最小平方高阶联想存贮器*

陈松灿 陈德建

(南京航空航天大学计算机系,南京 210016)

摘要 本文提出了一个新的高阶联想记忆模型. 该模型采用噪声模式优化联想功能,使得对于噪声输入模式在均方误差的意义下同样达到最优的联想效果和存贮性能,推广了 Chen 的的结果,计算机的模拟结果表明了这一点.

关键词 联想记忆(或存贮),均方误差.

联想存贮器是实现分布式信息存贮的主要手段和工具,它具有较高的容错性,即对于信息部分丢失的输入,只要原输入是已存贮,那么就能得到正确的联想. 此法已被广泛地应用于模式识别,数据流结构,数据库机及逻辑程序设计等方面.

文献[1]中提出了高阶相关的联想存贮器模型,但无法摆脱对噪声的敏感性而导致不稳定的联想. 本文提出了一种采用带噪的输入模式来建立高阶联想存贮器模型,并使得此模型在均方误差最小的意义下达到最优的联想. 推广了文献[1,2]的结果,降低了存贮器对噪声的敏感性,提高了容错能力和联想的稳定性.

1 最小平方高阶联想存贮器

本节中,我们将给出最小平方意义下的高阶联想存贮器的模型及理论推导.

设在记忆阶段,联想存贮器存贮 M 对模式$(\vec{V}_1,\vec{Y}_1),(\vec{V}_2,\vec{Y}_2,),\cdots,(\vec{V}_M,\vec{Y}_M)$. 其中

$$\vec{V}_h=(V_{h1},V_{h2},\cdots,V_{hn})^T \tag{1}$$

$$\vec{Y}_h=(Y_{h1},Y_{h2},\cdots,Y_{hm})^T \tag{2}$$

式中 $h=1,2,\cdots,M$,T 表示转置. 各向量分量取自实数集. 即 $\vec{V}_n\in R^n$,$\vec{Y}_h\in R^m$. $h=1,2,\cdots,M$,R^n,R^m 分别表示 n,m 维欧氏空间.

在回忆阶段,处理一个退化输入模式向量 $\widetilde{V}_n\in R^n(h=1,2,\cdots,M)$,将其提交存贮器,那么将联想出相应的 m 维向量 $\hat{Y}_h=(\hat{Y}_{h1},\hat{Y}_{h2},\cdots,\hat{Y}_{hm})$,$(h=1,2,\cdots,M)$. 其中

$$\hat{Y}_{hl}=\sum_{u_1}\sum_{u_2}\cdots\sum_{u_k}T(l,u_1,u_2,\cdots,u_k)\widetilde{V}_{hu_1}\widetilde{V}_{hu_2}\cdots\widetilde{V}_{hu_k} \tag{3}$$

$$l=1,2,\cdots,m,u_i=1,2,\cdots,n,i=1,2,\cdots,k.$$

---

* 本文1993-07-16收到,1994-04-07定稿

作者陈松灿,1962年生,副教授,主要研究领域为人工智能,模式识别. 陈德建,1972年生,助工,主要研究领域为模式识别.

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

© 中国科学院软件研究所 http://www.jos.org.cn

$T(l,u_1,u_2,\cdots,u_k)$是连接权值,它是一个超矩阵的元素. $\hat{Y}_h$ 是 $\tilde{V}_h$ 所回忆而得的输出向量,我们期望 $\hat{Y}_h$ 接近 $\vec{Y}_h$.

下面我们假定输入向量 $\tilde{V}_h$ 是 $\vec{V}_h$ 的噪声输入向量

$$\tilde{V}_h=\vec{V}_h+\vec{n}_h, h=1,2,\cdots,M. \tag{4}$$

其中 $\vec{n}_h=(n_{h1},n_{h2},\cdots,n_{hn})^T$ 是 n 维的随机向量,且具有如下特性

$$E\{n_{hi}\}=0 \tag{5}$$

$$E\{n_{hi}n_{hj}\}=\begin{cases}0 & i\neq j\\ \sigma_h^2 & i=j\end{cases}\qquad h=1,2,\cdots,M. \tag{6}$$

这里 $E\{\cdot\}$表示随机变量的期望算子. 我们希望求出超矩阵$(T(l,u_1,u_2,\cdots,u_k))_{l,u_1,\cdots,u_k}$,使得下式的目标函数为最小.

$$J(T)=E\{\sum_{h=1}^{M}\|\vec{Y}_h-\hat{Y}_h\|^2\} \tag{7}$$

根据(7)式定义,$J(T)$又可写为

$$J(T)=E\{\sum_{h=1}^{M}\sum_{l=1}^{m}(Y_{hl}-\hat{Y}_{hl})^2\} \tag{8}$$

令
$$\frac{\partial J(T)}{\partial T(s,\lambda_1,\lambda_2,\cdots,\lambda_k)}=0\qquad \forall\ s,\lambda_1,\lambda_2,\cdots,\lambda_k \tag{9}$$

从而有
$$\sum_{h=1}^{M}\sum_{l=1}^{m}E\{(Y_{hl}-\hat{Y}_{hl})\cdot\frac{\partial(Y_{hl}-\hat{Y}_{hl})}{\partial T(s,\lambda_1,\lambda_2,\cdots,\lambda_k)}\}=0 \tag{10}$$

其中,$s=1,2,\cdots,m,\lambda_i=1,2,\cdots,n,i=1,2,\cdots,k$.

