文章编号: 1001-0920(2005)11-1235-06

核化的自联想反馈神经网络

胡恺君[1], 吴锡生[2], 王士同[1,2]
(1. 江南大学 信息工程学院, 江苏 无锡 214036; 2. 南京理工大学 计算机系, 南京 210016)

摘 要: 将Mercer的核思想与自联想反馈神经网络的理论相结合, 提出了核化的自联想反馈神经网络KARN, 它是对多值自联想反馈神经网络MREM的有效扩展. 实验结果证明, 相对于自联想反馈神经网络, 核化的自联想反馈神经网络的主要优势在于有效扩大了网络的容量.

关键词: 核函数; 自联想; Hopfield神经网络; 多值反馈神经网络
中图分类号: TP273 **文献标识码**: A

Kernelized Associative Autonomous Recurrent Network

HU Kai-jun[1], *WU Xi-sheng*[1,2], *WANG Shi-tong*[1,2]
(1. School of Information Engineering, Southern Yangtze University, Wuxi 214036, China; 2. Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210016, China. Correspondent: WANG Shi-tong, E-mail: wxwangst@yahoo.com.cn)

Abstract: The kernelized autoassociative recurrent neural network (KARN) is presented, which is an extension of the autoassociative recurrent neural network MREM. Compared with MREM, the major advantage of KARN is the enlarged memory capacity. Experimental results manifest this advantage.

Key words: Kernel function; Autoassociation; Hopfield network; Multivalued recurrent network

1 引 言

Hopfield[1,2]于1982年首先提出了离散型Hopfield网络, 因其简单和快速收敛的特性, 该网络在各个领域广泛运用并快速发展. Casermeiro等[3]提出了多值自联想反馈神经网络(MREM), 它是对Hopfield神经网络的扩展, 使其不仅能像Hopfield网络那样存储二进制样本和两极样本, 而且能够存储多值样本, 如red, green, blue. MREM中神经元之间相互作用通过相似函数来体现, 其容量与传统的Hopfield神经网络相差不多. 如何有效地扩大MREM的存储容量是值得研究的问题.

为了进一步提高MREM的容量, 本文提出了核化的多值自联想反馈神经网络(KARN), 该网络利用核函数把维数较低的样本映射到高维空间, 再通过能量函数来计算经过核化的多值网络的容量. 理论与实验结果表明, 核化后的KARN与MREM相比, 能够有效地提高网络的容量, 并且在收敛速度方面也有所改善.

2 多值自联想反馈神经网络

令H是由P个结点组成的反馈神经神经网络, 其中每个神经元的状态为V_i, $i \in I$, $I = \{1, 2, \cdots, P\}$, 其取值都属于集合M, M中的元素可以是任意值, 如$M = \{\text{black}, \text{white}, \text{red}\}$. 网络在$k$时刻的状态可以这样给出: $V(t) = (V_1(t), V_2(t), \cdots, V_P(t)) \in M^P$, 这个网络是完全连接的, 并且$\omega_{ij}$是从$j$节点到$i$节点的连接权值. 网络的能量函数如下:

$$E = -\frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\omega_{ij} f(V_i, V_j) + \sum_{i=1}^{P}\theta_i(V_i). \quad (1)$$

收稿日期: 2004-11-17; 修回日期: 2005-04-04.
基金项目: 教育部跨世纪优秀人才支持计划项目(NCET-04-0496); 江苏省自然科学基金项目(BK2003017); 中科院软件所计算机科学重点实验室开放课题(SYSKF0406); 中科院自动化研究所模式识别国家重点实验室开放课题.
作者简介: 胡恺君(1980—), 男, 江苏无锡人, 硕士生, 从事计算机应用等研究; 王士同(1964—), 男, 江苏扬州人, 教授, 博士生导师, 主要从事人工智能、神经网络等研究.

