

环结构神经网络及其互联想性能

张永军

陈宗鹭

(国防科工委指挥技术学院 北京 101407) (中国科学院电子学研究所 北京 100080)

摘要 本文提出了一种称为环结构神经网络(LANN)模型及其学习算法。它能像Hopfield网络,双向联想记忆(BAM)网络和其它类似网络一样工作,特别是它能执行多类样本之间的互联想记忆。理论分析和计算机模拟都证明LANN具有很好的收敛性,是一种有效的网络结构。最后本文给出了计算机模拟结果。

关键词 神经网络,环结构,学习算法,模拟

1 引言

Hopfield提出的Hopfield模型成功地用于联想记忆和组合优化等问题^[1-3]。但是由于该网络结构的限制,主要用于自联想记忆。双向联想记忆(Bidirectional Associative Memory, BAM)网络在这一点上前进了一步,它能完成两类样本之间的联想记忆^[4]。这些网络都能用于模式识别,但是怎样完成多于两类的样本之间的联想,无论是Hopfield网络还是BAM网络都有局限性。本文提出的环结构神经网络(Loop Architectur Neural Network, LANN)就是针对这一任务而构成的。它能完成多类样本之间的联想记忆,并且在离散样本的情况下,证明能以很高的概率收敛。

2 环结构神经网络

LANN的基本结构如图1所示。图中为一个三层神经网络构成的环路。每一层把上一层的输出作为输入信号,而把自己的输出送到下一层的输入端。最后一层的输出反馈到第一层的输入端,从而构成了一个闭合的环路。设每一层有 N 个神经元,三层分别记为 X, Y 和 Z 。 X 层中的每一个神经元与 Y 和 Z 层中的每一个神经元互相连接, X 层内的神经元之间不互相连接。另外,每一层还有自己的外部输入端。 Y 层和 Z 层的结构与 X 层类似,每一层的结构如图2所示。这一结构很容易推广到具有 $L(L > 3)$ 层的环结构网络,加权矩阵由Hebb规则决定^[4],如从 X 层到 Y 层的加权可表示为

$$W_{xy} = \sum_{m=1}^L Y(m)X(m)^T,$$

1994-06-07 收到, 1994-12-02 定稿

张永军 男, 1950年生, 副教授, 现主要从事的专业有信号处理、雷达、微波成像和人工神经网络。

陈宗鹭 男, 1922年生, 研究员, 博士生导师, 现主要从事微波成像、雷达信号处理, 信息论和神经网络等的研究工作。

其中 $X(m)^T = \{x_1(m), x_2(m), \dots, x_N(m)\}^T, Y(m)^T = \{y_1(m), y_2(m), \dots, y_N(m)\}^T, m = 1, 2, \dots, M, M$ 为样本数, T 表示转置运算。

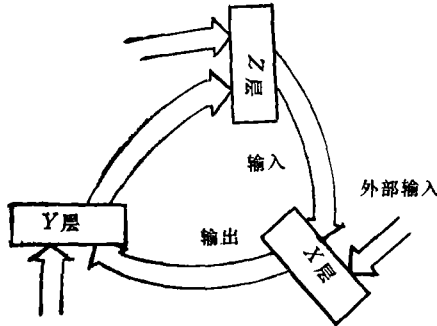


图1 三层环结构神经网络模型

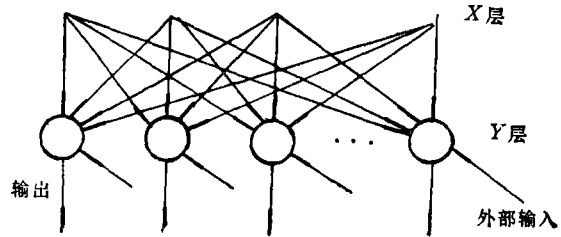


图2 单层网络的结构

3 LANN 的性能

当 $L = 1$ 时, 除网络加权矩阵对角线元素不为零外, LANN 与 Hopfield 网络一样。当 $L = 2$ 时, LANN 变为 BAM 网络。此时, 网络能像 Hopfield 网络或 BAM 网络一样完成自联想或异联想。当 $L \geq 3$ 时, LANN 与 Hopfield 网络和 BAM 网络一样具有很多相似的性能, 其主要特点是: LANN 能够实现三类或多于三类的样本之间的互联想。

