

PLN网络的改进及其应用*

张 钹

(清华大学计算机科学与技术系, 北京 100084)

张 铃

(安徽大学计算机系, 合肥 230039) (浙江大学计算机系, 杭州 310027)

摘要 本文在分析概率逻辑神经元(PLN)网络原型存在不足的基础上, 提出 PLN 元件的一个改进模型, 并指出原来的 PLN 网络模型、Hopfield 学习律以及 Boltzman 机的演化律等都是 PLN 网络改进模型的特例. 文中还给出 PLN 网络改进模型在联想记忆应用中的模拟实验结果, 说明改进模型无论在鲁棒性和收敛速度上比原型都有很大改进.

关键词 概率逻辑神经元(PLN)网络, A-学习律, 广义 A-学习律, 联想记忆.

在[1-4]中我们系统地讨论了概率逻辑神经元(PLN)网络的性质, 之后我们曾将该网络应用于模式识别, 结果发现其容错能力很差. 例如, 进行 0-9 数字的识别, 以标准的 $\{0, 1, \dots, 9\}$ 为样本, 构造一 PLN 网络 $N(H, I, R)$. 然后任取一待识别的样本 r , r 与某个数字(某个标准训练样本)有几处不同, 以 r 为输入, 进行识别, 重复 100 次, 结果发现由 r 出发收敛到 0-9 的各个数字的概率几乎相等. 即该 PLN 网络抗“噪音”的能力很差.

实验结果表明 PLN 元件原型有较大缺陷, 我们分析其基本原因如下. PLN 元件在原学习过程中, 若其输入是未学习过的, 其对应的输出总取为 u , 即以等概率取值 0 或 1, 这显然不很合理. 我们提出一个改进的 PLN 元件模型, 即广义 A-学习律, 并分析、研究了该学习律的性质, 与 Hopfield 学习律和 Boltzman 机演化律进行比较. 最后将它应用于联想记忆, 取得很满意的结果.

为叙述方便, 下面先简单介绍 PLN 网络的一些基本概念和性质.

PLN 元件(概率逻辑神经元元件)

设 PLN 元件 A 有 n 个输入变量 x_1, x_2, \dots, x_n (简记为输入向量 $x = (x_1, x_2, \dots, x_n)$), 一个输出 y , 分别取值于 $\{0, 1\}^n$ 和 $\{0, 1\}$.

y 与 x 的关系记为 $y = f(x)$.

假设: (1) A 可实现 $\{0, 1\}^n \rightarrow \{0, 1\}$ 的一切可能对应关系; (2) 关系 f 预先未确定, 是在记

* 本文 1993 年 11 月 7 日收到, 1993 年 12 月 22 日定稿

本项目得到国家攀登计划和 863 高技术研究与发展的支持. 作者张钹, 59 岁, 教授, 主要研究领域为人工智能、计算机应用技术. 张铃, 57 岁, 教授, 主要研究领域为人工智能, 应用数学.

本文通讯联系人: 张钹, 北京 100084, 清华大学计算机科学与技术系

忆及学习过程中逐步被(部分)确定.

设元件输入定义域为 I , 值域为 O .

令 $W = I * O$ 为训练样本空间, 即 $r \in W$ 有 $r = (x, y)$, 其中 $x \in I, y \in O$.

设 $K \subset W$, 称 K 是相容训练样本集, 若 $\forall r^1 = (x^1, y^1), r^2 = (x^2, y^2), r^1, r^2 \in K$, 当 $x^1 = x^2$ 有 $y^1 = y^2$. (文中不同向量用上标表示, 向量的分量用下标表示).

PLN 元件 A 经训练样本集训练过后, 其对应关系 f 定义如下, (即 A -学习律).

$$y = f(x) = \begin{cases} y^0 & \text{若存在 } r = (x, y^0) \in K \\ u & \text{其他} \end{cases} \quad (1)$$

其中 u 表示 y 以等概率输出 0 或 1.

PLN 网络

由有限个 PLN 元件组成的网络称为 PLN 网络, 设网络每层结点同步进行工作, 由某一时钟控制, 时间只取离散值, 如取 $t = 0, 1, 2, \dots$.

设 PLN 网络 N 由 m 个 PLN 元件 A_1, A_2, \dots, A_m 组成, A_i 输入向量定义域记为 $I(i)$, $i = 1, 2, \dots, m$, 则 $I = \cup I(i)$ 为网络输入向量定义域; $O = \{(y_1, \dots, y_m)\}$ 为网络输出向量值域.

