c) 的情况下, 可以明显看到 P_+ 和 P_- 已经发生相交重叠.

实际应用中, 间隙的消失可以根据是否存在属于 P_+ (或 P_-) 的局域场 h_i^m 错误地落入 $h < 0$ (或 $h > 0$) 的区域来判断, 为此, 定义当有一个 h_i^m 落入不正确的区域时的存储率为 α_1 . 考虑到 P_+ 和 P_- 关于坐标原点对称, 并且 m 个记忆模式共有 Nm 个 h_i^m , 这一条件具体可以写成

$$\int_0^{\infty} NmP_-(x) \geq 1, \quad (10)$$

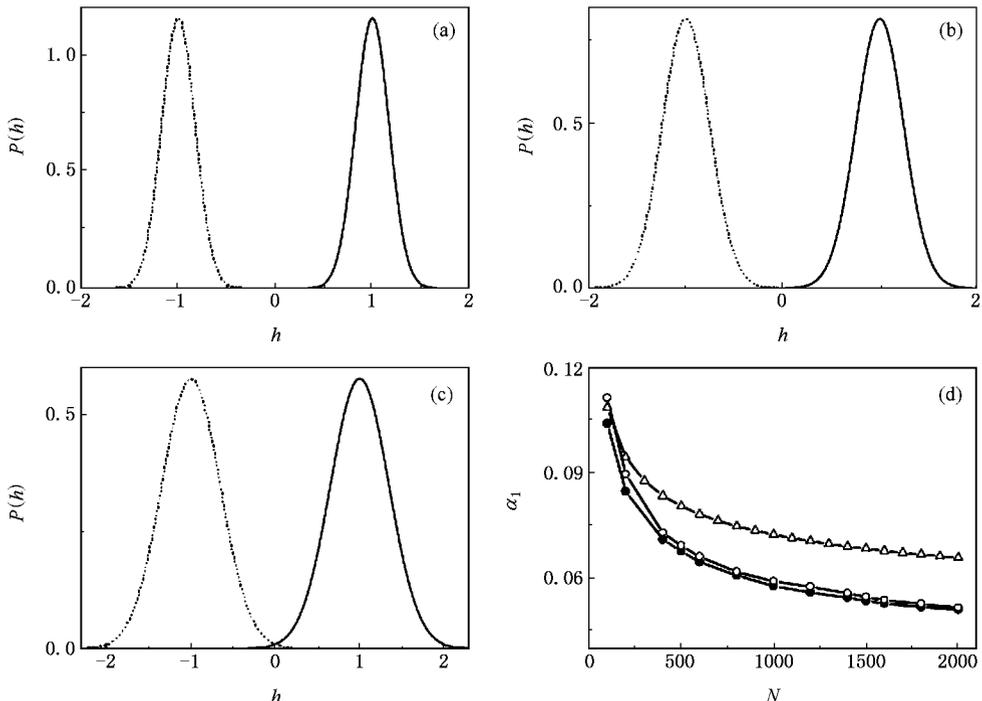


图 1 在存储率分别为(a) $\alpha = 0.03$, (b) $\alpha = 0.06$, (c) $\alpha = 0.12$ 情况下的局域场分布函数 $P_{\pm}(h)$, 虚线表示 P_- , 实线表示 P_+ , 以及(d) α_1 随 N 变化的函数曲线, 实心圆圈表示本文的理论结果, 空心圆圈表示数值计算结果, 空心三角表示 McEliece 等人的结果^[23]

将(9)式带入(10)式得到

$$N^2 \sqrt{\frac{\alpha}{2\pi}} \int_1^{\infty} \exp\left(-\frac{x^2}{2\alpha}\right) dx \geq 1, \quad (11)$$

利用数值积分可以得到 α_1 (即(11)式取等号时的存储率)与 N 的函数曲线, 例如 $N = 1000$ 的情况下得到 $\alpha_1 \approx 0.06$. 图 1(d) 给出了这样得到的 α_1 随 N 的变化曲线(实心圆圈). McEliece 等人曾经用和本文相似的思路推出 $\alpha_1 = 1/(2 \ln N)^{231}$, 如图中空心三角所示. 为了和理论结果相比较, 我们做了直接的数值计算, 结果由图中空心圆圈给出, 每个点都是 10 组不同的记忆模式在 $N = 1000$ 的网络下系综平均的结果(以下数值计算中网络的大小均固定在 $N = 1000$). 可见(10)式更符合数值计算结果.