其中: $i, j \in I$; $f: M \times M \to R$, $\theta: M \to R$, R 为实数; $\omega_{ij} = \sum_{k \in K} f(x_{ki}, x_{kj})$; 函数 $f(V_i, V_j)$ 是一个相似函数, 表示第 i 个和第 j 个神经元的相似程度. 最简单的相似函数是

$$f(x, y) = \begin{cases} 1, x = y; \\ 0, x \neq y. \end{cases}$$

或

$$f(x, y) = \begin{cases} 1, x = y; \\ -1, x \neq y. \end{cases}$$

$$\forall x, y \in M. \tag{2}$$

MREM 的动态调整是在离散时刻内异步进行的, 即在时刻 t 内, 只有一个神经元的状态被改变. 在 t 时刻, 计算所有神经元能量函数的值, 取其最小值作为需要更改状态的神经元. 设 t 时刻第 α 个神经元是能量最小的神经元, 对于所有其他神经元($i \neq \alpha$), $t+1$ 时刻的状态都与 t 时刻一样($V_i(t+1) = V_i(t)$, $\forall i \neq \alpha$). 为了决定第 α 个神经元在第 $t+1$ 时刻的状态, 定义第 α 个神经元在 t 时刻的潜在增加函数

$$U_{\alpha, l}(t) = \frac{1}{2} \sum_{i=1}^{P} [\omega_{\alpha i} F_{l, V_i(t)} + \omega_{i\alpha} F_{V_i(t), l}] - \frac{1}{2} \omega_{\alpha\alpha} F_{ll}. \tag{3}$$

这样潜在增加函数是一个向量 $U_\alpha = (U_{\alpha,1}, U_{\alpha,2}, \cdots, U_{\alpha,N})$, 它有 N 个元素 $U_{\alpha,l}(t)$ 由能量函数演化而来, 于是使用下面的神经元状态更新规则:

$$V_\alpha(t+1) = l \Leftrightarrow U_{\alpha,l}(t) \geq U_{\alpha,i}(t),$$

$$\forall l \in M. \tag{4}$$

要取得能量函数的最小值, 即要取得潜在增加函数的最大值, 规定如果第 α 个神经元与 M 集合中 l 的潜在增加函数的值为最大, 那么第 α 个神经元 $t+1$ 时刻的状态为 l; 如果在 M 集合中有几个元素与第 α 个神经元的潜在函数同样大, 则取这几个元素中最前的一个元素值为第 α 个神经元 $t+1$ 时刻的状态.

考虑到相似函数中 $F_{v_i(t), l}$ 与 $F_{l, v_i(t)}$ 的值是一样的, $F_{ll} = 1$ 和 $\frac{1}{2}\omega_{\alpha\alpha}$ 是共有的, 将其省略, 所以式(3) 最终可以写成

$$U_{\alpha, l}(t) = \frac{1}{2} \sum_{i=1}^{P} (\omega_{\alpha i} + \omega_{i\alpha}) F_{V_i(t), l}. \tag{5}$$

同 Hopfield 神经网络的分析[4] 相似, 多值反馈神经网络的存储容量也是用估计网络产生错误概率的方法来计算. 文献[3] 通过对能量函数的讨论, 得到容量大小, 也就是样本数目与样本维数之比, 即 K/P 的比例. MREM 模型已被证明可以动态收敛于一个极小值状态[5,6]. 通过研究发现, MREM 的样本容量与传统 Hopfield 神经网络的样本容量大小相似, 为了解决这个问题, 本文尝试通过引入核函数和核化概念来扩大网络的容量.

3 核化的自联想多值反馈神经网络

对于多值反馈神经网络的算法, 基于核思想对其进行改进, 将满足 Mercer 条件的核函数代替两向量间的内积运算, 把向量映射到高维空间, 从而扩大网络的存储容量.

3.1 对潜在增加函数核化

注意到式(5) 是将向量 $\omega_{\alpha i}$ 和向量 $F_{V_i(t), l}$ 做内积, 于是根据核化的思想将式(5) 核化, 得到核化的潜在增加函数.

多项式核函数 $K(X \cdot Y) = (X \cdot Y^T + c)^m$, X 与 Y 为做内积向量, c 和 m 为常量.

径向基核函数 $K(X \cdot Y) = \exp(-\|X - Y\|^2 / 2\sigma_1^2)$, X 与 Y 为做内积向量, σ_1 为常量.

将式(5) 核化为:

$$U'_{\alpha, l}(t) = \frac{1}{2}((\omega_\alpha \cdot F^T_{V(t), l}) + c)^m, \tag{6}$$

$$U'_{\alpha, l}(t) = \frac{1}{2}\exp(-\|\omega_\alpha - F_{V(t), l}\|^2 / 2\sigma_1^2). \tag{7}$$

3.2 KARN 权值

KARN 的权值计算与 Hopfield 神经网络的权值计算相似, 都是求样本向量的外积, 再把 K 个样本通过外积得到的矩阵求和. 这里使用相似函数

$$\omega_{ij} = \sum_{k \in K} f(x_{ki}, x_{kj}), \tag{8}$$

其中 x_{ki} 和 x_{kj} 是第 k 个样本的第 i 个和第 j 个神经元.

