第 21 卷 第 1 期
1995 年 1 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 21, No. 1
Jan., 1995

短 文

感知器的动态稀疏化学习

汪 涛　邢小良
（浙江大学计算机系　杭州　310027）

摘　要

本文提出了一种感知器的动态稀疏化（dynamic dilution）概念，同时估计权值和减少神经元间的连接权个数. 动态稀疏化有效地克服了传统的静态稀疏化(先确定权值，然后减少连接权个数)的缺陷. 计算机实验结果说明了算法的优越性.

关键词： 感知器，学习算法，动态稀疏化，静态稀疏化.

1　引言

神经网络以其大规模并行处理、容错性、自学习和自组织的特点，得到了国内外学者的广泛重视.在现有的各种神经网络模型[1]中，连接权的个数往往很多，例如，在 Hopfield 网络中神经元相互连接，层次网络中每个神经元与前一层中的所有神经元连接. 其主要缺点是，随着神经元数目的增加，连接权的个数将迅速增加，尤其是高阶网络，给神经网络的大规模集成电路实现带来困难，同时提高了软件模拟的复杂度. 事实上，脑科学研究表明，人脑中神经元间的连接是非常稀疏的[2]，神经元数目与突触数目之比达到 10^6—10^7 左右，明显不同于现有的神经网络模型的连接情况.

为了减少连接权数目，静态稀疏方法[3]在学习算法确定连接权值以后，删除一些连接. 其弱点是稀释后的网络性能较差. 局部连接网络[4,5]只连接拓扑结构上相邻神经元，达到减少连接权个数的目的. 但是，这种网络的适用范围受到限制，而且需要人为地定义邻域关系.

神经网络的一个重要性质是知识的分布式表达，即通过学习算法将蕴含在训练样本中的信息，分布到神经元间的连接权上. 因此，人为地删除部分连接权会丢失某些信息，这是静态稀疏化的弱点根源所在. 但是，如果确定权值和稀释网络同时进行，就会使得权值重新分配，有效地保护连接权上的信息. 基于这种思想，本文提出了动态稀疏化的概念，并且应用于感知器的学习.

本文于 1993 年 1 月 4 日收到

2 感知器的动态稀疏化算法

感知器是一种简单的神经网络模型，只能处理线性可分问题．感知器学习算法的目标是求一组连接权 $W=[w_0,w_1,\cdots,w_N]$,使M个学习样本是$(X^k,t_k)(k=1,2,\cdots,M)$满足

$$\langle X^k,W\rangle\begin{cases}>0 & \text{如果 } t_k=+1\\ <0 & \text{如果 } t_k=-1\end{cases}\quad k=1,2,\cdots,M \tag{1}$$

其中，t_k 是期望输出，$X^k=[x_0^k,x_1^k,\cdots,x_N^k]$ 是$N+1$ 维输入值，它的第一个分 量 $x_0^k=1$,对应于网络的阈值 w_0．运算符$\langle\cdot,\cdot\rangle$表示点积．令

$$Z^k=\begin{cases}+X^k & \text{如果 } t_k=+1\\ -X^k & \text{如果 } t_k=-1\end{cases}\quad k=1,2,\cdots,M \tag{2}$$

这时,感知器学习算法可归结于求权向量W,使得以下不等式方程组成立

$$\langle Z_k,W\rangle=b_k>0,k=1,2,\cdots,M \tag{3}$$

直接 Ho-Kashyap(DHK) 算法[6]是一种求解方程(3)的迭代估计方法,并且能在有限步内收敛,其性能明显优于感知器学习算法[7]．为了降低 DHK 算法的计算复杂度，Hassoun[8] 提出了一种自适应 AHK 算法,利用梯度下降法,求目标函数 $J(W,b)$的最小值，

$$J(W,b)=(\langle Z^k,W\rangle-b_k)^2 \tag{4}$$

并且满足约束条件 $b_k>0$．其中 $b=[b_1,b_2,\cdots,b_M]$，Z^k 是当前输给感知器的学习样本．

动态稀疏的基本思路是增加一个与权向量W相对应的$N+1$ 维向量S，每个分量的取值是 $s_i\in\{0,1\}$．$s_i=0$ 表示权值 w_i 被删除．令 $Y=[y_0,y_1,\cdots y_N]$表示保留的连接权值组成的向量，$y_i=s_iw_i(i=0,1,\cdots,N)$．利用 HK 算法的目标函数(4)，定义动态稀疏学习的目标函数为