约定用 $x_a \in I(i)$ 表 x_a 是 A_i 的某一输入变量, 而 $x \in I(i)$ 表 x 是 A_i 的输入向量.

连接结构

设 x_a, y_j 在 t 时刻的值分别记为 $x_a(t), y_j(t)$. 若 A_j 的输出 y_j 反馈连接到 A_i 的某一输入变量 x_a , 则 $\forall t$ 有 $x_a(t) = y_j(t)$, 这种反馈连接关系用有序指标对 (α, j) 表示.

令

$$H = \{(\alpha, j) | \exists i, x_a \in I(i), y_j \in O \text{ 且 } t \geq 1 \text{ 有 } x_a(t) = y_j(t)\}$$

即 H 是所有反馈连接对应的有序指标对组成的集合.

显然网络的连接结构由 $I(i), i = 1, 2, \dots, m$ 及 H 集合完全确定. 我们简记网络 N 为 $N(H, I)$.

A_i 的运行规则: 若在 t 时刻网络处于状态 $x(t) \in I$, 则在 $(t+1)$ 时刻元件 A_i 的输出 $y_i(t+1)$ 为:

$$y_i(t+1) = f_i(x(t))$$

其中 $f_i(\cdot)$ 是元件 A_i 的对应关系.

设 $s \in I$ 是一状态, s 对应于 $I(i)$ 中分量构成的子向量记为 $s(I(i))$, 称为 s 在 $I(i)$ 上的投影. 类似可以定义 s 对其他分量集合的投影.

PLN 网络的训练样本集

仍以 $W = I * O$ 定义 PLN 网络的训练样本空间, 其中 I, O 分别是 PLN 网络的定义域和值域.

有关相容训练样本集等概念均可推广到网络情况, 不再赘述.

设 K 为一网络训练样本集, 令:

$$K(I) = \{r(I) | r \in K\}$$

稳定状态: 设状态 s , 若在 t_0 时处于 s^0 状态, 当 $t \geq t_0$ 仍有 $s(t) = s^0$, 则称 s^0 是网络的稳定状态.

网络 N 的所有稳定状态集合记为 $S(N)$.

训练样本集 K 称为与网络 $N(H, I)$ 是一致的 $\Leftrightarrow \forall r = (x, y) \in K, (i, j) \in H$ 有当 $t \geq 1$ 时 $x_i(t) = y_j(t)$ 成立.

1 广义 A-学习律

1.1 A-学习律的分析

在 A-学习律中规定, 对任一元件 A_i , 只要其输入 x^0 是未学习过的 (即 $(x^0, y) \notin K; K$ 是训练样本集), 则其输出 $f_i(x^0) = u$, 这显然不太合理, 因为它不能表示未学知识与已学的知识的联系. 按理经过一段时间学习后, 网络的识别性能应有所提高, 比如, 有一输入 x^1 虽是未学习过的, 但与某一样本 $r^1(I)$ 很相近, 按理其输出应以较大的概率取 r^1 对应的输出, 而不是等概率取 0 或 1.

我们根据上述直观的想法, 对 A-学习律进行推广和改进.

1.2 广义 A-学习律

设 PLN 网络有 n 个元件 $A_i, i = 1, 2, \dots, n$.

其训练样本集 K 为, $K = \{r^1, r^2, \dots, r^m\}$ 是相容的.

$$\text{令 } K^0(i) = \{r^a | r^a(y_i) = 0, r^a \in K\}$$

$$K^1(i) = \{r^a | r^a(y_i) = 1, r^a \in K\}$$

定义向量 x 到向量集合 B 的距离函数

$$\rho(x, B)$$

$$\text{令 } D_i^0(x) = \rho(x, K^0(i))$$

$$D_i^1(x) = \rho(x, K^1(i))$$

设 $g(t_1, t_2): [0, \infty) \times [0, \infty) \rightarrow (-\infty, \infty)$ 是 t_1 的非增函数, t_2 的非减函数.

设 $f: (-\infty, \infty) \rightarrow [0, 1]$ 的单调非减函数且

$$\lim_{t \rightarrow -\infty} f(t) = 0, \lim_{t \rightarrow \infty} f(t) = 1.$$

最后令:

$$p_i(x) = f(g(D_i^1(x), D_i^0(x))) \quad (2)$$

定义. (广义 A-学习律) 设网络第 i 个元件, 当输入为 x 时, 它以概率 $p_i(x)$ 输出 1.