4 KARN 的稳定点分析

在 Hopfield 神经网络中, 样本向量的相反向量也是网络的稳定点, 但这个稳定点却是不需要的. 为了不使网络产生其他不需要的稳定点, 将初始样本扩充 N 个 M 集合中的元素.

初始样本 $X = \{x_1, x_2, \cdots, x_P\}$, $M = (m_1, m_2, \cdots, m_N)$, $X' = \{x_1, x_2, \cdots, x_P, m_1, m_2, \cdots, m_N\}$. 例如两极样本向量 $X = \{-1, 1, 1, -1\}$, $M = (-1, 1)$, $X' = (-1, 1, 1, -1, -1, 1)$, $-X' = (1, -1, -1, 1, 1, -1)$.

通过这样的变化, 相反向量和原来的样本向量在计算其权矩阵时, 得到的矩阵是不一致的, 从而有效地避免了网络最后稳定到错误的样本上.

5 KARN 的容量分析

多值反馈神经神经网络的能量函数为

$$E = -\frac{1}{2} \sum_{i=1}^{P+N} \sum_{j=1}^{P+N} \sum_{k=1}^{K} f(X_{ki}, X_{kj}) f(V_i, V_j). \tag{9}$$

考虑 k 个样本向量被引入网络, 假设向量 V^* 与输入的样本向量一致, 而 V^λ 与 V^* 之间只有第 λ 个神经元不一致, 其他的神经元都一致. 定义

$$D = 2\Delta E = 2(E_{V^\lambda} - E_{V^*}). \tag{10}$$

如果 $D > 0$, 则说明 V^λ 加入网络后网络没有产生错误, 当 V^λ 这个含有噪声的向量被还原时, 能正确还原出其初始状态. 可根据式(9) 将式(10) 改写为

$$D = \sum_{i=1}^{P+N}\sum_{j=1}^{P+N}\sum_{k=1}^{K} f(X_{ki}, X_{kj}) f(V_1^*, V_i^*) - \sum_{i=1}^{P+N}\sum_{j=1}^{P+N}\sum_{k=1}^{K} f(X_{ki}, X_{kj}) f(V_i^\lambda, V_j^\lambda). \tag{11}$$

当 $j > 1$ 时, 定义 $V_j^* = V_j^\lambda$, 即 $V_1^* \neq V_1^\lambda$. 这样可将式(11) 再次简化为

$$D = 2\sum_{i=2}^{P+N}\sum_{k=1}^{K} f(X_{ki}, X_{kj})(f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)). \tag{12}$$

用与样本向量 X_{k0} 一致的 V^* 来替换 X_{k0}, 则有

$$D = 2\sum_{i=2}^{P+N} f(V_1^*, V_i^*)(f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)) + 2\sum_{i=2}^{P+N}\sum_{k\neq k0}^{K} f(X_{k1}, X_{k0})(f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)). \tag{13}$$

5.1　使用核函数进行核化

为了提高网络的样本容量, 根据本文提出的核化思想, 使用上述多项式核函数和径向基核函数对式(13) 进行核化. 令

$$S = \{f(V_1^*, V_1^*), f(V_1^*, V_2^*), \cdots, f(V_1^*, V_{N+P}^*)\},$$
$$S_i = f(V_1^*, V_i^*); \tag{14}$$
$$T = \{f(V_1^\lambda, V_1^\lambda), f(V_1^\lambda, V_2^\lambda), \cdots, f(V_1^\lambda, V_{N+P}^\lambda)\},$$
$$T_i = f(V_1^\lambda, V_i^\lambda); \tag{15}$$
$$Q = \{f(X_{k1}, X_{k1}), f(X_{k1}, X_{K2}), \cdots, f(X_{k1}, X_{K(N+P)})\},$$
$$Q_i = f(X_{k1}, X_{Ki}). \tag{16}$$

根据式(14)～(16) 可将式(13) 核化为

$$D = 2(K(S,S) - K(S,T) + 2K(Q,(S-T))). \tag{17}$$

其中 k 为核函数.