但是, 从图 1(d) 中也可以发现, 如果以每个记忆模式都被严格存储为网络的不动点为判断标准, Hopfield 网络的存储能力相对而言很低. 在神经网络的研究中, 实际上采用了一个更加广义的联想记忆定义. 为此首先定义 $d_s = 1 - \frac{1}{N} \sum_{i=1}^N \text{sign}(\xi_i) \text{sign}(\xi_i')$ 为两个状态 ξ_i 和 ξ_i' 之间的距离, 它描述这两个状态的相似程度. 如果一些初始状态被吸引到一些不动点吸引子上, 而这些吸引子和某个记忆模式吸引

子 $\{\xi_i^c\}$ 的重合度高于预先规定的一个阈值 d_c ,则也认为这个记忆模式被该神经网络存储,也就是说 $\{\xi_i^c\}$ 不一定要满足不动点条件. 阈值 d_c 一般被设置为 $d_c = 0.05^{[8]}$,因为生物学上的一些实验表明,生物神经系统对于 5% 以内差异的目标是不能分辨的.

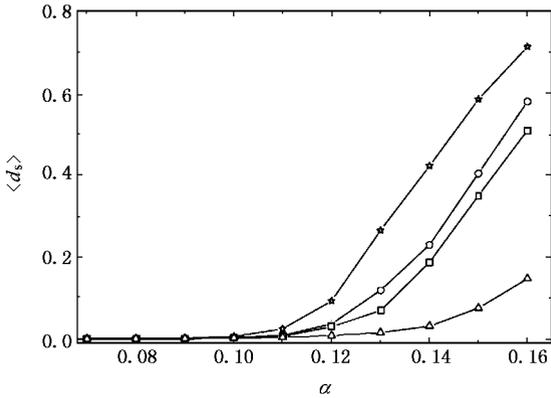


图2 平均距离 $\langle d_s \rangle$ 随存储率 α 的函数关系

以 Hebb 学习规则设计的 Hopfield 网络有这样的特点,当存储率不很高的情况下,即便记忆模式本身不是网络的不动点,网络会把他们演化到与这个模式的距离非常近的一个不动点吸引子上. 图 2 中给出了以待记忆模式为初值而演化足够长的时间后得到的不动点与对应记忆模式之间的平均距离 d_s 随存储率的函数关系,其中,三角符号表示硬极限传输函数情况下的结果,五角星、空心圆圈、空心方块符号则分别表示连续传输函数在 $\beta = 6, 10, 14$ 的情况下的结果. 可以看出,当 $\alpha < \alpha_1$ 时,平均距离均为零,随后平均距离随着存储率的增加而增加,当 $\alpha \sim 0.14$ 时硬极限传输函数对应的网络平均距离超过阈值 d_c . 我们注意到在推广的定义下,利用平均场理论计算出的硬极限传输函数 Hopfield 网络的最大存储率是 $\alpha_2 = 0.138^{[24, 25]}$,也就是说,以图 2 的方式可以很方便地得到网络的最大存储率. 因此,图 2 还显示在采用连续传输函数的情况下,网络的最大存储率(仍然以 $d_c = 0.05$ 作为判断标准)随着 β 的减小而减小. 也就是说,在采用 Hebb 学习规则的情况下,连续传输函数的引入实际上降低了网络的最大存储能力.

2.2. Hopfield 网络记忆模式的吸引域

一个被设计为吸引子的记忆模式不一定能被联

想出来,从动力学的角度看,这个吸引子还需要有足够大的吸引域. 吸引域越大,能够被吸引到吸引子上的状态就越多,从而这个模式被存储的质量就越高. 定义 P_{total} 为随机选取的模式被吸引到记忆模式吸引子上的概率. 显然,这个概率可以度量被网络存储的记忆模式的总的吸引域.

图 3 显示了连续传输函数情况下增益因子 β 对 Hopfield 网络吸引域的影响,其中纵坐标是 P_{total} ,横坐标是 β ,方块、圆圈、三角符号分别表示存储率 $\alpha = 0.035, 0.045, 0.06$ 的结果. 这里,每个点都是随机选取 10000 个初值下的计算结果.

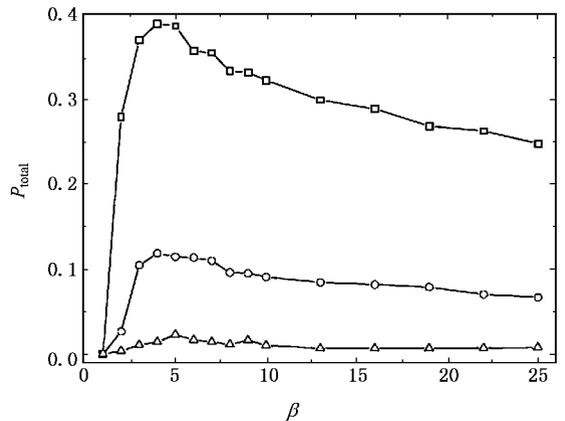


图3 连续传输函数随增益因子 β 对 Hopfield 网络吸引域的影响

从图 3 可以看出, P_{total} 在 $\beta \sim 6$ 的地方出现极大值,注意到 $\beta \rightarrow \infty$ 对应硬极限传输函数,这一结果显示选择具有合适增益因子的连续传输函数可以增大记忆模式的总吸引域.