显然
$$\frac{\partial(Y_{hl}-\hat{Y}_{hl})}{\partial T(s,\lambda_1,\lambda_2,\cdots,\lambda_k)}=\begin{cases}\tilde{V}_{hu_1}\tilde{V}_{hu_2}\cdots\tilde{V}_{hu_k}, & \text{当 } l=s,u_i=\lambda_i,i=1\sim k \text{ 时},\\ 0, & \text{否则}.\end{cases} \tag{11}$$

故(10)式化为
$$\sum_{h=1}^{M}E\{(Y_{hs}-\hat{Y}_{hs})\tilde{V}_{hu_1}\tilde{V}_{hu_2}\cdots\tilde{V}_{hu_k}\}=0 \tag{12}$$

利用条件(5)(6)式,(12)式化简为

$$\sum_{h=1}^{M}Y_{hs}V_{h\lambda_1}V_{h\lambda_2}\cdots V_{h\lambda_k}=\sum_{h=1}^{M}\sum_{u_1}\sum_{u_2}\cdots\sum_{u_k}T(s,u_1,\cdots,u_k)\prod_{i=1}^{k}V_{hu_i}\cdot V_{h\lambda_i}+T(s,\lambda_1,\lambda_2,\cdots,\lambda_k)\sum_{h=1}^{M}\sigma_h^{2k} \tag{13}$$

其中,$s=1,2,\cdots,m,\lambda_i=1,2,\cdots,k$. 记

$$T=(T(s,\lambda_1,\lambda_2,\cdots,\lambda_k))_{s,\lambda_1,\cdots,\lambda_k} \tag{14}$$

$$V=(V(h,\lambda_1,\lambda_2,\cdots,\lambda_k))_{h,\lambda_1,\cdots,\lambda_k} \tag{15}$$

$$V(h,\lambda_1,\lambda_2,\cdots,\lambda_k)=V_{h\lambda_1}V_{h\lambda_2}\cdots V_{h\lambda_k} \tag{16}$$

$$Y=(Y_{hs})_{m\times M} \tag{17}$$

则(17)式对于一切 $s,\lambda_1,\lambda_2,\cdots,\lambda_k$ 有

$$YV^T=T(VV^T+(\sum_{h=1}^{M}\sigma_h^{2k})\cdot I) \tag{18}$$

其中,矩阵左上方的标记 T 表转置,J 与 T 同阶且为单位阵. 故有

$$T=YV^T(VV^T+(\sum_{h=1}^{M}\sigma_h^{2k})I)^+ \tag{19}$$

© 中国科学院软件研究所　http://www.jos.org.cn

上式中"＋"表示广义逆，且是唯一的.

讨论:(i)若至少有一个 $\sigma_n>0$，则 $VV^T+(\sum_{h=1}^{M}\sigma_h^{2k})I$ 是非奇异的，从而广义逆变为一般矩阵的逆，故 T 直接可求. (ii)若 $\sigma_n=0, h=1,2,\cdots,M$. 则(19)式化为

$$T=YV^T(VV^T)^+ \tag{20}$$

它可看作是高阶 Kohonen 联想存贮阵，特别当 $k=1$ 时，即为 Kohonen 联想存贮器模型. (iii)当 $k=1, \sigma_h=\sigma, h=1,2,\cdots,M$ 时，(18)式退化为 Murakami[3]的最小平方联想存贮器. 该作者已证明，采用噪声优化的联想存贮阵在噪声敏感性、稳定性和联想误差方面均优于 Kohonen 模型，同样优于该作者所提出的稳定化模型[4]. 然而，Murakami 所建立的输入与输出间的关系均是由二阶联想阵 M 所决定，是线性的. 本文的高阶联想超矩阵则是一非线性映射. 在文献[5]中已初步证明了高阶超阵的联想性能，存贮容量优于低价联想存贮器，而本文恰好是上述诸联想存贮模型的推广.

