最小平方函数链式联想存储器*

陈松灿

(南京航空航天大学计算机系　南京　210016)

摘要　本文借助于输入的函数扩展思想,将其引入到一般类型的联想存储器模型,通过噪声输入模式来优化所提出的模型,推广了一类联想存储器模型,实验结果表明最小平方函数链式联想存储器(FLAM)比之KOHONEN模型、MURAKAMI模型在性能上优越.

关键词　联想存储,函数链扩展,神经网络,最小平方误差.

目前有许多有趣的联想存储器模型,信息在整个存储器中以分布方式存储,这些联想存储器模型能够通过原存模式的部分去回忆出整个已存模式,极快地获得回忆信息,具有一定的容错性,因此被广泛地应用于模式识别,数据流体系,数据库机及逻辑程序设计.

同KOHONEN提出的MOORE－PENROSE广义逆联想存储器[1]是最有趣的模型之一,很简单,然而对输入噪声极端敏感,从而造成不可接受的联想误差.MURAKAMI提出了最小平方联想存储器[2],试图利用噪声模式来优化存储器的全局性能,并从理论上证明了所提模型的优越性,但仍然没有排除一类问题的解决,如XOR及高阶奇偶校验问题.原因在于这些模型作为异联想存储器,其输入输出间的关系是线性的,从而无法实现非线性映射问题.本文利用输入的函数扩展方式,结合噪声输入模式来优化联想存储器的联想性能,不仅使模型具有了非线性映射能力,并在均方误差最小意义下达到最优的性能.

1　最小平方函数链联想存储器

1.1　函数扩展方式

函数的扩展思想来自于PAO所提出的函数链思想[3],通过对输入模式作函数扩展,即在输入层引入输入模式分量的非线性扩展项,使简单的二层网络具有了很强的非线性映射能力,如使单层感知器具有了非线性可分能力,使BP模型从结构上简化为二层,免除了中间隐节点层,大大提高了学习速度,并使输入输出间的关系可显式表达.但是PAO的函数链网络采用的是δ学习规则,因此,动态行为差,在硬件实现上相对困难得多.本文仅借助其函数扩展思想,对联想存储器利用均方误差最小,获得模式的存储编码阵M.其硬件可仿Widrow的自适应滤波器实现方法.

---

* 本文部分研究获"攀登计划"资助.作者陈松灿,1962年生,副教授,主要研究领域为模式识别,神经网络,信号处理.

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1994－08－05收到,1994－12－29定稿

在本文中，所加非线性扩展项均是二阶的.不过在实际应用中扩展的项数视情形而定.扩展的方式可见文献[3].

1.2 最小平方函数链 AM

本节中，我们将构造出最小平方函数链 AM. 在最小均方误差意义下，使其获得最优的联想特性.

在记忆阶段，有 K 对模式$\{(X_k,Y_k)\}_{k=1}^{K}$，输入经扩展后成$\{(\overline{X}_k,Y_k)\}_{k=1}^{K}$，其中 $\overline{X}_k$ 中的扩展项形如 $x_i\times x_j(j>i)$，x_i,x_j 为模式分量. 将后者提交联想存储器，并以矩阵形式 M 记忆. 我们称 $\overline{X}_k$为扩展的关键模式向量，Y_k为相应的数据向量.

$$X_k=(x_{k1},x_{k2},\cdots,x_{kn})^T \tag{1}$$

$$Y_k=(y_{k1},y_{k2},\cdots,y_{km})^T \tag{2}$$

$$\overline{X}_k=(x_{k1},\cdots,x_{kn},x_{k1}x_{k2},\cdots,x_{ki}x_{kj},\cdots)^T \qquad k=1,2,\cdots,K \tag{3}$$

上述诸式中 T 表转置，且 $X_k\in\{-1,1\}^n$ 或 R^n，$Y_k\in\{-1,1\}^m$ 或 R^m，因而 $\overline{X}_k\in\{-1,1\}^p$ 或 R^p，$p\geqslant n$.

设在回忆阶段，一个退化关键模式(或带噪模式)$X_h^d\in R^N$ 扩展后成 $\overline{X}_h^d(h=1,2,\cdots,K)$ 提交存储器，Z_k 为相应的输出，且 $Z_h\in R^m$，故有

$$Z_h=M\overline{X}_h^d \tag{4}$$

其中

$$Z_h=(z_{h1},z_{h2},\cdots,z_{hm})^T \tag{5}$$

由于 $\overline{X}_h^d$ 中含有非线性扩展项，因此，输入输出间的关系是非线性的. M 为 $m*p$ 阶矩阵($p\geqslant n$)，优化存储器的思想是希望 Z_h 尽可能接近 Y_h，使它们的误差尽可能小，从而构造出使误差最小的 M. 下面我们认为，退化输入关键模式向量 X_h^d 是 X_h 中渗入了噪声所引起的，因此

$$X_h^d=X_h+n_h \tag{6}$$

其中

$$n_h=(n_{h1},\cdots,n_{hn})^T \tag{7}$$

附加噪声 n_h 是一个 n 维随机向量，且有如下特性

$$E\{n_{hi}\}=0 \tag{8}$$

$$E\{n_{hi}n_{hj}\}=\begin{cases}\sigma^2 & i=j\\ 0 & i\neq j\end{cases} \tag{9}$$