不过这并不能表明采用连续传输函数就一定能得到比硬极限传输函数更好的网络特性. 一方面,并不是任意的 β 都能使记忆模式的吸引域增大,即使是在最优的选择下, P_{total} 的改善也并不显著. 就如下一节中将会看到的,在整体优化网络中连续传输函数的引入有时也会减小记忆模式吸引域;另一方面, P_{total} 所定义的吸引域并不是刻画记忆模式存储质量的一个非常合适的指标. 联想记忆的目的是当一个与特定存储模式有一定相似性的状态输入给网络,网络能够联想起所对应的特定模式,如果一个完全没有相似性的状态被吸引到这个模式上,这实际上相当于一种错误联想,而 P_{total} 并不能区分这两种情况. 为此,定义 P_{eff} 为随机选取的、与记忆模式其中之一在一定的距离 d_s 内的初始状态被吸引到对应的记忆模式上的概率,图 4 给出了存储率 $\alpha = 0.05$

情况下 P_{eff} 随距离 d_s 的变化曲线,其中实心圆圈表示硬极限传输函数(相当于 $\beta = \infty$)下的结果,空心圆圈表示连续传输函数在 $\beta = 6$ 下的结果.图中的每个数据是在随机选取 10000 个初态计算的结果, P_{eff} 是对所有记忆模式的平均.可以看出,当 d_s 比较小时所有的初态都被正确吸引到了记忆模式上,当 d_s 超过阈值 $d_s \sim 0.36$, P_{eff} 迅速减小.因此这个阈值决定了一定能够被记忆模式吸引子吸引的那些初态与吸引子的相似度,也就是说,将任意一个与某个记忆模式吸引子具有 64% 重合元素的状态为初值输入网络后,网络都能正确联想起对应的记忆模式.由此可见,这个阈值度量记忆模式的有效吸引域.图 4 所给出的结论是:硬极限传输函数和连续传输函数(即使采用对 P_{total} 来说是最优的 β)对记忆模式的有效吸引域几乎没有影响.

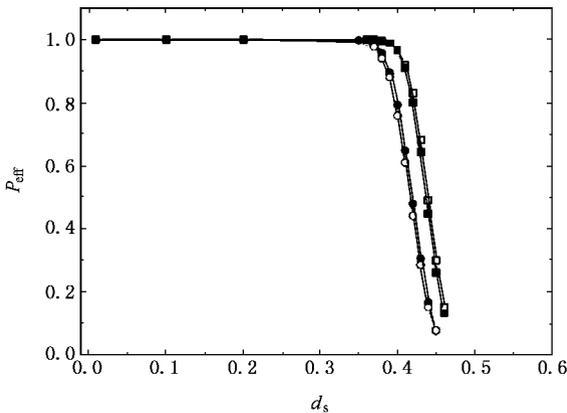


图 4 $\alpha = 0.05$ 时, P_{eff} 随距离 d_s 的变化曲线.其中实心圆圈表示 Hopfield 网络在硬极限传输函数下的结果,空心圆圈表示 Hopfield 网络在连续传输函数 $\beta = 6$ 下的结果;实心方块表示硬极限传输函数下整体优化网络的结果,空心方块表示连续传输函数在 $\beta = 0.04$ 下整体优化网络的结果

3. 整体优化网络

利用蒙特卡罗变异-优化选择规则可以设计连接权值为整数的非对称权值矩阵,本文研究权值矩阵元限定为 $J_{ij} = \pm 1$ 的整体优化网络,详细设计算法可参看文献 [15],这里不再赘述.需要特别强调,蒙特卡罗变异-优化选择规则是基于全局优化的设计思想,它不同于 Hebb 规则和感知器规则等学习规则,运用传统的这些规则时,记忆模式是一个接着一个地放进去学习,因此它们都是局域学习规则,而蒙特卡罗变异-优化设计规则针对全体记忆模式寻找

最优解(仿逆规则虽然也是整体学习规则,但它只适用于对称网络的设计),具体做法是通过随机变异-优化选择的反复操作使神经元连接权值向着整体最优的方向演化,最终达到预设的目标,如所有的记忆模式满足 $\bar{h}_i^c \geq c$.参数 c 是预先给定的正数,它控制网络的动力学特性.在硬极限传输函数的情况下,随着参数 c 由小变大,网络的动力学行为依次表现混沌相、记忆相、混合相等三种不同的动力学特性.在具有混沌相的网络中,记忆模式已经被设计成网络的不动点吸引子,这些不动点的吸引域成为一个个镶嵌在“混沌海”中的小岛,因此这种网络非常适合模式识别^[26];在具有记忆相的网络里,所有的初态都将演化到记忆模式中的其中之一上,也就是说这种网络中没有任何多余的吸引子或亚稳态存在.设计在混合相区的网络虽然也能正确地存贮记忆模式为其不动点吸引子,但同时也出现了伪吸引子.

