

具非线性映射特性的广义逆联想存储器*

陈松灿 高 航 杨国庆

(南京航空航天大学计算机系 南京 210016)

摘要 本文针对基于 Moore-Penrose 广义逆实现的联想存储模型(如 Kohonen 模型、Murakami 模型)缺乏对已存数据完全的联想回忆能力和非线性映射能力,通过在这些模型中引入一个扩展层(隐节点层)使原模型具有对已存数据的完全回忆能力和一定的非线性映射能力,通过矩阵的奇异值分解,从理论上阐明了改进模型的性能优越性.模拟结果证实了这一点.

关键词 广义逆,联想存储,神经网络,奇异值分解,非线性映射.

到目前为止,众多研究者已提出了许多有趣的联想存储模型.联想存储模型具有相似输入可回忆出相似输出的能力,原理简单,因此被广泛地应用于模式识别、数据库、容错诊断和逻辑程序设计等领域.然而这类模型无法从原有结构上使模型具有更高的联想特性,更不用说是非线性映射能力了.关于联想存储,一般有 2 种存储编码方式,即相关阵(CAM)编码法和 Moore-Penrose 广义逆编码法(GI).现已证明基于 GI 的存储性能优于基于 CAM 的存储性能.因此本文的目的是旨在提高基于 GI 编码的联想存储器的性能(如 Kohonen 模型^[1]和 Murakami 模型^[2]).Kohonen 模型对噪声甚为敏感,Murakami 通过带噪模式优化 Kohonen 模型的性能,获得了比前者优越得多的性能.但关于这方面的研究目前进展仍不大.本文力图通过在原结构中引入一个扩展层来获取更高的联想特性和非线性映射特性.从而拓展了他们的研究.

1 改进的 GI 类模型

GI 类模型的结构是二层的,即一个输入层和一个输出层.在原结构上添加一个扩展层后如图 1 所示.

原有输入层仍保持与输出层的直接相连,其连接权阵 M 用 GI 编码(即这部分为 Kohonen 模型).现设有 K 个输入输出训练数据对 (X_i, Y_i) , $X_i \in R^n$ 为输入列向量, $Y_i \in \{-1, 1\}^m$ 为输出列向量,则

$$M = YX^+ \quad (1)$$

而对于 Murakami 模型有

* 本文研究得到国家基础研究“攀登计划”基金资助.作者陈松灿,1962 年生,副教授,主要研究领域为模式识别,信息处理,神经网络.高航,1964 年生,讲师,主要研究领域为模式识别,多媒体技术.杨国庆,1949 年生,教授,主要研究领域为模式识别,管理信息系统.

本文通讯联系人:陈松灿,南京 210016,南京航空航天大学计算机系

本文 1995-10-26 收到修改稿

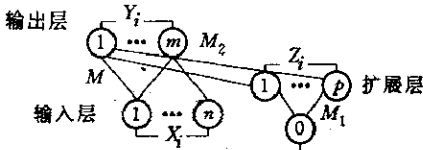


图1 添加了1个扩展层的改进模型

$$M^{Mu} = YX^T (XX^T + K\sigma^2 I_n)^+ \quad (1')$$

其中“ T ”表示矩阵或向量转置，“ $+$ ”表示广义逆。 σ^2 为输入列向量 X_i 的噪声方差。当 $\sigma=0$ 时，Murakami 模型即为 Kohonen 模型。记 $X = (X_1, X_2, \dots, X_K)$, $Y = (Y_1, Y_2, \dots, Y_K)$ 。我们期望实现

$$Y_i = f(MX_i) \quad i=1, 2, \dots, K \quad (2)$$

其中
$$f(x) = \begin{cases} 1 & \text{若 } x > 0; \\ -1 & \text{若 } x < 0; \\ \text{不变} & \text{若 } x = 0. \end{cases} \quad (3)$$