例:

令 $D_i(x, y)$ 表示向量 x, y 在 $I(i)$ 上的投影的海明距离.

令 x 到集合 B 的距离 ρ 为

$$\rho(x, B) = \inf_{r \in B} D_r(x, r) \tag{3}$$

取

$$g(t_1, t_2) = \frac{t_2}{t_1 + t_2} \tag{4}$$

$$f(t) = \begin{cases} \frac{(2t)^k}{2} & \text{当 } 0 \leq t \leq \frac{1}{2} \\ \frac{2 - (2(1-t))^k}{2} & \text{当 } \frac{1}{2} \leq t \leq 1 \end{cases} \quad k \text{ 是参数} \tag{5}$$

$$p_i(x) = F(t_1, t_2) = f(g(t_1, t_2)) \tag{6}$$

下面讨论以(6)定义的 PLN 网络的两个极限情况.

令 $k \rightarrow 0$ 得

$$f(t) = \begin{cases} 0 & t=0 \\ \frac{1}{2} & 0 < t < 1 \\ 1 & t=1 \end{cases} \quad \text{即} \quad y_i(t+1) = \begin{cases} 0 & D_i^0(x(t)) = 0 \\ 1 & D_i^1(x(t)) = 0 \\ u & \text{其他} \end{cases} \tag{7}$$

(7)式就是原来的 A-学习律((7)式与(1)式相同).

令 $k \rightarrow \infty$ 得

$$f(t) = \begin{cases} 0 & 0 \leq t < \frac{1}{2} \\ \frac{1}{2} & t = \frac{1}{2} \\ 1 & \frac{1}{2} < t \leq 1 \end{cases} \quad \text{即} \quad y_i(t+1) = \begin{cases} 0 & D_i^0(x(t)) < D_i^1(x(t)) \\ u & D_i^0(x(t)) = D_i^1(x(t)) \\ 1 & D_i^1(x(t)) > D_i^0(x(t)) \end{cases} \tag{8}$$

(8)式的物理概念很清楚,即当输入 $x(t)$ 与 y_i 取 0 集合的海明距离小于与 y_i 取 1 集合的海明距离时, $(t+1)$ 时刻的输出 $y_i(t+1)$ 就取 0, ...

这与我们开始时的直观分析相吻合.

命题 1. 若 $\rho(x, B)$ 取(3)式, $g(t_1, t_2)$ 取(4)式, 按(6)式得到的广义 A-学习律, 当训练样本集是相容且一致的, 则每个训练样本是稳定状态.

证明: 设 $x(t) = r^{-1}(I) \in K(I)$, 不妨设 $r^{-1} \in K^0(i)$

按定义得 $D_i^0(x(t)) = 0$, 由(3)-(5)式知 $p_i(x(t)) = 0$.

由定义得 $y_i(t+1) = 0$ 的概率为 1, 即 $y_i(t+1) = r^{-1}(y_i)$. 显然, 对各个 i 都有 $y_i(t+1)$

$$=r^1(y_i), i=1, 2, \dots, n.$$

即 $y(t+1)=r^1(O)$ ($r^1(O)$ 表 r^1 在输出 O 上投影)

又因为 K 是一致的, 故

$$x(t+1)=H(y(t+1))=x(t)$$

其中 H 表反馈连接关系.

最后得 $x(t)$ 是稳定状态, 证毕.

1.3 广义 A-学习律与 Hopfield 网络演化律的关系

Hopfield 网络中状态分量取 $1, -1$.

训练样本集 $K = \{r^1, r^2, \dots, r^m\}$

其演化规律为:

$$y_i(t+1) = \text{sgn}\left(\sum_{j=1}^n w_{ij} x_j(t)\right), \quad i=1, 2, \dots, n \quad (9)$$

其中 w_{ij} 由 Hebb 规则确定

$$w_{ij} = \sum_{u=1}^m r_i^u r_j^u$$

故

$$\begin{aligned} y_i(t+1) &= \text{sgn}\left(\sum_{j=1}^n w_{ij} x_j(t)\right) \\ &= \text{sgn}\left(\sum_{j=1}^n \sum_{u=1}^m r_i^u r_j^u x_j(t)\right) \\ &= \text{sgn}\left(\sum_{u=1}^m r_i^u \left(\sum_{j=1}^n r_j^u x_j(t)\right)\right) \\ &= \text{sgn}\left(\sum_{u=1}^m r_i^u \langle r^u, x(t) \rangle\right) \end{aligned}$$

其中 $\langle r, x \rangle$ 表 r 在 I 上投影与 x 的内积.