由此可见,只要被设计在记忆相区,整体优化网络就不具有任何的伪吸引子.消除了伪吸引子也就消除了反馈网络的周期性运动而使网络输出总是趋于稳态,同时,由于所有的初态都会被吸引到某个记忆模式吸引子上,记忆模式吸引子的吸引域从而被最大化.这也是蒙特卡罗变异-优化选择规则的主要优点之一,在以往的文献中,有很多专门研究如何减小伪吸引子的数目^[11, 27, 28],用连续传输函数代替硬极限传输函数是以前的研究者实现这一目标的重要方法.

设计在记忆相区的整体优化网络则完全消除了伪吸引子而不需要引入连续传输函数.事实上,在整体优化网络中引入连续传输函数反而会使消失的伪吸引子重新出现.图 5 显示了具有不同增益因子 β 的连续传输函数下 P_{total} 随参数 c 的变化特性,这也反映了不同的传输函数对记忆模式吸引子吸引域的影响.这里的计算结果是在 $\alpha = 0.05$ 情况下得到的.图中的实心圆圈符号给出的是硬极限传输函数($\beta = \infty$)下 P_{total} 随参数 c 的变化特性,空心圆圈和空心方块分别给出连续传输函数在 $\beta = 0.6$ 和 $\beta = 0.04$ 下的结果.在硬极限传输函数下, P_{total} 明显地分成三个不同特性的区域,即 $P_{\text{total}} = 0$, $P_{\text{total}} = 1$ 和 $P_{\text{total}} < 1$ 的区域,其中 $P_{\text{total}} = 0$ 对应混沌相区,此区域中记忆模式吸引子的吸引域相对于“混沌海”小到可以忽略, $P_{\text{total}} = 1$ 是记忆相区,这个区域中记忆模式的吸引域完全充满了状态空间,也就是说这个相区不存在额外的吸引子(即使存在它们的吸引域尺度和记

忆模式吸引子的吸引域尺度相比也是可忽略的小), $P_{\text{total}} < 1$ 标志混合相, 在这个相区开始出现伪吸引子.

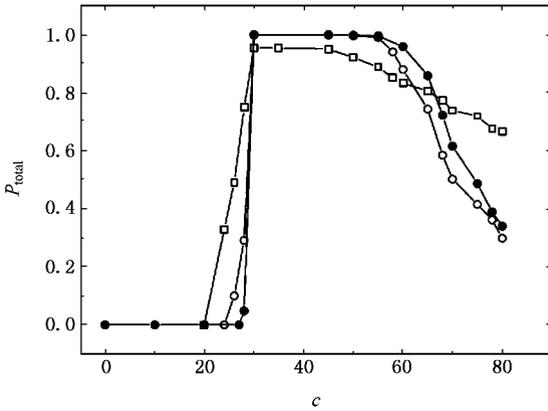


图5 具有不同增益因子的传输函数下整体优化网络 P_{total} 随参数 c 的变化特性

图5的结果表明连续传输函数的引入对记忆模式吸引子的吸引域乃至网络的特性的影响基本上是负面的, 当 β 较大而更接近硬极限传输函数时, P_{total} 的特性基本保持不变, 严格一些讲, 连续传输函数的引入轻微地缩小了记忆相区的宽度, 并且在混合区减小了记忆模式的吸引域. 当 β 较小时, 虽然在混合区参数 c 较大的区域记忆模式的吸引域大为增加, 但是在参数 c 较小的区域却有所减小, 更为严重的是连续传输函数的引入使得记忆相区 $P_{\text{total}} = 1$ 的特性消失, 也就是说使得完全消失的伪吸引子在这个区域中重新出现.

混合区中参数 c 较大的区域记忆模式吸引子的吸引域虽然大为增加, 但是同样并不意味着有效存储能力的增加. 图4中也显示了传输函数对整体优化网络记忆模式有效吸引域的影响, 实心方块是硬极限传输函数, 空心方块是 $\beta = 0.04$ 的结果, 这里参数 $c = 85$. 可见, 虽然在对应的参数下连续传输函数使得记忆模式的总吸引域扩大, 但有效吸引域并没有明显的变化(实际上相对于硬极限传输函数还略为减小). 另外需要指出 β 的数量级与 Hopfield 网络情况有所不同, 这仅仅是因为两种网络所采用的权值矩阵的矩阵元的定义造成的, β 的大小只能根据特定的网络本身的权值矩阵元的大小定义, 这一点在下一节中会被清楚地揭示.