从扩展层(有 p 个节点且待定)到输出层的互连权阵用 M_2 标记,它用 GI 编码实现。若记扩展层中相应于 X_i 的输出为 Z_i , $Z = (Z_1, Z_2, \dots, Z_K)$ 则

$$M_2 = \eta Y Z^+ \quad (4)$$

其中 $\eta > 0$ 待定。

实现非线性映射和提高存储能力的关键在于从输入层到扩展层的存储编码阵 M_1 的实现,并同时保证 Z_i 有某些特殊性质。为此,我们先在输入层中添加一个 0 号节点,它仅与扩展层中的 p 个节点直接相连,权值均为 1,其输入值为 $-\|X_i\|^2/2$ 。其中 $\|\cdot\|$ 表示欧氏模。 M_1 和 Z 的设置如下:

设在 $(X_i, Y_i), i=1, 2, \dots, K$ 中满足 $Y_i \neq f(MX_i)$ 的训练数据对数为 p 。记它们为 S_p , 并设为

$$S_p = \{(X_{i_j}, Y_{i_j}), j=1, 2, \dots, p \text{ 且 } i_j \in \{1, 2, \dots, K\}\} \quad (5)$$

则 M_1 如下形成

$$M_1 = \begin{bmatrix} X_{i_1} & X_{i_2} & \dots & X_{i_p} \\ 1 & 1 & \dots & 1 \end{bmatrix} \quad (6)$$

而 Z_i 如下设置若 $(X_i, Y_i) \notin S_p$, 则 $Z_i = 0$; 若 $(X_i, Y_i) \in S_p$, 且 $i = i_k$ 则 Z_i 的第 k 个分量置为 1, 其余分量置为 0。即 $Z_{i_k} = 1$ 和 $Z_{i_j} = 0 \quad \forall j \neq k$, 从而

$$\forall i \neq j \quad Z_i^T Z_j = 0 \quad (7)$$

为使任一输入 X_i 获取满足上述特性的 Z_i , 我们引入如下函数

$$G(v) = (g(v_1), g(v_2), \dots, g(v_p))^T$$

其中
$$g(v_k) = \begin{cases} 1 & \text{若 } v_k > \frac{-d_{i_k} + \|X_{i_k}\|^2}{2} \quad k=1, 2, \dots, p; \\ 0 & \text{否则} \end{cases} \quad (8)$$

且 $0 \leq d_i < \text{Min}_{j \neq i} \{\|X_i - X_j\|^2\}, i=1, 2, \dots, K$ 。那么对于任一输入 X_i , 恒有

$$G\left(M_1 \begin{bmatrix} X_i \\ -\|X_i\|^2/2 \end{bmatrix}\right) = Z_i \quad (9)$$

事实上因为

$$M_1 \begin{bmatrix} X_i \\ -\|X_i\|^2/2 \end{bmatrix} = \begin{bmatrix} X_{i_1}^T & 1 \\ X_{i_2}^T & 1 \\ \dots & \dots \\ X_{i_p}^T & 1 \end{bmatrix} \begin{bmatrix} X_i \\ -\|X_i\|^2/2 \end{bmatrix} = \begin{bmatrix} X_{i_1}^T X_i - \|X_i\|^2/2 \\ X_{i_2}^T X_i - \|X_i\|^2/2 \\ \dots \\ X_{i_p}^T X_i - \|X_i\|^2/2 \end{bmatrix}$$

则对任意的 $k \in \{1, 2, \dots, p\}$ 有

$$g(X_i^T X_i - \frac{\|X_i\|^2}{2}) = g(\frac{2X_i^T X_i - \|X_i\|^2}{2})$$

$$= g(\frac{\|X_{i_k}\|^2 - \|X_i - X_{i_k}\|^2}{2}) \stackrel{\text{由(8)}}{=} \begin{cases} 1 & \text{若 } i = i_k; \\ 0 & \text{否则} \end{cases}$$