E 为期望算子，$h=1,2,\cdots,K$. 为了决定 M，我们定义均方误差函数 $J(M)$是

$$\begin{aligned}J(M)&=E\{\sum_{h=1}^{K}\|Y_h-M\overline{X}_h^d\|^2\}\\&=E\{\|Y-M\overline{X}^d\|^2\}\end{aligned} \tag{10}$$

$\overline{X}^d$，Y 分别是 $p*K$，$m*K$ 阶矩阵，其定义分别为.

$$\overline{X}^d=(\overline{X}_1^d,\overline{X}_2^d,\cdots,\overline{X}_K^d) \tag{11}$$

$$Y=(Y_1,Y_2,\cdots,Y_K) \tag{12}$$

它们分别称为关键噪声输入矩阵和数据矩阵. 利用假定(8)及(9)，$J(M)$可变换为

$$\begin{aligned}J(M)&=E\{trace[(Y-M\overline{X}^d)(Y^t-\overline{X}^{dT}M^T)]\}\\&=trace\{(YY^T-YE(\overline{X}^d)M^T-ME(\overline{X}^d)Y^T+ME\{\overline{X}^d\overline{X}^{dT}\}M^T)\end{aligned} \tag{13}$$

其中

$$\overline{X}=(\overline{X}_1,\overline{X}_2,\cdots,\overline{X}_K) \tag{13'}$$

对(13)式关于 M 求导并令导数为零，则可获得

$$ME\{\overline{X}^d\overline{X}^{dT}\}=YE\{\overline{X}^T\}=Y\,\overline{X}^T \tag{14}$$

利用 *Moore*—*Penrose* 广义逆有

$$M=Y\overline{X}^T(E(\overline{X}^d\overline{X}^{dT}))^+ \tag{15}$$

"＋"表广义逆，T 表转置.

记　$\overline{X}_h=(x_{h1},\cdots,x_{hn},\cdots,x_{hi}x_{hj},\cdots,)^T$

$$\overline{X}^d=(\overline{X}_1^d,\cdots,\overline{X}_k^d)=\begin{bmatrix} x_{11}+n_{11} & \cdots & x_{h1}+n_{h1} & \cdots & x_{K1}+n_{K1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1n}+n_{1n} & \cdots & x_{hn}+n_{hn} & \cdots & x_{Kn}+n_{Kn} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (x_{1i}+n_{1i})(x_{1j}+n_{1j}) & \cdots & (x_{hi}+n_{hi})(x_{hj}+n_{hj}) & \cdots & (x_{Ki}+n_{Ki})(x_{Kj}+n_{Kj}) \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

$$=\overline{X}+\overline{N}+N_1+N_2 \tag{16}$$

其中 $\overline{X}$ 同前(13')式　$\overline{N}=(\overline{n}_1,\overline{n}_2,\cdots,\overline{n}_K)$　(17)

且 $E(\overline{N})=0$，

$$N_1=\begin{bmatrix} O_{n\times K} \\ \cdots x_{hi}n_{hj}\cdots \end{bmatrix}_{p\times K},\quad N_2=\begin{bmatrix} O_{n\times K} \\ \cdots n_{hi}x_{hj}\cdots \end{bmatrix}_{p\times K} \tag{18}$$

且 $E(N_1)=E(N_2)=0$

故

$$E(\overline{X}^d\overline{X}^{dT})=\overline{X}\overline{X}^T+E(\overline{N}\overline{N}^T)+E(N_1N_1^T)+E(N_2N_2^T)$$
$$+E(\overline{N}N_1^T)+E(\overline{N}N_2^T)+E(N_2\overline{N}^T)+E(N_2\overline{N}^T) \tag{19}$$

我们假设 $X_h\in\{-1,1\}^n$，$h=1,2,\cdots,K$. 故 $\overline{X}_h\in\{-1,1\}^p$，$Y_h\in R^m(h=1,2,\cdots,K)$

从而

$$E(N_1N_1^T)=E(N_2N_2^T)=\begin{bmatrix} O_{n\times n} & & O \\ & K\sigma^2 & \\ 0 & \ddots & \\ & & K\sigma^2 \end{bmatrix} \tag{20}$$

$$E(\overline{N}\overline{N}^T)=\begin{bmatrix} K\sigma^2 & & & & & 0 \\ & \ddots & & & & \\ & & K\sigma^2 & & & \\ & & & K\sigma^4 & & \\ & & & & \ddots & \\ 0 & & & & & K\sigma^4 \end{bmatrix} \tag{21}$$

尽管我们希望求得使 $J(M)$ 最小的 M，但为简洁，我们在此退一步，不考虑(19)式中后面的四项. 故(19)式简化为

$$E(\overline{X}^d\overline{x}^{dT}) = \overline{X}\overline{X}^T + \begin{bmatrix} K\sigma^2 & & & & & \\ & \ddots & & & 0 & \\ & & K\sigma^2 & & & \\ & & & K\sigma^4+2K\sigma^2 & & \\ & 0 & & & \ddots & \\ & & & & & K\sigma^4+2K\sigma^2 \end{bmatrix}_{p*p}$$

$$= \overline{X}\overline{X}^T + K\sigma^2 H_p \tag{22}$$

其中

$$H_p = \begin{bmatrix} I_n & & & \\ & \sigma^2+2 & & 0 \\ 0 & & \ddots & \\ & & & \sigma^2+2 \end{bmatrix} \tag{22'}$$

为对角非奇异阵.