下面分析 LANN 的收敛特性, 证明 LANN 能以很高的概率收敛到正确的稳态解。设网络的初始输入为 $X(m_0)$ (或 $Y(m_0), Z(m_0)$) 则 Y 层中第 i 个神经元的加权和 $\text{net}y_i(m_0)$ 为

$$\begin{aligned} \text{net}y_i(m_0) &= \sum_{j=1}^N W_{xy(i,j)} x_j(m_0) \\ &= \sum_{j=1}^N \sum_{m=1}^M y_i(m) x_j(m) x_j(m_0) \\ &= N y_i(m_0) + \sum_{m \neq m_0} [b_y(m, m_0)] y_i(m), \end{aligned} \quad (1)$$

式中 $\text{net}y_i(m_0)$ 为 Y 层第 i 个神经元的净输入, $b_y(m, m_0) = \sum_{j=1}^N x_j(m) x_j(m_0)$ 。

$\text{net}y_i(m_0)$ 由两项组成, 第一项正好为初始值 $X(m_0)$ 对应的 Y 层中的样本 $Y(m_0)$ 的第 i 个分量乘以 N , 在离散网络中, $y_i(m_0)$ 取 -1 或 1 , 第一项的绝对值总为 N 。第二项为一交叉项, 在各样本分量 $x_i(m), y_i(m)$ ($i = 1, 2, \dots, N; m = 1, 2, \dots, M$) 等概率地取 -1 和 1 的条件下, 不难证明 $b_y(m, m_0)$ 的均方根值为 \sqrt{N} , 而 $(M - 1)$ 项这样的项求和的均方根值为 $\sqrt{N(M - 1)}$ 。一般来说, N 远大于 M , 所以 $\text{net}y_i(m)$ 第一项的绝对值远大于第二项的平均绝对值, 故 $\text{net}y_i(m)$ 以很高的概率与 $\text{net}y_i(m_0)$ 同号。因此 $\text{net}y_i(m_0)$ 通过非线性函数可得

$$\text{sgn}[\text{net}y_i(m_0)] = y_i(m_0), \quad (2)$$

正好与样本 $Y(m_0)$ 对应的值一致。这样就完成了从样本 $X(m_0)$ 联想出对应样本 $Y(m_0)$ 的任务。进一步地可以从样本 $Y(m_0)$ 联想出 Z 层中的对应样本。

若样本分量取 -1 和 1 的概率不相等,用 $P\{\mu = a\}$ 表示 μ 为 a 的概率,设

$$P\{x_i(m) = 1\} = P\{y_i(m) = 1\} = P\{z_i(m) = 1\} = p, \quad (3a)$$

$$P\{x_i(m) = -1\} = P\{y_i(m) = -1\} = P\{z_i(m) = -1\} = q, \quad (3b)$$

且 $p + q = 1, p \neq q$ 时,情况要复杂一些,此时(1)式第一项的绝对值没有变化。令 $\mu_{ij}(n) = y_i(m)x_j(m)x_j(m_0)$, 则第二项中每一项的概率为

$$p_0 = P\{\mu_{ij}(m) = 1\} = 2pq^2 + p(q^2 + p^2), \quad (4a)$$

$$q_0 = P\{\mu_{ij}(m) = -1\} = 2p^2q + q(q^2 + p^2), \quad (4b)$$

第二项中 $N(M-1)$ 项随机相加,由二项式分布的概率可知,有 n 个 $\mu_{ij}(m)$ 等于 1 的概率为

$$P\{\sum \sum \mu_{ij}(m) = 2n - N(M-1)\} = C_{N(M-1)}^n p_0^n q_0^{N(M-1)-n}. \quad (5)$$