即得(9)式. 显然 $Z_{i_1}, Z_{i_2}, \dots, Z_{i_p}$ 都是单位向量, 因此互相正交. 从上述过程, 我们可实现对 K 个数据对 (X_i, Y_i) 的存储或记忆. 下面分析该存储模型的回忆性能. 现将 X_i 提交输入端, 输出层对该 X_i 的输入加权和记为 NET_i , 则

$$NET_i = MX_i + M_2 G(M_1 \begin{pmatrix} X_i \\ -\|X_i\|^2/2 \end{pmatrix})$$

$$= YX^+ + \eta YZ^+ Z_i \tag{10}$$

其中 Z^+ 将退化为 Z^T . 事实上从广义逆定义有

$$Z^+ = Z^T (ZZ^T)^+ \text{ 且秩}(Z) = p = \text{秩}(Z^+)$$

故

$$ZZ^T = \sum_{i=1}^K Z_i Z_i^T = \sum_{j=1}^p Z_{i_j} Z_{i_j}^T = I_p$$

从而 $(ZZ^T)^+ = I_p$, 由(10)得

$$NET_i = \begin{cases} YX^+ X_i + \eta Y_i & \text{若 } i \in \{i_1, i_2, \dots, i_p\}; \\ YX^+ X_i & \text{否则} \end{cases} \tag{11}$$

由构造知当 $i \in \{i_1, i_2, \dots, i_p\}$ 时, 就有 $Y_i = f(MX_i)$ 成立. 因此对于任何 $i \in \{i_1, i_2, \dots, i_p\}$, 能从 X_i 回忆出 Y_i . 若 $i \in \{i_1, i_2, \dots, i_p\}$ 则

$$NET_i = YX^+ X_i + \eta Y_i \tag{12}$$

为分析(12)式的性能, 我们将利用矩阵的奇异值分解. 设 X 有如下的 SVD 分解

$$X = \sum_{i=1}^l \lambda_i^{-\frac{1}{2}} p_i q_i^T \tag{13}$$

其中 l 是 X 的秩; p_i, q_i 分别是 n 维和 K 维正交规范列向量; λ_i 为 $X^T X$ 或 XX^T 的特征值, 即满足

$$XX^T p_i = \lambda_i p_i \text{ 和 } X^T X q_i = \lambda_i q_i, \quad i = 1, 2, \dots, l$$

且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l > 0$.

设 $NET_i = (net_{i1}, net_{i2}, \dots, net_{im})^T$

$u_{ij} = (YX^+ X_i)_j$ 表示向量 $YX^+ X_i$ 的第 j 个分量;

$v_{ij} = \eta y_{ij}$ $j = 1, 2, \dots, m; i = 1, 2, \dots, K$

$\{p_{l+1}, \dots, p_n\}$ 为 $\{p_1, p_2, \dots, p_l\}$ 的正交补基. 由于 $X_i \in R^n$, 故可由 $\{p_1, p_2, \dots, p_n\}$ 线性表示. 设

$$X_i = \sum_{j=1}^n a_{ij} p_j \tag{14}$$

故由 $X^+ = \sum_{i=1}^l \lambda_i^{-\frac{1}{2}} q_i p_i^T$ 得 $u_{ij} = ((Y_1, Y_2, \dots, Y_K) (\sum_{i=1}^l \lambda_i^{-\frac{1}{2}} q_i p_i^T) (\sum_{j=1}^n a_{ij} p_j))_j$

$$= y_{ij} \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kj} + \sum_{r \neq j} y_{rj} \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kr} \tag{15}$$

其中 q_{kj} 为 q_k 的第 j 个分量, $j = 1, 2, \dots, K$. 因此

$$net_{ij} = u_{ij} + v_{ij} = (\eta + \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kj}) y_{ij} + \sum_{r \neq j} y_{rj} \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kr} \quad (16)$$

$i=1, 2, \dots, K, j=1, 2, \dots, m.$

只要使 η 满足下式

$$\eta > \sum_{r \neq j} \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kr} - \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kj} \quad (17)$$