最后我们的 M 为

$$M = Y\overline{X}^T(\overline{X}\overline{X}^T + K\sigma^2 H_p)^+ \tag{23}$$

讨论(1)当 $\sigma^2=0$ 时,$M=Y\overline{X}^T(\overline{X}\overline{X}^T)^+$ 退化为函数链式递归联想存储器;它是 KOHONEN 联想存储器的推广.并具有非线性映射能力.

(2)当输入不作扩展时,

$$M = YX^T(XX^T + K\sigma^2 I_n)^+ \tag{24}$$

此即为 MURAKAMI 的最小平方联想存储器模型.他证明了由(24)式表示的联想存储器在性能上高于 KOHONEN 的存储器,联想更稳定.

(3)当 $\sigma^2\neq 0$ 时,$(\overline{X}\overline{X}^T + K\sigma^2 H_p)^+$ 成为普通逆,因此 M 易于获得.

2 例 子

本节要说明,尽管 MURAKAMI 的 LSAM 在联想性能上较 KOHONEN 的为优,但同样无法实现 XOR 及高阶奇偶问题的异联想存储编码.

设 XOR 问题的异联想如下:

$$X_1=(1,1)^T \qquad Y_1=(1)$$
$$X_2=(-1,-1)^T \qquad Y_2=(1)$$
$$X_3=(-1,1)^T \qquad Y_3=(-1)$$
$$X_4=(1,-1)^T \qquad Y_4=(-1)$$

对于 LSAM 而言,由于 $YX^T=0$ 故而 $M=0$

因此不论在无噪或是有噪情形下,MURAKAMI 的 LSAM(同时包括 KOHONEN 模型)关于 XOR 的联想失败.对于高阶同样如此.

而在我们的模型下,设

$$\overline{X}_1=(x_{11},x_{12},x_{11}x_{12})^T=(1,1,1)^T$$
$$\overline{X}_2=(x_{21},x_{22},x_{21}x_{22})^T=(-1,-1,1)^T$$
$$\overline{X}_3=(x_{31},x_{32},x_{31}x_{32})^T=(-1,1,-1)^T$$
$$\overline{X}_4=(x_{41},x_{42},x_{41}x_{42})^T=(1,-1,-1)^T$$

Y 同前

则若 $\sigma^2=0$，那么 $M=(0,0,1)$. 故有 $Y=M\overline{X}=(1,1,-1,-1)$联想获得完全的成功.

若 $\sigma^2\neq 0$ 则$(\overline{X}\overline{X}^T+K\sigma^2H_p)^+=(4I_3+4\sigma^2H_p)^+$成为普通逆.

求得 M 为 $M=(0,0,1/(1+\sigma^2)^2)$

显然在带噪情形下

$$Y=M\overline{X}=(1/(1+\sigma^2)^2,1/(1+\sigma^2)^2,-1/(1+\sigma^2)^2,-1/(1+\sigma^2)^2)$$

若对 Y 的各元素以零作为阈值，则同样获得正确联想. 关于高阶奇偶问题的异联想能获得同样的效果. 此外，对一般性问题（如随机产生输入输出并构成对应）的计算机初步模拟，随着扩展项的增加，性能逐步提高，且当不增加扩展项时，性能与文献[2]中的模型相当.

3　结束语

所提出的模型借助输入的函数扩展，达到了较高程度的非线性映射能力. 因此它可广泛地应用于模式识别等领域.

致谢　感谢审稿者指出了文中的错误及提出的建议！

参考文献

1　T Kohonen, M Ruohonen. Representation of associated data by matrix operator. IEEE Trans. Comput, 1973, C-**22**: 701～702.

2　K Murakami, T Aibara. Least squares associative memory and a theoretical comparison of its performance. IEEE Trans. SMC, 1989, **19**(**5**): 1230～1235.

3　Y H Pao. Adaptive pattern recognition and neural networks. New York: Addison Wesley Press, 1988.

LEAST SQUARES FUNCTIONAL—LINK ASSOCIATIVE MEMORY

Chen Songcan

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

Abstract　This paper introduces the idea of input functional expansion into a type of general associative memories, and proposes a new least squares functional link associative memory model, the performance of the model is optimized through noise input patterns. The experimental results show the associative effect of the proposed model is better than those of the Kohonen's model and the Murakami's model.

Key words　Associative memory, functional link expansion, neural networks, least squares error.