5.2　多项式核函数的计算

5.2.1　K(S, S) 的计算

$$K(S,S) = (S \cdot S^T + c)^m = (N + P - 1 + c)^m,$$
$$\sum_{i=2}^{P+N} f(V_1^*, V_i^*) f(V_1^*, V_i^*) = N + P - 1. \tag{18}$$

5.2.2　K(S, T) 的计算

$$K(S,T) = (S \cdot T^T + c)^m.$$

首先计算 $\sum_{i=P+1}^{P+N} f(V_1^*, V_i^*)(f(V_1^\lambda, V_i^\lambda), i \in (P+1, P+N)$, 扩充元素为 $\{m_1, m_2, \cdots, m_N\}$, 这里存在 $m_p \neq m_q, p, q \in \{1, 2, \cdots, N\}$. 如果 $V_1^* = m_p, V_1^\lambda = m_q$, 则当 $i = P + p$ 或 $i = P + q$ 时, $f(V_1^*, V_i^*) f(V_1^\lambda, V_i^\lambda) = -1$; 否则, $f(V_1^*, V_i^*) f(V_1^\lambda, V_i^\lambda) = 1$. 所以可得

$$\sum_{i=P+1}^{P+N} f(V_1^*, V_i^*)(f(V_1^\lambda, V_i^\lambda)) = (N-2)\times 1 - 2 = N - 4,$$
$$2 \le i \le P.$$

考虑随机样本, 因为样本中每个元素出现的几率相等, 所以定义 $f(V_1^*, V_i^*) f(V_1^\lambda, V_i^\lambda)$ 是一个随机值 ϕ. $f(V_1^*, V_i^*) f(V_1^\lambda, V_i^\lambda)$ 的概率分布如表 1 所示, 根据表 1, 可以计算出当 $\phi = 1$ 或 $\phi = -1$ 时的概率.

表 1　$f(V_1^*, V_i^*) f(V_1^\lambda, V_i^\lambda)$ 的概率分布

第 1 项	第 2 项	概率	值
$V_1^* = V_i^*$	$V_1^\lambda = V_i^\lambda$	0	1
$V_1^* = V_i^*$	$V_1^\lambda \neq V_i^\lambda$	$1/N$	-1
$V_1^* \neq V_i^*$	$V_1^\lambda = V_i^\lambda$	$1/N$	-1
$V_1^* \neq V_i^*$	$V_1^\lambda \neq V_i^\lambda$	$(N-2)/N$	1

因为 $V_1^* \neq V_i^*, V_1^\lambda \neq V_i^\lambda, i > 1$, 而且是独立分布, 所以 $p(V_1^* = V_i^*, V_1^\lambda = V_i^\lambda) = p(V_1^* = V_i^*) \times p(V_1^\lambda = V_i^\lambda) = 1/N$. 根据表 1 得到 $p(\phi = -1) = 2/N$, $p(\phi = 1) = (N-2)/N$, 由此得到 ϕ 的数学期望和方差分别为 $E(\phi) = (N-4)/N$. $V(\phi) = 8(N-4)/N^2$. 所以

$$\sum_{i=2}^{P+N} f(V_1^*, V_i^*) f(V_1^\lambda, V_i^\lambda) = N - 4 + \sum_{i=2}^{P} \phi_i,$$

由此可得

$$K(S,T) = \left(N - 4 + \sum_{i=2}^{P} \phi_i + c\right)^m. \tag{19}$$

5.2.3　K(Q, (S - T)) 的计算

$$K(Q,(S-T)) = (Q \cdot (S-T)^T + c)^m.$$

首先令 $A = f(X_{k1}, X_{ki})(f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)), i \in (2, P)$, 注意到 A 是一个随机变量为 ϵ 独立分布, 可以得到 A 的概率分布如表 2 所示.

可以写出 A 的概率分布为 $p(\epsilon = 2) = p(\epsilon = -2) = 1/N$, $p(\epsilon = 0) = (N-2)/N$. 通过计算得到数学期望与方差为 $E(\epsilon) = 0, V(\epsilon) = 8/N$, 所以有

$$\sum_{i=2}^{P}\sum_{k\neq k_0}^{K} f(X_{k_1}, X_{k_i})(f(V_1^*, V_i^*)) - f(V_1^\lambda, V_i^\lambda) = \sum_{i=1}^{(P-1)(K-1)} \epsilon_i, \; i \in (2, p).$$