就可保证 $f(NET_i) = Y_i \quad \forall i \in \{i_1, i_2, \dots, i_p\}$

但在实际设计中, (17) 式给出的 η 下界不够确切. 现作进一步分析. 由于 $i \in \{i_1, i_2, \dots, i_p\}$ 故 (17) 式右端大于零. 则有

$$\begin{aligned} \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} (\sum_{r \neq j} q_{kr} - q_{kj}) &\leq (\sum_{r \neq j} |q_{kr}| + |q_{kj}|) \sum_{k=1}^l |a_{ik}| \lambda_k^{-\frac{1}{2}} \\ &\leq (1 + \sum_{r=1}^K |q_{kr}|) \sqrt{\sum_{k=1}^n a_{ik}^2} \sqrt{\sum_{k=1}^l \lambda_k^{-1}} \\ &\leq (1 + \sqrt{K}) \|X_i\| \sqrt{\sum_{k=1}^l \lambda_k^{-1}} \end{aligned} \quad (18)$$

因为 λ_k 是 XX^T 的特征值, 则 λ_k^{-1} 为 $(XX^T)^+$ 的特征值

由文献[3]知 $\sum_{k=1}^l \lambda_k^{-1} = tr((XX^T)^+)$

其中 $tr(\cdot)$ 表示矩阵的迹. 结合 (18) 式取 η 为

$$\eta \geq (1 + \sqrt{K}) * \sqrt{tr((XX^T)^+)} \max \|X_i\| \quad (19)$$

即可保证 K 个所学模式全部得到正确回忆. 此时对于 Kohonen 模型有

$$net_{ij}^{ko} = y_{ij} \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kj} + \sum_{r \neq j} \sum_{k=1}^l a_{ik} \lambda_k^{-\frac{1}{2}} q_{kr} \quad (20)$$

对于 Murakami 模型有

$$net_{ij}^{mu} = y_{ij} \sum_{k=1}^l a_{ik} \lambda_k^{\frac{1}{2}} (\lambda_k + K\sigma^2)^{-1} q_{kj} + \sum_{r \neq j} y_{rj} \sum_{k=1}^l a_{ik} \lambda_k^{\frac{1}{2}} (\lambda_k + K^2\sigma^2)^{-1} q_{kr} \quad (21)$$

比较 (16)、(20) 和 (21) 式可知, 改进模型的信噪比比 Kohonen 模型和 Murakami 模型优越得多, 且可以通过参数 η 来进一步作出改善. 除了能全部正确地回忆出所学数据之外, 改进模型还具有非线性映射能力. 事实上, 我们已证实了这一点. 下面仅就 XOR 问题给出一个说明.

$$\begin{array}{cccc} \text{设} & X_1 = (-1, -1)^T & X_2 = (-1, 1)^T & X_3 = (1, -1)^T & X_4 = (1, 1)^T \\ & Y_1 = (1)^T & Y_2 = (-1)^T & Y_3 = (-1)^T & Y_4 = (1)^T \end{array}$$

显然 $M=0$ 和 $M^{mu}=0$. 故 Kohonen 模型和 Murakami 模型均无法记忆这些模式数据. 原因是它们具有高度的非线性性. 对于我们的模型, 由于 $\forall i=1, 2, 3, 4, Y_i \neq f(MX_i)$. 故设 $p=$

$$4, \text{ 此时由构造知 } Z = I_4, M_2 = Y, M_1 = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \text{ 取 } d_1 = d_2 = d_3 = d_4 = 1 < \min_{j \neq i}$$

$\{\|X_i - X_j\|^2\}, \eta = 1$. 从而获得了完全正确的回忆.

对于任何其他非线性问题,改进模型总能解决.对噪声敏感性的说明.若 $X' \in R^n$ 是一个带噪输入列矢量,只要 $\|X' - X_i\|^2 \leq d_i$, X' 就能回忆出 Y_i 来.因此容错能力取决于 X_i 间的欧氏距离.