令

$$K^1(i) = \{r^u \mid r_i^u = 1, r^u \in K\}$$

$$K^{-1}(i) = \{r^u \mid r_i^u = -1, r^u \in K\}$$

$$\Sigma(i, 1) = \Sigma_1 \langle r^u, x(t) \rangle, \Sigma_1 \text{ 表对 } r^u \in K^1(i) \text{ 求和.}$$

$$\Sigma(i, -1) = \Sigma_{-1} \langle r^u, x(t) \rangle, \Sigma_{-1} \text{ 表对 } r^u \in K^{-1}(i) \text{ 求和.}$$

则

$$y_i(t+1) = \begin{cases} 1 & \text{当 } \Sigma(i, 1) > \Sigma(i, -1) \\ -1 & \text{当 } \Sigma(i, 1) < \Sigma(i, -1) \end{cases}$$

令

$$R^1(i) = \frac{1}{m_1} \Sigma_1 r^u \quad (R^{-1}(i) = \frac{1}{m_2} \Sigma_{-1} r^u)$$

其中 $m_1(m_2)$ 是 $K^1(i)(K^{-1}(i))$ 元素的个数, 而 $R^1(i)(R^{-1}(i))$ 是 $K^1(i)(K^{-1}(i))$ 的重心. 则有

$$\frac{1}{m_1} \sum_1 \langle r^u, x \rangle = \sum_1 \langle \frac{r^u}{m_1}, x \rangle = \langle R^1(i), x \rangle$$

$$\langle \frac{1}{m_2} \sum_{-1} \langle r^u, x \rangle = \sum_{-1} \langle \frac{r^u}{m_2}, x \rangle = \langle R^{-1}(i), x \rangle$$

得

$$\Sigma(i, 1) > \Sigma(i, -1) \Leftrightarrow \langle R^1(i), x \rangle > \langle R^{-1}(i), x \rangle$$

现定义向量 x 到集合 B 的距离 ρ^* 如下

$$\rho^*(x, B) = (x - \langle R(B), x \rangle) / 2 \tag{10}$$

其中 $R(B)$ 表 B 的重心.

于是得

$$\Sigma(i, 1) > \Sigma(i, -1) \Leftrightarrow \langle R^1(i), x \rangle > \langle R^{-1}(i), x \rangle$$

$$\Leftrightarrow \rho^*(x, K^1(i)) < \rho^*(x, K^{-1}(i))$$

于是(9)式最后化成

$$y_i(t + 1) = \begin{cases} -1 & \rho^*(x(t), K^1(i)) > \rho^*(x(t), K^{-1}(i)) \\ 1 & \rho^*(x(t), K^1(i)) < \rho^*(x(t), K^{-1}(i)) \end{cases} \tag{11}$$

与(8)式比较

$$y_i(t + 1) = \begin{cases} 0 & \rho(x(t), K^1(i)) > \rho(x(t), K^0(i)) \\ u & \rho(x(t), K^1(i)) = \rho(x(t), K^0(i)) \\ 1 & \rho(x(t), K^1(i)) < \rho(x(t), K^0(i)) \end{cases} \tag{12}$$

比较(11)式与(12)式, 知两者除定义的距离 ρ 不同外, 其形式是完全一致的, 故在这个意义下, 我们可以将 Hopfield 网络的演化律看成是广义 A-学习律的特例.

1.4 广义 A-学习律与 Boltzman 机演化律的关系

Boltzman 机(BM)是一种概率神经网络, 其元件之间的连接方式也是权-阈值形式, 但每个元件的输出是按如下的概率形式演化.

第 i 元件 y_i (当其输入向量为 x 时) 以概率 p_i 输出 1, 以 $(1-p_i)$ 输出 0.

$$p_i = \frac{1}{1 + e^{-I_i / T}} \tag{13}$$

其中

$$I_i = \sum_{j=1}^n w_{ij} x_j - \theta_i = \sum_{j=0}^n w_{ij} x_j, \quad x_0 = 1$$

由 1.3 节的推导知,当取 $w_{ij} = \sum_{a=1}^m r_i^a r_j^a$ 时, I_i 可化成下面形式.

$$I_i(x) = [m_1 \langle R^1(i), x \rangle - m_2 \langle R^{-1}(i), x \rangle]$$

其中 $m_1(m_2)$ 是 $K^1(i)(K^{-1}(i))$ 的规模.