将 $\sum_{j=1}^M \sum_{m \neq m_0} \mu_{ij}(m)$ 均方值表示成 p_0 的函数,则

$$G(p_0) = 4N(M-1)p_0(1-p_0) + N^2(M-1)^2(2p_0-1)^2, \quad (6)$$

其中 $G(p_0)$ 表示 $\sum_{j=1}^M \sum_{m \neq m_0} \mu_{ij}(m)$ 的均方值。

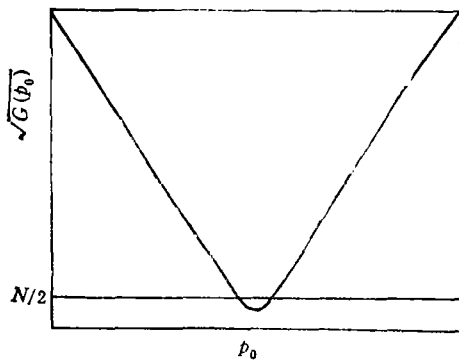


图3 $\sqrt{G(p_0)}$ 随 p_0 变化的曲线

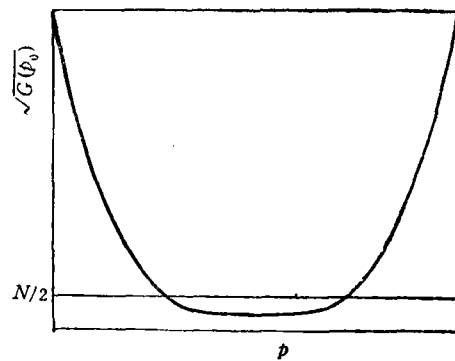


图4 $\sqrt{G(p_0)}$ 随 p 变化的曲线

对 $G(p_0)$ 求 p_0 的导数,并令 $G'(p_0) = 0$, 得 $p_0 = 1/2$, 且容易知道, $p_0 = 1/2$ 为 $G(p_0)$ 的最小值。求 $G(p_0)$ 对 p 的导数,并令 $G'_p[p_0(p)] = 0$, 则得在 $p \in (0, 1)$ 的区域内, $G'_p[p_0(p)]$ 只存在一个解 $p = 1/2$, 且为 $G[p_0(p)]$ 的最小值。由此可知,当 $p = 1/2$ 时,(1)式的第二项的影响最小,此时网络的联想效果最好。

图3给出了 $\sqrt{G(p_0)}$ 随 p_0 变化的曲线;图4给出了 $\sqrt{G(p_0)}$ 随 p 变化的曲线。由图3和图4可以看出,当 p 和 q 不对称时,即 $p \neq q$ 时,(1)式的第一项的绝对值将可能小于第二项的绝对值,此时一般会联想出错误的结果。为保证网络联想出正确的样本,这时存储的样本数应适当减少。存储样本数 M 与 p 的关系由(6)和(4a)式决定。

由(6)式知, $\sqrt{N(M-1)} \leq \sqrt{G(p_0)} \leq N(M-1)$, 只有 $\sqrt{G(p_0)} \ll N$, 网络才能

联想出正确的样本矢量, 即

$$4N(M-1)p_0(1-p_0) + N^2(M-1)^2(2p_0-1)^2 \ll N^2.$$

因上式左面第一项的最大值为 $N(M-1)$, 所以, $G(p_0)$ 接近 N^2 主要是第二项起作用, 为分析方便起见, 忽略第一项可得

$$|p_0 - q_0| \ll 1/(M-1). \quad (7)$$

图 5 给出了 (4a) 式的变换关系的曲线, 由图可以清楚地看出 p 在较大的范围内, 能保证 (7) 式成立, 这一特点有利于网络的联想操作. 若输入的样本 X' 不为 LANN 存储的样本 $X(m_0)$, 但与 $X(m_0)$ 具有最小的汉明距离 d_H , 类似的推导可知, (7) 式可改写为