2 模 拟

实现计算机模拟关键在于 X^+ 的算法设计.有了 X^+ 便可获知 M . 而 M_1 根据(6)式而设置. $M_2 = YZ^+ = YZ^T$ 成为相关编码阵,它们较易实现.现由文献[3]中的 Greville 算法对 X^+ 的迭代算法如下.

设 $A \in R^{n \times K}$, 记 $A_k = (A_{k-1}, a_k)$, $k=2, 3, \dots, n$, 则 A_k 的广义逆迭代公式为

$$A_k^+ = (A_{k-1} a_k)^+ = \begin{bmatrix} A_{k-1}^+ - d_k b_k^T \\ b_k^T \end{bmatrix} \quad (22)$$

其中 a_k 为 A 的第 k 列.

$$b_k^T = \begin{cases} c_k^+ & \text{若 } c_k \neq 0; \\ (1 + d_k^T d_k)^{-1} d_k^T A_{k-1}^+ & \text{若 } c_k = 0 \end{cases} \quad (23)$$

$$d_k = A_{k-1}^+ a_k, \quad k=2, 3, \dots, n$$

$$c_k = a_k - A_{k-1} d_k, \quad k=2, 3, \dots, n$$

当 $c_k=0$ 且 $d_k=0$ 时,自然有 $a_k=0$ 且 $b_k=0$. 不过在我们的情形中, $a_k \neq 0$. 现令 $a_k = X_k$, $k=1, 2, \dots, K$. 则经 n 次迭代后便获得了 X^+ , 从而完成了 K 个训练数据对的学习或记忆.

现设 $X_k \in R^n$, $k=1, 2, \dots, K$ 随机地产生于 $[-a, a]$ ($a > 0$), 相应的 $Y_k \in \{-1, 1\}^m$, $k=1, 2, \dots, K$ 亦可随机产生也可人为预置. 设 $n=m=20$, 分别取 $K=10, 20, 30, 40, 50$ 获得了 Kohonen, Murakami 和改进模型的性能数据如表 1 所示. 结果与分析十分吻合.

表 1

模型	K 回忆率	K				
		10	20	30	40	50
Kohonen		6	3	17	20	23
Murakami		7	13	21	28	32
所提模型		10	20	30	40	50

其中 Murakami 模型中的 $\sigma^2 = 0.03$.

注:在 Kohonen, Murakami 模型中, $(X_i, Y_i) \in R^n \times R^m$ ($i=1, 2, \dots, K$), 我们的模型中 $(X_i, Y_i) \in R^n \times \{-1, 1\}^m$. 为便于比较,我们将它们全统一在 $R^n \times \{-1, 1\}^m$ 空间中讨论. 实际上,这种限制在应用中极其平凡,因此结果仍具相当的代表性.

参考文献

- 1 Kohonen T. Self-organization and associative memory. New York: Springer-Verlag, 1984.
- 2 Murakami K, Aibara T. Least squares associative memory and a theoretical comparison of its performance. IEEE. Syst. Man Cybern., 1989, 19(1):1230~1234.
- 3 蒋正国, 施国梁. 矩阵理论及其应用. 北京: 北京航空航天大学出版社, 1988.

GENERALIZED INVERSE ASSOCIATIVE MEMORY WITH NONLINEAR MAPPING CHARACTERISTICS

Chen Songcan Gao Hang Yang Guoqing

(Department of Computer Science Nanjing University of Aeronautics and Astronautics Nanjing 210016)

Abstract Considering some AM (Associative Memories) (such as Kohonen and Murakami models) implemented with generalized inverse lack complete recall to the stored data and nonlinear mapping abilities, in this paper an improved associative memory model with complete recall and nonlinear mapping abilities is presented by adding one extended (hidden) layer to original AM models. The theoretical analysis based on matrix SVD (singular value decomposition) and the simulation on computer show the performance superiority of the improved model.

Key words Generalized inverse, associative memories, neural network, singular value decomposition, nonlinear mapping.