4. 传输函数影响网络特性的机理

传统上, 常用能量函数(即李雅普诺夫函数)分析和理解神经网络的特性^[29, 30]. 能量函数分析方法的缺点是不能很好地提供直观图像, 同时, 在非对称网络情况下也不能一般地定义出能量函数. 本文引入一种新的方法, 即采用网络的吸引子对应的神经元局域场分布分析方法来研究网络的动力学结构并理解神经元传输函数对网络的影响.

具体地, 首先利用设计规则设计出权值矩阵并选择不同的传输函数演化网络, 直到网络趋于稳态, 即演化到吸引子上(吸引子可以是不动点, 周期轨道, 混沌轨道)然后分别计算记忆模式吸引子和伪吸引子对应的局域场分布. 以某个待记忆模式为初态演化得到的吸引子如果和这个记忆模式的差异不大于 5%, 则这个吸引子是记忆模式对应的有效吸引子. 以描述记这些有效吸引子的神经元状态根据(2)式计算的局域场为记忆模式对应的局域场. 用随机初态为初值而不能演化到记忆模式吸引子上的吸引子是伪吸引子, 以描述记这些伪吸引子的神经元状态根据(2)式计算的局域场为伪吸引子对应的局域场. 图6(a), (b)分别显示 Hopfield 网络和整体优化网络记忆模式吸引子和伪吸引子对应的局域场分布, 分布函数都被归一化, 网络的存储率都是 $\alpha = 0.05$. 图6中, 空心三角和实心三角分别表示硬极限传输函数和连续传输函数下记忆模式吸引子对应的局域场分布, 空心圆圈和实心圆圈分别表示硬极限传输函数和连续传输函数下伪吸引子对应的局域场分布. 增益因子 β 在 Hopfield 网络中取值为 $\beta = 6$, 在整体优化网络中取值为 $\beta = 0.04$, 同时后者所用到的控制参数 c 被选定在 $c = 85$, 这两个增益因子分别导致对应的网络记忆模式吸引子的总吸引域最大, 如图3, 图5所示.

对于硬极限传输函数, 由第2节的讨论已知, Hopfield 网络在 $\alpha = 0.05$ 的存储率的情况下记忆模式本身就是不动点吸引子, 而在整体优化网络情况下, 每个记忆模式已被严格设计为网络的不动点吸引子, 因此记忆模式吸引子对应的局域场分布刚好就是 h_i^c 的分布, 由学习规则或设计算法所完全确定. 无论是 Hebb 规则还是蒙特卡罗变异-优化规则其实质都是一方面确保 $h_i^c > 0$ 如果 $\xi_i^c = 1$ 或者 $h_i^c < 0$ 如果 $\xi_i^c = -1$, 并且尽量使局域场分布远离坐标原

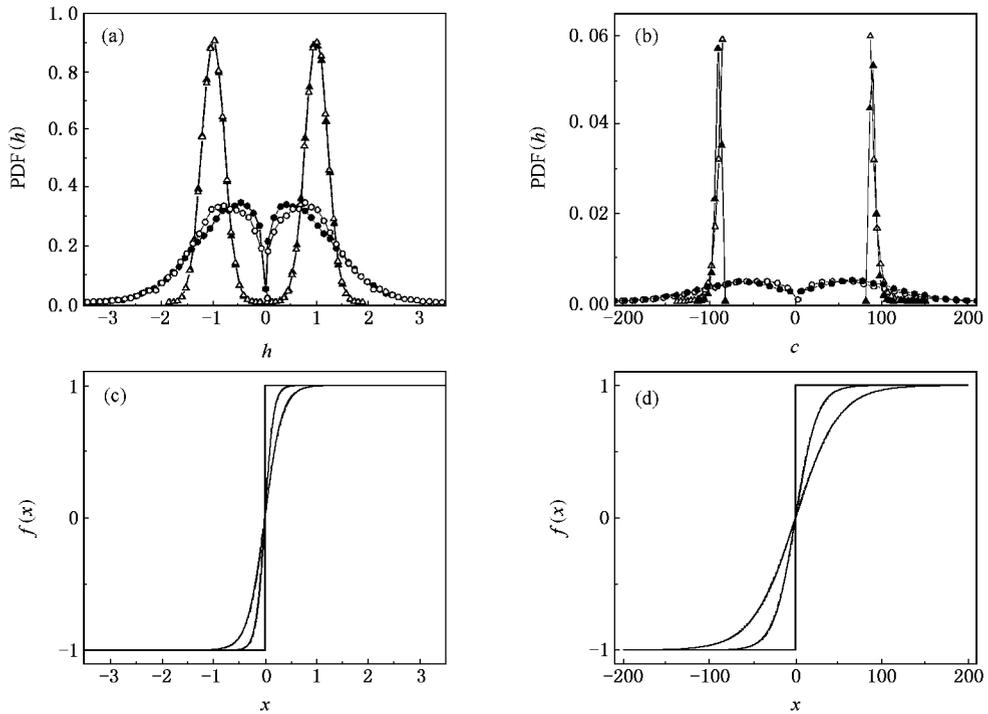


图6 记忆模式吸引子和伪吸引子对应的局域场分布 (a)Hopfield网络 (b)整体优化网络 (c)和(d)分别显示对应于(a)和(b)局域场尺度的传输函数.从y轴正方向数起,依次是(c) $\beta = \infty$ (相当于符号函数), $\beta = 6$ 和 $\beta = 3$; (d) $\beta = \infty, 0.04, 0.02$