按(11)式定义的 ρ^* , 可将 $I_i(x)$ 化为

$$I_i(x) = 2(m_2 d_i^{-1}(x) - m_1 d_i^1(x))$$

其中

$$d_i^1(x) = \rho^*(x, K^1(i)), \quad d_i^{-1}(x) = \rho^*(x, K^{-1}(i)) \tag{14}$$

取

$$g(t_1, t_2) = 2(m_2 t_2 - m_1 t_1) \tag{15}$$

$$f(t) = \frac{1}{1 + e^{-t/T}} \tag{16}$$

利用(14)–(16)式得

$$p_i(x) = f(g(d_i^1(x), d_i^{-1}(x))) = \frac{1}{1 + e^{-I_i(x)/T}}$$

即其对应的广义 A-学习律,就是 Boltzman 机的演化律.
故 Boltzman 机可以看成是 PLN 网络的特例.

一个神经网络,当作为一个联想记忆器时,我们希望所有训练样本都是网络的稳定状态. 这个条件,称为“可识别条件”.

称以 $p_i(x) = f(I_i(x))$ 为第 i 元件输出 1 的概率的概率网络为 B 型网络.

下面讨论在什么条件下 B-型网络是可识别的. 为此引入下面条件.

若对训练样本集 $K, \forall i, K^1(i)$ 与 $K^{-1}(i)$ 均是线性可分的(即存在超平面使 $K^1(i), K^{-1}(i)$ 分属超平面的两侧), 简称为 K 是线性可分的.

命题 2. B-型网络 N 是可识别网络 $\Leftrightarrow \forall i, K^{-1}(i)$ 与 $K^1(i)$ 是线性可分的.

证明: 因 f 是 $(-\infty, \infty) \rightarrow [0, 1]$ 的单调非减函数.

设 $x_* = \sup\{x | f(x) = 0\}$ $x^* = \inf\{x | f(x) = 1\}$

显然, $x_* \leq x^*$

设训练样本集 $K = \{r^1, r^2, \dots, r^m\}$, 因 N 是可识别的, 则 $\forall i, \forall x' \in K^1(i) (x' \in K^{-1}(i))$
有

$$I_i(x') \geq x_* \quad (I_i(x') \leq x_*)$$

即

$$\sum w_{ij} x_j' \begin{cases} \geq x_* & \text{当 } x' \in K^1(i) \\ \leq x_* & \text{当 } x' \in K^{-1}(i) \end{cases}$$

取 $\alpha, x_* \leq \alpha \leq x^*$

作超平面 $\Pi_i: \sum w_{ij}x_j - \alpha = 0.$

显然, Π_i 分离 $K^1(i)$ 与 $K^{-1}(i).$

反之, 若训练样本集 K 是线性可分的, 则容易证明存在单调非减函数 f , 当取 $p_i(x) = f(I_i(x) - \alpha_i)$ 为第 i 元件输出 1 的概率时, 对应的 B 型网络是可识别网络.

推论 1. 若 BM 是可识别的, 则其训练样本集是线性可分的.

换句话说, 当训练样本不是线性可分时, 其 B 型网络不是可识别的, 即不是所有训练样本都是稳定状态.

例: 设 B 型网络 N 有 4 个元件, 训练样本集 K

$$K = \{(-1, 1, -1, 1), (1, -1, 1, -1), (-1, -1, -1, 1), (1, 1, 1, -1)\}$$

$$\text{设对某一个 } i, \quad K^1(i) = \{(-1, 1, -1, 1), (1, -1, 1, -1)\}$$

$$K^{-1}(i) = \{(-1, -1, -1, 1), (1, 1, 1, -1)\}$$

显然, $K^1(i)$ 与 $K^{-1}(i)$ 不是线性可分的, 故对这个训练样本集, 不存在可识别 B 型网络. 但是若用 (3) - (6) 公式定义广义 A -学习律, 得其对应网络是可识别的 (由命题 1 得).

由此可见, 用 (3) - (6) 定义的广义 A -学习律, 在这一点上 (可识别性质) 比用 (14) - (16) 定义的广义 A -学习律强, 即比 Boltzman 机强.