表 2 A 的概率分布

第 1 项	第 2 项	第 3 项	概 率	值
$X_{k_1}=X_{k_j}$	$V_1^*=V_i^*$	$V_1^\lambda=V_i^\lambda$	0	0
$X_{k_1}=X_{k_j}$	$V_1^*=V_i^*$	$V_1^\lambda\neq V_i^\lambda$	$1/N^2$	2
$X_{k_1}=X_{k_j}$	$V_1^*\neq V_i^*$	$V_1^\lambda=V_i^\lambda$	$1/N^2$	-2
$X_{k_1}=X_{k_j}$	$V_1^*\neq V_i^*$	$V_1^\lambda\neq V_i^\lambda$	$(N-2)/N^2$	0
$X_{k_1}\neq X_{k_j}$	$V_1^*=V_i^*$	$V_1^\lambda=V_i^\lambda$	0	0
$X_{k_1}\neq X_{k_j}$	$V_1^*=V_i^*$	$V_1^\lambda\neq V_i^\lambda$	$(N-1)/N^2$	2
$X_{k_1}\neq X_{k_j}$	$V_1^*\neq V_i^*$	$V_1^\lambda=V_i^\lambda$	$(N-1)/N^2$	-2
$X_{k_1}\neq X_{k_j}$	$V_1^*\neq V_i^*$	$V_1^\lambda\neq V_i^\lambda$	$(N-1)(N-2)/N^2$	0

令

$$B = f(X_{k_1},X_{k_i})(f(V_1^*,V_i^*) - f(V_1^\lambda,V_i^\lambda)),$$
$$i\in(P+1,P+N),$$

则 B 也是一个随机变量为 ϵ 的独立分布函数，限于篇幅，这里直接给出结果

$$B = \sum_{i=(P-1)(K-1)+1}^{P(K-1)} \epsilon_i.$$

当 $i\in(P+1,P+N)$ 时，B 的值与 $i\in(2,P)$ 时一样，且属于同一个分布，数学期望和方差一致．结合 A 与 B，可以得到

$$K(\mathrm{S},\mathrm{T}) = (\mathrm{S}\cdot\mathrm{T}^{\mathrm{T}}+c)^m = \Big(\sum_{i=1}^{P(K-1)}\epsilon_i + c\Big)^m. \tag{20}$$

5.3 对于使用多项式核函数后容量的讨论

通过式(18)～(20)得到式(19)的值为

$$D = 2((N+P-1+c)^m - (N-4+c+\sum_{i=2}^{P}\phi_i)^m + \Big(\sum_{i=1}^{P(K-1)}(\epsilon_i+c)^m\Big). \tag{21}$$

根据中心极限定理，当 P 和 K (样本个数) 很大时，$\sum_{i=1}^{P(K-1)}\epsilon_i$ 变成一个高斯分布，期望为 0，方差为 $8P(K-1)/N$．$\sum_{i=2}^{P}\phi_i$ 也是一个高斯分布，期望为 $(P-1)(N-4)/N$，方差为 $8(P-1)(N-2)/N^2$．网络中错误的概率为

$$P_e = \mathrm{prob}(D<0) =$$
$$\mathrm{prob}\Big((N+P-1+c)^m - (N-4+c+\sum_{i=2}^{P}\phi_i)^m + \Big(\sum_{i=1}^{P(K-1)}\epsilon_i + c\Big)^m < 0\Big),$$

此时可以代入 m 值来计算错误率．为了讨论方便，不妨设 $m=1$，此时

$$\mathrm{prob}\Big(N+P-1+c-(N-4+c+\sum_{i=2}^{P}\phi_i) + \sum_{i=1}^{P(K-1)}\epsilon_i + c < 0\Big) =$$
$$\mathrm{prob}\Big(P+3+c<\sum_{i=2}^{P}\phi_i - \sum_{i=1}^{P(K-1)}\epsilon_i\Big)\Rightarrow$$
$$\mathrm{prob} = (P+3+c<\varphi).$$

其中 φ 为 $\sum_{i=2}^{P}\phi_i$ 和 $\sum_{i=1}^{P(K-1)}\epsilon_i$ 叠加后的一个分布，因为二者都是高斯分布，所以

$$\varphi\to N(\mu,\sigma),\ \mu = \frac{(P-1)(N-4)}{N},$$
$$\sigma = \sqrt{\frac{8(P-1)(N-2)}{N^2} + \frac{8P(K-1)}{N}}.$$