$$|p_0 - q_0| \ll [1/(M-1)][(N-2d_H)/N]. \quad (8)$$

4 模拟结果

用 8×8 方阵组成的阿拉伯数字对 LANN 模型做了大量的计算机模拟. 模拟中我们取 $N = 64$, $L = 3$. 黑的象素点对应的样本分量值为 1, 白的象素点对应的样本分量值为 -1. 这些数字的方阵取 1 的象素约占 15%~28%, 以此作为取 1 的概率, 通过 (4a) 式的变换后, p_0 的分布在 0.33~0.45 的范围内. 图 6 给出了 $M = 3$ 时的模拟结果. X 层的存储样本为 {0}, {1} 和 {2}; Y 层的存储样本为 {3}, {4} 和 {5}; Z 层的存储样本为 {6}, {7} 和 {8}. 图中第一列为输入值, 第二至第八列分别为多次迭代的输出值; 第一, 二和三行分别对应着 X 层, Y 层和 Z 层的输入和输出样本, 为了避免取 1 的象素集中在方阵的中部, 对方阵的图形做了适当的位移, 以保证象素取 1 的概率近似为 p , 图中的输入样本与数字 {0} 的汉明距离为 $d_H = 7$. 模拟实验表明, 进一步增加存储样本数 M 或者增加输入样本与存储样本之间的汉明距离, 都有可能导致错误的联想结果. 这一结论与 (7) 式和 (8) 式的分析结果基本吻合.

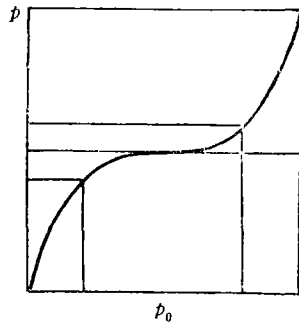


图 5 p 与 p_0 的关系曲线

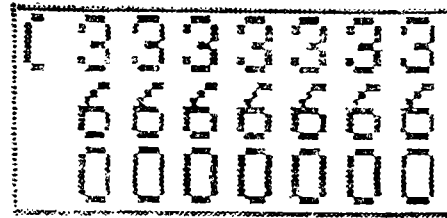


图 6 输入样本 {0} 部分数据的联想结果

5 结论

本文给出了一种新的 LANN 及其学习算法, 这种网络能完成多于两类的样本之间的联想操作. 在离散数据的情况下, 理论分析和计算机模拟结果都表明, LANN 具有很快的收敛速度, 并且 LANN 能从部分数据或受到干扰的数据联想恢复出正确的完整样本.

参 考 文 献

- [1] Farhat N H, Psaltis D, Prata A, *et al.* Applied Optics, 1985, 24(10): 1469—1475.
- [2] Hopfield J J. Neurons with graded response have collective computational properties like those of two states neurons. Proceedings of the National Academy of Sciences, USA; 31, May, 1984, 3088—3092.
- [3] Anderson J A. IEEE Trans. on Systems, Man, and Cybernetics, 1983, SMC-13(5): 799—815.
- [4] Kosko B. Applied Optics, 1981, 26(23): 4974—4979.

LOOP ARCHITECTURE NEURAL NETWORK

Zhang Yunjun

(*Institute of Command and Technology, COSTIND, Beijing 101407*)

Chen Zongzhi

(*Institute of Electronics, Academia Sinica, Beijing 100080*)

Abstract This paper provides a new architecture of neural network, called Loop Architecture Neural Network (LANN), and its learning rules. One of its distinguished features from other network, such as Hopfield and bidirectional associative memories, is that it can perform the associative memory among multiple categories. Analysis and simulated results have proved that it is an effective network with excellent convergence.

Key words Neural network, Loop architecture, Learning rule, Simulation