2 收敛半径的估计

本节讨论, 当我们引入广义 A -学习律之后, PLN 网络的容错能力大大提高, 使网络中每个吸引子 (稳定状态) 附近都形成一定半径的吸引区域. 下面我们对吸引半径进行定量估计.

以吸引子为中心的球形邻域的每个状态在网络的演化过程中最后都收敛于这个吸引子, 满足此条件的最大球的半径称为吸引半径.

下面讨论以 (8) 式定义的广义 A -学习律的网络的吸引子的吸引半径.

令 $r \in K$, 固定 r , 设 $r \in K^*(i)$ (其中 $K^*(i)$ 为 $K^0(i)$ 或 $K^1(i)$, 由 $r(y_i) = 0$ 或 1 而定), $i = 1, 2, \dots, n.$

令

$$l(r, i) = \min_{s \in K/K^*(i)} \frac{1}{2} D_i(r, s)$$

将 $\{l(r, i)\}$ 从小到大排列, 不妨设为:

$$l(r, 1) \leq l(r, 2) \leq \dots \leq l(r, n)$$

令 $\varphi(r)$ 是满足下列的标号

$$l(r, j) > j, \quad j = 1, 2, \dots, \varphi(r)$$

$$l(r, \varphi(r) + 1) \leq \varphi(r) + 1$$

命题 3. $r \in K$, 则 r 的收敛半径 $\geq \varphi(r)$

证明: 设 $x \in S(r, \varphi(r))$

其中 $S(r, \varphi(r))$ 表示以 r 为中心 $\varphi(r)$ 为半径的球.

令

$$C_i(x) = \frac{D_i^0(x)}{(D_i^0(x) + D_i^1(x))}$$

则由 $\varphi(r)$ 定义得, 当 $i \geq \varphi(r)$ 时有

$$l(r, i) \geq l(r, \varphi(r)) > \varphi(r)$$

$$\min_{s \in K/K^* i} D_i(r, s) > 2\varphi(r)$$

因为 $x \in S(r, \varphi(r))$ 得 $D_i(x, r) \leq \varphi(r)$

故

$$C_i(x(t)) = \begin{cases} < \frac{1}{2} & r(y_i) = 0 \\ > \frac{1}{2} & r(y_i) = 1 \end{cases}$$

由(8)式知

$$y_i(t+1) = \begin{cases} 0 & r(y_i) = 0 \\ 1 & r(y_i) = 1, \quad i = \varphi(r), \varphi(r) + 1, \dots, n. \end{cases}$$

故 $y_i(t+1)$ 与 $r(y)$ 的海明距离 $\leq \varphi(r) - 1$

于是有 $x(t+1) \in S(r, \varphi(r) - 1)$

当 $i \geq \varphi(r) - 1$ 时, 同理可得

$$y_i(t+2) = \begin{cases} 0 & r(y_i) = 0 \\ 1 & r(y_i) = 1 \end{cases}$$

得 $y(t+2)$ 与 $r(y)$ 的海明距离 $\leq \varphi(r) - 2$, 归纳得

$$x(t + \varphi(r)) \in S(r, \varphi(r) - \varphi(r)) = S(r, 0) = r$$

即

$$x(t + \varphi(r)) = r$$

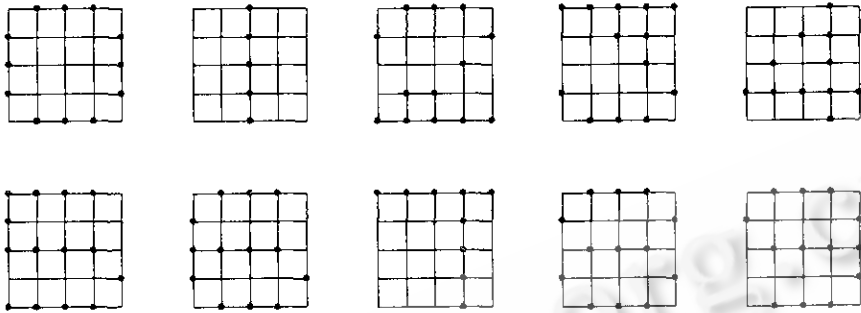
即至多经过 $\varphi(r)$ 步, x 必收敛于 r , 故得 r 的吸引半径 $\geq \varphi(r)$, 命题证毕.