2 模拟及分析

2.1 模 拟

模拟主要在 SUN 工作站上进行. 目的是求得超阵 T. 为了在计算机上实现，采用梯度最陡下降法逐次迭代，直至稳定. 即 T 中元素的第 $n+1$ 次迭代是

$$T_{n+1}(s,\lambda_1,\cdots,\lambda_k)=T_n(s,\lambda_1,\cdots,\lambda_k)-\eta\frac{\partial J(T)}{\partial T(s,\lambda_1,\cdots,\lambda_k)} \qquad \forall\ s,\lambda_1,\cdots,\lambda_k.$$

其中，$T_0(s,\lambda_1,\cdots,\lambda_k)$ 在(－0.2,0.2)或(－0.01,0.01)间随机产生，而所附加的噪声在(－0.3,0.3)之间随机产生. 因此，噪声的均值为 0，均方为 0.03，η 是学习率，它影响着迭代的收敛与否，收敛速度，η 一般在 0.01～0.1 间取值，以确保收敛，但速度很慢. 本文仅考虑 $k=1,2,3$ 的情形，其实验结果如表 1 所示：

表 1 取 $n=3, m=3, \sigma_h$ 分别取 0 和 0.03

$\sigma_h^2=0.018, n=3, m=3$				$\sigma_h^2=0, n=3, m=3$			
H \ k	1	2	3	H \ k	1	2	3
1	0.0016	0.0008	0.00002	1	0.0057	0.0025	0.0005
2	0.1240	0.0666	0.0512	2	0.017	0.0042	0.0011
3	0.7114	0.6030	0.5339	3	0.1161	0.0752	0.0389
4	1.5246	0.5143	0.0504	4	1.3013	0.0467	0.0456
5	0.8058	0.6390	0.2027	5	0.9982	0.0619	0.0063
6	2.768	1.567	1.21	6	2.6699	1.7965	0.5260
7	1.9524	1.795	1.17	7	2.0010	1.6122	0.3596
8	6.515	4.158	3.688	8	6.7669	2.0346	1.6853

表中 H 表示所学模式个数，k 为网络的阶数. 表中诸数据表示总的误差，即

$$J(T)=E\{\sum_{h=1}^{M}\sum_{l=1}^{m}(Y_{hl}-\hat{Y}_{hl})^2\}$$

2.2 分析

由表 1 可知. 当 $\sigma_h^2=0$ 时，此时噪声为零，联想效果随 k 的增大总误差越来越小. 这与传统的联想存贮器效果完全相符，因此，那些联想存贮器是本存贮器模型的一个特例.

© 中国科学院软件研究所 http://www.jos.org.cn

当 $\sigma_h^2=0.03$ 时，其联想总趋势完全与 $\sigma_h=0$ 时的情形一样，随着 k 的增大，总误差越来越小，即联想效果亦越来越佳. 但从表中还看出在 $\sigma_h^2=0$ 与 $\sigma_h^2=0.03$ 两种情形中，在学习了第 4 个模式及第 6 个模式后，误差一下子变大了，其原因是①输入、输出模式均在(−1,1)之间随机产生，随机对应，因而会造成模式对之间相互对应的矛盾. ②当输入模式连续且增加 30%的噪声后，就会产生如下情形：

输入模式　　　　期望输出

$$X_1=\begin{pmatrix}0.1\\-0.1\end{pmatrix}\longrightarrow\begin{pmatrix}0.5\\0.3\end{pmatrix}$$

$$X_2=\begin{pmatrix}0.4\\-0.2\end{pmatrix}\longrightarrow\begin{pmatrix}0.2\\-0.1\end{pmatrix}$$

当在 X_1 中加入 30%的噪声后，成为 X_2，从而造成歧义现象，即相同的模式对应两个不同的期望输出. 事实上，在连续模式下，增加 30%的噪声显然过大，而在离散模式下，不存在上述问题.

3 结束语

本文所提出的高阶联想模型，利用噪声输入模式优化联想特性，使存贮性能在噪声存在时，仍能获得最佳. 它大大推广了传统的 Kohonen，Hopfield 著名联想存储器. 但由于高阶联结的复杂性，因此高阶网络在硬件的实现上目前尚存在不足.

参考文献

1 Chen H H, Lee Y C. High order correlation model for associative memory. Proc. AIP Conf. on Neural Networks for Computing, Snowbird, 1986. 86−89.

2 Kohonen T. Correlation matrix memories. IEEE Trans. Comput., 1972, C−**21**:353−359.

3 Murakami K. Least squares associative memory and a theoretical comparison of its performance. IEEE Trans. SMC., 1989, **19**:210−215.

4 Anon. An Improvement on the moore−penrose generalized inverse associative memory. IEEE Trans. SMC., 1987, **17**:699−707.

5 Prodos D L. Capacity of neural nets. Electronics Letters, 1988, **24**:454−455.

LEAST SQUARE HIGH ORDER ASSOCIATIVE MEMORY

Chen Songcan　　Chen Dejian

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

Abstract　In this paper, a new high order associative memory model is proposed. The associative function of the model is optimized by using noise patterns such that the best associative effect and memory performance for noise input patterns os obtained in mean square error sense, and Chen's results are generalized. The simulation on confirms them.

Key words　Associative memory (or storage), mean squares errors.

© 中国科学院软件研究所　http://www.jos.org.cn