点,因为 h_i^u 的绝对值,即 \bar{h}_i^u 决定了记忆模式的有效吸引域,也就决定了存储的质量.

但是,由于Hebb学习规则和蒙特卡罗变异-优化规则是两种不同的设计规则,所以由这两种规则设计的网络,它们的局域场分布存在很大差别,主要体现在记忆模式吸引子对应的局域场分布上.因此,神经元传输函数对于网络的局域场的响应是不同的.当以连续传输函数取代硬极限传输函数后,局域场会被扰动,扰动的效果一方面取决于局域场分布的特性,一方面取决于增益因子 β 的大小, β 越小扰动的程度就越大,而 β 值大小的度量依赖于记忆模式吸引子对应的局域场分布的间隙宽度.

首先我们来分析连续传输函数对这两种网络记忆模式对应的局域场分布的影响.如图6(a)和(b):Hebb规则情况下记忆模式对应的局域场是以 $h_i^u = \pm 1$ 为中心的两个高斯分布,而由蒙特卡罗变异-优化规则设计的网络,记忆模式对应的局域场分布被严格限制在 $\bar{h}_i^u \geq c$ 的区域.在图6(a)中,间隙宽度大约为 $|h| \sim 0.5$,而如图6(c)所示,增益因子为 $\beta = 6$ 的连续传输函数引起的扰动(相对于硬极限传输函数)刚好能达到这样的量级,这样的扰动对记忆

模式对应的局域场没有构成大的影响,但对于许多局域场落入间隙内的伪吸引子则构成大的扰动.如果增益因子进一步减小,如减小到如图6(c)所示的 $\beta = 3$,则引起的扰动已经足以破坏记忆模式吸引子的稳定性.在Hopfield网络中,“间隙”是随着存储率的增加而减小的,这也意味着能够扰动伪吸引子而基本不扰动记忆模式吸引子的 β 随着存储率的增加而增大.对于图6(b)所示的整体优化网络情况下,“间隙”的宽度大约为 $|h| \sim 100$,要达到这样量级的扰动,如图6(c)所示,增益因子大约应当在 $\beta = 0.04$ 左右,这个量级的 β 同样只对记忆模式吸引子构成微扰,而对于伪吸引子则构成大扰动.

我们说在上述参数下连续传输函数对于记忆模式吸引子的扰动可以看成微扰,可以从以下两个方面进一步说明.一方面,简单的计算可以直接验证连续传输函数没有破坏任何一个记忆模式作为不动点的条件,也就是说,根据第1节的定义的 d_s ,同一个记忆模式和对应的吸引子之间的距离在硬极限传输函数和连续传输函数情况下都保持为零,这也意味着连续传输函数的引入没有改变记忆模式吸引子局域场 h_i^u 的符号.另一方面,从图6(a)和(b)中可以看

到两种传输函数下记忆模式吸引子对应的局域场分布也没有显著的变化.这就是说,当连续传输函数的增益因子作用范围和记忆模式吸引子对应的局域场分布之间的间隙宽度相比在相同的范围的时候,增益因子对记忆模式吸引子的局域场只能构成微扰,不会改变吸引子的稳定性.

下面,我们分析连续传输函数对网络的伪吸引子对应的局域场分布的作用,以及由此引起的对网络功能的影响.如图(a)和(b)所示,在这两种网络中,伪吸引子对应的局域场分布区域和记忆模式吸引子对应的局域场分布区域有明显的差别.虽然坐标原点处有小的“凹陷”,但整体而言伪吸引子对应的局域场分布可以看成是以坐标原点为中心的高斯函数.这种分布特点使得连续传输函数引起的扰动对于伪吸引子而言非常显著,使得当引入连续传输函数后原来的许多伪吸引子对应的局域场符号发生改变,从而使得许多伪吸引子的稳定性被破坏,从而减少了伪吸引子的数量,这一点也可以通过简单的计算直接验证.值得指出的是,这一特性从分布函数不能体现出来,分布函数是被归一化了的,不反映伪吸引子的数量变化.扰动前后的伪吸引子对应的局域场分布没有发生明显变化的事实说明伪吸引子存在于相同的区域中.以前的一些文献根据伪吸引子减小的事实推断出吸引子的吸引域一定会增加,认为这样将增加随机选取的初态被吸引到记忆模式上的机会^[9-12].然而这种观念不够准确,连续传输函数的引入并没有破坏所有硬极限传输函数下的伪吸引子的稳定性,同时一些新的伪吸引子也随之产生,这样记忆模式吸引子和伪吸引子的吸引域要重新组合,没有理由断定记忆模式的吸引域一定会扩大.