将 φ 变为标准正态分布

$$Z = \frac{P+3+c-\mu}{\sigma}\Rightarrow$$
$$\frac{P+3+c-\dfrac{(P-1)(N-4)}{N}}{\sqrt{\dfrac{8(P-1)(N-2)}{N^2}+\dfrac{8P(\alpha P-1)}{N}}} = Z_\alpha\Rightarrow$$
$$\alpha = \frac{1}{P} + \frac{(A^2/Z_\alpha^2 - B)}{PC}. \tag{22}$$

其中

$$A = P+3+c-(P-1)(N-4)/N,$$
$$B = 8(P-1)(N-2)/N^2,$$
$$C = -8P/N,$$

α 是网络的容量 $\alpha=K/P$；Z_α 一般取 1.645．可以通过式(22)得到 α．当 $m=2$ 或 2 以上时，式(21)将很复杂，受篇幅限制，这里不展开讨论．

5.4 径向基核函数的计算

5.4.1 $K(\mathrm{S},\mathrm{S})$ 的计算

$$K(\mathrm{S},\mathrm{S}) = \exp(-\|\mathrm{S}-\mathrm{S}\|^2/2\sigma_1^2) = 1. \tag{23}$$

5.4.2 $K(\mathrm{S},\mathrm{T})$ 的计算

$$K(\mathrm{S},\mathrm{T}) = \exp(-\|\mathrm{S}-\mathrm{T}\|^2/2\sigma_1^2).$$

首先计算

$$\sum_{i=2}^{P}(f(V_1^*,V_i^*) - f(V_1^\lambda,V_i^\lambda))^2,$$

令 $R = f(V_1^*,V_i^*) - f(V_1^\lambda,V_i^\lambda)$，设随机变量为 ϕ，R 是一个随机变量为 ϕ 的独立分布．于是可以得到 ϕ 的概率，当 $2\leq i\leq P$ 时，根据表 3 的 R 概率分布得到 ϕ 的概率

$$P(\phi=2) = P(\phi=-2) = 1/N,$$
$$P(\phi=0) = (N-2)/N.$$

表 3 R 的概率分布

第 1 项	第 2 项	概率	值
$V_1^*=V_i^*$	$V_1^\lambda=V_i^\lambda$	0	0
$V_1^*=V_i^*$	$V_1^\lambda\neq V_i^\lambda$	$1/N$	2
$V_1^*\neq V_i^*$	$V_1^\lambda=V_i^\lambda$	$1/N$	-2
$V_1^*\neq V_i^*$	$V_1^\lambda\neq V_i^\lambda$	$(N-2)/N$	0

令 $Z = R^2$, 可以得到 Z 的概率, Z 也可写成随机变量为 ϕ 的独立分布, 即 $p(\phi = 4) = 2/N$, $p(\phi = 0) = (N - 2)/N$. 于是得到 Z 的数学期望和方差分别为 $E(\phi) = 8/N$, $V(\phi) = (32N - 64)/N^2$. 当 $2 \le i \le P$ 时,

$$\sum_{i=2}^{P}(f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda))^2 = \sum_{i=2}^{P}\phi_i$$

然后计算当 $P + 1 \le i \le P + N$ 时, $\sum_{i=P+1}^{P+N}(f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda))^2$ 的值. 注意到扩充元素为 $\{m_1, m_2, \cdots, m_N\}$, 存在 $m_p \ne m_q$, $p, q \in \{1, 2, \cdots, N\}$. 如果 $V_1^* = m_p$, $V_1^\lambda = m_q$, 则当 $i = P + p$ 时,

$$f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda) = 2,$$

当 $i = P + q$ 时

$$f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda) = -2,$$

否则

$$f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda) = 0,$$

$$\sum_{i=P+1}^{P+N}(f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda))^2 = 8,$$

$$K(\mathrm{S}, \mathrm{T}) = \exp\left(-\left(\sum_{i=2}^{P}\phi_i + 8\right)\Big/2\sigma_1^2\right). \quad (24)$$

5.4.3 K(Q, (S - T)) 的计算

首先计算当 $i \in (2, P)$ 时, $\sum_{k \ne k_0}^{K}\sum_{i=2}^{P}(f(X_{k_1}, X_{k_i}) - (f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)))^2$ 的值. 令

$$J = f(X_{k_1}, X_{k_i}) - (f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)),$$