命题 3 说明, 引入广义 A-学习律后, 使 PLN 网络的容错能力大为提高. 下一节我们将广义 A-学习律应用于联想记忆, 计算机模拟结果说明当适当选取参数 k , 可使网络有良好的容错能力和很快的收敛速度.

3 模拟实验结果

为了说明广义 PLN 网络的性能, 我们作了如下计算机模拟实验.

训练样本是 0-9 的标准数字,以 5×5 点阵表示,如下图所示.



PLN 网络共有 25 个输入端, x_0, \dots, x_{24} , 神经元节点 A_0, \dots, A_{24} , 各节点对应的点阵中点如下图所示.

0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
20	21	22	23	24

反馈连接 $H = \{(i, i) | i = 0, 1, \dots, 24\}$

各节点输入是

$$I(0) = I(1) = \dots = I(4) = \{x_0, \dots, x_9\}$$

$$I(5) = I(6) = \dots = I(9) = \{x_5, \dots, x_{14}\}$$

$$I(10) = I(11) = \dots = I(14) = \{x_5, \dots, x_{14}, x_{15}, \dots, x_{19}\}$$

$$I(16) = I(17) = \dots = I(19) = \{x_{10}, \dots, x_{19}\}$$

$$I(21) = I(22) = \dots = I(24) = \{x_{15}, \dots, x_{24}\}$$

$$I(15) = I(16) \cup \{x_0, x_4\}$$

$$I(20) = I(21) \cup \{x_0, x_4\}$$

广义 A-学习律采用公式(4)-(6),为了说明网络的容错能力测试样本比标准训练样本具有多余的黑点(即噪音), q 表示多余的黑点数目,每次重复运行 100 次.表中所列以数字 '0' 为训练样本,在不同 k 值下收敛到 0-9 各数字的次数以及平均收敛步数.

结果表明,当 k 值选择适当时,可以明显提高网络的容错能力和收敛速度.

$q=1$

	0	1	2	3	4	5	6	7	8	9	平均收敛步数
$k=1$	25	6	3	4	3	14	14	6	11	14	158.35
$k=2$	95	0	0	0	0	0	1	1	2	0	3.66
$k=3$	100	0	0	0	0	0	0	0	0	0	1.71

$q=2$											
	0	1	2	3	4	5	6	7	8	9	平均收敛步数
$k=1$	26	5	2	3	3	15	18	3	13	12	161.64
$k=2$	85	5	0	0	1	2	3	0	2	1	8.06
$k=3$	99	0	1	0	0	0	0	0	0	0	1.70

$q=3$											
	0	1	2	3	4	5	6	7	8	9	平均收敛步数
$k=1$	15	8	3	3	4	22	18	3	12	12	167.72
$k=2$	55	6	3	4	1	2	11	1	10	7	13.72
$k=3$	88	2	0	1	0	1	3	0	3	1	3.61
$k=4$	97	1	0	0	0	0	0	0	1	1	2.13

参考文献

- 1 Zhang Bo, Zhang Ling, Zhang Huai. A quantitative analysis of the behaviors of the PLN network. *Neural Networks*, 1992;5(4):639-644.
- 2 张钹,张铃. 概率逻辑神经网络收敛性的分析. *计算机学报*, 1993;16(1):1-12.
- 3 张铃,张钹. 论概率逻辑神经网络(I)—结构性定理. *模式识别与人工智能*, 1991;4(3):1-10.
- 4 张铃,张钹. 论概率逻辑神经网络(II)—Aleksander 算法性质. *模式识别与人工智能*, 1992;5(2):90-96.

THE IMPROVEMENT OF PLN NETWORK AND ITS APPLICATIONS

Zhang Bo

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

Zhang Ling

(Department of Computer Science, Anhui University, Hefei 230039)

(Department of Computer Science, Zhejiang University, Hangzhou 310027)

Abstract In light of the deficiency of the original model of Probabilistic Logic Neuron (PLN) network, this paper presents a revised model and shows that its robustness and speed are improved. The authors also indicate that the original model of PLN network, the Hopfield's learning rule and the evolution rule of Boltzman machine are special cases of the new model. Finally, the application of the generalized PLN network to associative memory is given. The computer simulation results show that robustness and convergence speed of the new model are much better than the original one.

Key words Probabilistic Logic Neuron (PLN) network, A-learning rule, generalized A-learning rule, associative memory.