记忆模式的吸引域是否扩大主要决定于两种吸引子对应的局域场分布特征.比如在图(b)所示的整体优化网络,伪吸引子对应的局域场和记忆模式吸引子对应的局域场分别分布在很不同的区域,并且后者远离坐标原点,这样,具有较大的增益因子的连续传输函数只能影响部分伪吸引子,由于对局域场的扰动量很小,使得失去稳定性的伪吸引子被吸引到了留下来的伪吸引子上,因为它们处在相同的区域中.这种机制可以造成两种吸引子的吸引域基本不变,甚至有可能使得记忆模式吸引子的吸引域相对缩小,因为伪吸引子的相互合并可能增加伪吸引子的整体吸引域,从而解释了图5中 $\beta = 0.6$ 的结果.当 β 进一步减小,如 $\beta = 0.04$ 传输函数引起的扰

动就会影响到大部分的伪吸引子,扰动的尺度也足以使它们进入记忆模式吸引子区域,虽然这种情况下记忆模式的吸引子也会被扰动(但是如前所述这种扰动只是轻微地改变了记忆模式对应的局域场分布)这样,被破坏稳定性的伪吸引子就会有更多的机会落入记忆模式吸引子上,从而增大了记忆模式的吸引域并解释了如图5中 $\beta = 0.04, c > 65$ 的情况下记忆模式总吸引域扩大的现象.

而Hebb规则设计的网络中,两种吸引子对应的局域场分布区域重叠度较高,尤其当网络的存储率较大的情况下,因此被扰动破坏的伪吸引子有更多的机会被扰动到记忆模式吸引子的吸引域内,从而扩大了记忆模式的总吸引域.

吸引子对应的局域场的绝对值 \bar{h}_i^{μ} 决定了有效吸引域的大小^[15,31],在 $\beta = 6$ 的Hopfield网络和 $\beta = 0.04, c = 85$ 的整体优化网络下,如上所述记忆模式吸引子对应的局域场基本没有被改变,这也就意味着具有这样的增益因子的连续传输函数并没有改变记忆模式的局域稳定性,从而解释了连续传输函数的引入几乎不影响记忆模式的有效吸引域的原因.

5. 总结与讨论

反馈神经网络的性能主要决定于网络的动力学结构,也就是网络吸引子对应的局域场分布.控制局域场分布的有效手段是学习规则或设计规则,对于Hebb规则设计的Hopfield网络来讲,记忆模式吸引子对应的局域场分布仅仅是存储率的函数,无法实行进一步的外部控制,而蒙特卡罗变异.优化选择规则可以在给定的存储率下根据需要控制记忆模式的局域场分布,以达到进一步优化和改善网络的性能的目的.利用特定的设计规则设计的网络,其记忆模式对应的局域场分布也可以通过引入不同的传输函数而有所改变,不过这种改变不是“定向的”,也就是说不能根据需要而控制,这样就会造成网络的某些性能指标被改善或变差.

引入连续传输函数导致Hopfield网络的最大存储能力减小,也导致整体优化网络的“记忆相”区变小甚至消失.连续传输函数虽然可能导致记忆模式的总的吸引域变大(也可能变小),但是并没有改变它们的有效吸引域.有效吸引域内的状态和对应的吸引子具有确定的相似度,这样的状态和记忆模式建立联系在物理上是“合理”的,而那些与记忆模式

没有相似性却因为某种扰动被吸引到某个记忆模式的吸引域内,这实际上属于错误的“联想”。因此有效吸引域才是度量网络性能的合适的指标。

那些被认为由连续传输函数带来的益处,如消除伪吸引子、消除网络的周期性摆动等都可以通过合适的设计方法(如应用蒙特卡罗变异-优化选择规则)在硬极限传输函数的情况下得到好得多的效果。简单的计算还可以证明网络的收敛时间也不会因为引入连

续传输函数而有所提高。总的来讲,连续传输函数对反馈神经网络的功能的改善几乎没有积极的贡献,设计规则的优劣才是决定网络性能的关键因素。

本文同时表明吸引子局域场分析的方法是研究神经网络性能的有效和更为直观的手段,利用这一手段,可以清晰地理解神经元传输函数起作用的机制,也可以得到关于网络结构与网络特性相联系的线索和知识。