可知 J 是一个随即变量为 ϵ 的独立分布. 限于篇幅, 直接给出结果

$$p(\epsilon = 1) = (2N - 3)/N^2,$$
$$p(\epsilon = -1) = ((N - 1)(N - 2) + 1)/N^2,$$
$$p(\epsilon = -3) = (N - 1)/N^2,$$
$$p(\epsilon = 3) = 1/N^2.$$

令 $U = J^2$, 则 U 的数学期望和方差为 $E(\epsilon) = 8/N$, $V(\epsilon) = (N^2 + 80N - 64)/N^2$. 可以得到

$$\sum_{k \ne k_0}^{K}\sum_{i=2}^{P}(f(X_{k_1}, X_{k_i}) - (f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)))^2 = \sum_{i=1}^{(P-1)(K-1)}\epsilon_i$$

然后计算当 $i \in (P + 1, P + N)$ 时, $f(X_{k_1}, X_{k_i}) - (f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda))^2$ 的值. 令

$$I = f(X_{k_1}, X_{k_i}) - (f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)),$$

可知 I 也是一个随机变量为 ϵ 的独立分布. 同样直接给出结果

$$\sum_{k \ne k_0}^{K}\sum_{i=P+1}^{P+N}(f(X_{k_1}, X_{k_i}) - (f(V_1^*, V_i^*) - f(V_1^\lambda, V_i^\lambda)))^2 = \sum_{i=(P-1)(K-1)+1}^{P(K-1)}\epsilon_i,$$

当 $i \in (P + 1, P + N)$ 时, I 的值与 $i \in (2, P)$ 时一样, 并且属于同一个分布, 数学期望和方差一致, 故

$$K(\mathrm{Q}, (\mathrm{S} - \mathrm{T})) = \exp\left(-\sum_{i=1}^{P(K-1)}\epsilon_i\Big/2\sigma_1^2\right). \quad (25)$$

5.5 对于使用径向基核函数后容量的讨论

通过式(23)～(25)得到式(17)的值为

$$D = 1 - \exp\left(-\left(\sum_{i=2}^{P}\phi_i + 8\right)\Big/2\sigma_1^2\right) + \exp\left(-\sum_{i=1}^{P(K-1)}\epsilon_i\Big/2\sigma_1^2\right). \quad (26)$$

根据中心极限定理, 当 P 和 K 很大时, $\sum_{i=2}^{P}\phi_i$ 变成一个高斯分布, 期望为 $8(P - 1)/N$, 方差为 $(P - 1)(32N - 64)/N^2$. $\sum_{i=1}^{P(K-1)}\epsilon_i$ 也是一个高斯分布, 期望为 $P(K - 1)(N + 80)/N$, 方差为 $P(K - 1)(N^2 + 80N - 64)/N^2$. 网络中错误的概率为

$$P_e = \mathrm{prob}(D < 0) =$$
$$\mathrm{prob}\left(1 - \left(-\left(\sum_{i=2}^{P}\phi_i + 8\right)\Big/2\sigma_1^2\right) + \exp\left(-\sum_{i=1}^{P(K-1)}\epsilon_i\Big/2\sigma_1^2\right) < 0\right) \Rightarrow$$
$$\mathrm{prob}\left(1 < \exp\left(-\left(\sum_{i=2}^{P}\phi_i + 8\right)\Big/2\sigma_1^2\right) - \exp\left(-\sum_{i=1}^{P(K-1)}\epsilon_i\Big/2\sigma_1^2\right)\right). \quad (27)$$

为计算简便, 先对不等式

$$1 < \exp\left(-\left(\sum_{i=2}^{P}\phi_i + 8\right)\Big/2\sigma_1^2\right) - \exp\left(-\sum_{i=1}^{P(K-1)}\epsilon_i\Big/2\sigma_1^2\right)$$

两边取对数, 可以得到

$$\ln 1 <$$
$$\ln\left[\exp\left(-\left(\sum_{i=2}^{P}\phi_i + 8\right)\Big/2\sigma_1^2\right)\right] -$$
$$\ln\left[\exp\left(-\sum_{i=1}^{P(K-1)}\epsilon_i/2\sigma_1^2\right)\right] \Rightarrow$$
$$0 < \left(-\left(\sum_{i=2}^{P}\phi_i + 8\right)\Big/2\sigma_1^2\right) - \left(-\sum_{i=1}^{P(K-1)}\epsilon_i/2\sigma_1^2\right),$$