- [1] Toulouse G 1989 *J. Phys. A: Math Gen* **22** 1959
- [2] Hopfield J J 1984 *Proc. Natl. Acad. Sci. U.S.A.* **81** 3088
- [3] Shiino M, Fukai T 1990 *J. Phys. A* **23** L1009
- [4] Brunel N, Zecchina R 1994 *Phys. Rev. E* **49** R1823
- [5] Kinsella J 1992 *Network: Comput. Neural Syst.* **3** 27
- [6] Barber M J, Clark J W, Anderson C H 2003 *Phys. Rev. E* **68** 041912
- [7] Hagan M T, Demuth H B, Beale M H 2002 *Neural Network Design* (Beijing: China Machine Press) [in Chinese] [Hagan M T, Demuth H B, Beale M H 2002 神经网络设计(北京:机械工业出版社)]
- [8] Wei H K 2005 *Theory and Method of Neural Net Structure Design* (Beijing: Defence Industry Press) [in Chinese] [魏海坤 2005 神经网络结构设计的理论和方法(北京:国防工业出版社)]
- [9] Marcus C M, Waugh F R, Westervelt R M 1990 *Phys. Rev. A* **41** 3355
- [10] Marcus C M, Westervelt R M 1989 *Phys. Rev. A* **40** 501
- [11] Waugh F R, Marcus C M, Westervelt R M 1990 *Phys. Rev. Lett.* **64** 1986
- [12] Marcus C M, Waugh F R, Westervelt R M 1991 *Phys. Rev. A* **43** 3131
- [13] Hopfield J J 1982 *Proc. Natl. Acad. Sci. U.S.A.* **79** 2554
- [14] Hopfield J J, Feinstein D I, Palmer R G 1983 *Nature* **304** 158
- [15] Zhao H 2004 *Phys. Rev. E* **70** 066137
- [16] Lyons W B *et al* 2001 *Physiol. Meas.* **22** 635
- [17] Hopfield J J, Tank D W 1985 *Biol. Cybern.* **52** 141
- [18] Nasrabadi N M, Choo C Y 1992 *IEEE Trans. on Neural Networks* **3** 5
- [19] Pajares G, Cruz J M, Aranda J 1998 *Pattern Recognition* **31** 5610574
- [20] Zhang Q, Gao L, Wang C, Xu J 2003 *Acta Phys. Sin.* **52** 1600 (in Chinese) [张 强、高 琳、王 超、许 进 2003 物理学报 **52** 1600]
- [21] Ma Y Q, Zhang Y M, Gong C D 1993 *Acta Phys. Sin.* **42** 1356 (in Chinese) [马余强、张 明、龚昌德 1993 物理学报 **42** 1356]
- [22] Xiong T, Zhang B L, Chang S J *et al* 2005 *Acta Phys. Sin.* **54** 2435 (in Chinese) [熊 涛、张便利、常胜江等 2005 物理学报 **54** 2435]
- [23] McElice R J *et al* 1987 *IEEE Trans. on Information Theory* IT-33 4
- [24] Amit D J, Gutfreund H, Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530
- [25] Amit D J, Gutfreund H, Sompolinsky H 1987 *Ann. Phys.* **173** 30
- [26] Jin T, Zhao H 2005 *Phys. Rev. E* **72** 066111
- [27] Krebs P R, Theumann W K 1999 *Phys. Rev. E* **60** 4580
- [28] McGraw P N, Menzinger M 2003 *Phys. Rev. E* **67** 016118
- [29] Tian W, Wang Y N 2005 *Chin. Phys.* **14** 72
- [30] Zhang Q, Ma R N, Wang C, Xu J 2003 *Chin. Phys.* **12** 22
- [31] Gardner E 1988 *J. Phys. A* **21** 257

The role of neuron transfer function in artificial neural networks *

Wang Rui-Min¹⁾ Zhao Hong^{2)†}

1) (*Department of Physics , Lanzhou University , Lanzhou 730000 , China*)

2) (*Department of Physics , Xiamen University , Xiamen 361005 , China*)

(Received 29 April 2006 ; revised manuscript received 3 June 2006)

Abstract

By analysis of local field distribution of the neurons in stationary state of associative memory neural networks , the role of the analog neuron transfer function in affecting the neural network performance is re-investigated . Different from the research done before , we find that the analog transfer function has no obvious advantages over the hard limit transfer function . Furthermore , analog transfer function sometimes produces a negative impact on certain functions of the network , such as the maximal storage capacity . We show that in pursuing the same performance a proper design rule is more essential than the choice of the transfer function .

Keywords : associative memory , neural networks , attractor , local field distribution

PACC : 0545 , 8710 , 8730G

* Project supported by the National Natural Science Foundation of China (Grant No. 10475067) and the Doctor Education Fund of Ministry of Education of China .

† Corresponding author. E-mail : zhaoh@xmu.edu.cn