整理得到

$$8 < \sum_{i=1}^{P(K-1)}\epsilon_i - \sum_{i=2}^{P}\phi_i. \quad (28)$$

根据式(30), 可将式(29)改写为

$$\mathrm{prob}\left(8 < \sum_{i=1}^{P(K-1)}\epsilon_i - \sum_{i=2}^{P}\phi_i\right) = \mathrm{prob}(8 < \Psi),$$

其中 φ 为 $\sum_{i=1}^{P(K-1)} \epsilon_i$ 和 $\sum_{i=2}^{P} \phi_i$ 叠加后的一个分布. 因为二者都是高斯分布, 所以

$$\varphi \to N(\mu, \sigma),$$

$$\mu = P(K-1)\frac{N+80}{N} - (P-1)\frac{8}{N},$$

$$\sigma = \sqrt{\frac{(32N-64)(P-1)}{N^2} + \frac{P(K-1)(N^2+80N-64)}{N^2}}.$$

将 φ 变为标准正态分布

$$Z = \frac{8-\mu}{\sigma} \Rightarrow Z_\alpha =$$

$$\frac{8 - P(\alpha P - 1)\frac{N+80}{N} + (P-1)\frac{8}{N}}{\sqrt{\frac{(32N-64)(P-1)}{N^2} + \frac{P(\alpha P-1)(N^2+80N-64)}{N^2}}}. \tag{29}$$

其中: α 是网络的容量, $\alpha = K/P$; Z_α 一般取 1.645. 可以通过式(29) 得到 α.

6 实验结果与分析

式(22) 和(29) 在理论上给出了 KARN 网络在使用多项式核函数和径向基核函数情况下的容量分析, 从两式中不易直接看出 KARN 与 MREM 在容量比较上的优势, 因此通过一些样本来试验. 经过计算, 得到反馈神经网络容量的比较如表 4 所示. 错误概率 P_e 取 0.05, Z_α 取 1.645. 可以看到, 使用多项式核函数, 当 $c=0$ 时, 网络可以退化为 MREM, 当 $c=1$ 时, 网络的容量有一定的提高, 这说明 KARN 是对 MREM 的有效扩展.

表 4 反馈神经网络容量比较(P_e = 0.05)

		N	P	α
MREM		4	1 000	0.186 4
		5	5 000	0.148 1
KARN (多项式)	$c=0$	4	1 000	0.186 4
		5	5 000	0.148 1
	$c=1$	4	1 000	0.186 7
		5	5 000	0.1494
KARN (径向基)		4	1 000	0.405 5
		5	5 000	0.354 2

通过表 4 的数据比较, 可以明显地看到, 当使用核函数时, 网络的容量 α 都有不同程度的增加, 尤其是当用到径向基核函数时, 网络的容量明显增加, 这便达到了预期目的.

7 结 论

MREM 不仅能像 Hopfield 网络那样存储二进制样本和两极样本, 而且能够存储多值的样本, 但是由于其容量与传统的 Hopfield 神经网络相差不多, 所以本文将 Mercer 的核思想与 MREM 相结合, 提出了 KARN. 与 MREM 相比, KARN 将样本映射到高维空间, 有效地提高了网络的存储容量, 这样便改进了多值反馈神经在网络容量方面的特性.

参考文献(References)

[1] Hopfield J J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities[J]. *Proc of National Academy of Science*, 1982, 79(10): 2552-2558.

[2] Hopfield J J. Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons[J]. *Proc of the National Academy of Sciences*, 1984, 81(12): 3088-3092.

[3] Enrique M C, Jose M P. MREM: An Associative Autonomous Recurrent Network[J]. *J of Intelligent and Fuzzy Systems*, 2002, 12: 163-173.

[4] John Hertz, Anders Krogh, Richard G Palmer. Introduction to the Theory of Neural Computation[J]. *Lecture Notes Volume I. Addison Wesley*, 1991, 1.

[5] Casermeiro E M. *Red Neuronal Recurrente Multivaluadaparael Reconnocimiento Depatronesy la Optimizacion Combinatorial*[D]. Malaga: Dissertation University, 2000.

[6] Casermeriro E M, Munoz P J, Garcia Bernal M A. An Associative Multivalued Recurrent Network [J]. *Lecture Notes in Artificial Intelligence*, 2002, 2557: 509-518.