$$J(W,S,b)=(\langle Z^k,Y\rangle-b_k)^2 \tag{5a}$$

满足约束条件 $s_j=0$ 或 $s_j=1,j=0,1,\cdots,N$;

$$\sum_i s_i=P;\ b_k>0. \tag{5b}$$

其中，P 是网络稀疏化后保留的连接权个数,$(N-P)/N$ 是网络的稀疏率．这是一个约束最优化问题,是在连续空间(W,b)和离散空间(S)上的优化问题．我们首先将它转化为一个等价的非约束优化问题[9]，

$$E(W,S,b)=J(W,S,b)+r[\sum_i s_i^2(1-s_i)^2+(\sum_i s_i-P)^2] \tag{6}$$

其中 r 是惩罚参数．为了获得有效解,即满足约束条件(5b),惩罚参数 r 必须足够大．其缺点是当参数 r 非常大时,优化过程会过多地考虑满足约束条件的解,而忽略求 $J(W,S,b)$的最小化;如果参数 r 太小,获得的解可能不满足约束．

一个切实可行的方法[9]是：从较小的 r 值开始,然后逐渐增大 r,直到算法收敛．在给定参数 r 值时,用梯度法求 $E(W,S,b)$的最小值．$E(W,S,b)$ 对各参数的导数是

$$dE/db_k=-2e_k \tag{7a}$$

$$dE/dw_i = 2e_k s_i z_i^k \tag{7b}$$

$$dE/ds_i = 2e_k w_i z_i^k + 2r[s_i(s_i - 1)(2s_i - 1) + (\sum_i s_i - P)] \tag{7c}$$

其中误差 $e_k = \langle Z^k, Y\rangle - b_k$.

根据梯度下降算法和 (7a),同时考虑到满足约束条件 $b_k > 0$，我们分两种情况迭代估计各参数：①如果 $b_k(t) + \beta_1 e_k(t) > 0$,则

$$b_k(t+1) = b_k(t) + \beta_1 e_k(t) \tag{8a}$$

$$w_i(t+1) = w_i(t) - \beta_2[\langle Z^k, Y\rangle - b_k(t+1)]s_i(t)z_i^k$$

$$= w_i(t) + \beta_2(\beta_1 - 1)e_k(t)s_i(t)z_i^k \tag{8b}$$

$$s_i(t+1) = s_i(t) - \beta_3[\langle Z^k, Y\rangle - b_k(t+1)]w_i(t)z_i^k$$
$$- \beta_3 r[s_i(t)(s_i(t) - 1)(2s_i(t) - 1) + g(t)]$$
$$= s_i(t) + \beta_3(\beta_1 - 1)e_k(t)w_i(t)z_i^k$$
$$- \beta_3 r[s_i(t)(s_i(t) - 1)(2s_i(t) - 1) + g(t)] \tag{8c}$$

②如果 $b_k(t) + \beta_1 e_k(t) \leqslant 0$,则

$$b_k(t+1) = b_k(t) \tag{9a}$$

$$w_i(t+1) = w_i(t) - \beta_2[\langle Z^k, Y\rangle - b_k(t+1)]s_i(t)z_i^k$$
$$= w_i(t) - \beta_2 e_k(t)s_i(t)z_i^k \tag{9b}$$

$$s_i(t+1) = s_i(t) - \beta_3[\langle Z^k, Y\rangle - b_k(t+1)]w_i(t)z_i^k$$
$$- \beta_3 r[s_i(t)(s_i(t) - 1)(2s_i(t) - 1) + g(t)]$$
$$= s_i(t) - \beta_3 e_k(t)w_i(t)z_i^k$$
$$- \beta_3 r[s_i(t)(s_i(t) - 1)(2s_i(t) - 1) + g(t)] \tag{9c}$$

其中 $\beta_1, \beta_2, \beta_3$ 是步长，$g(t) = \sum_i s_i(t) - P$，(t) 表示当前的参数估计值，$(t+1)$表示下一步的估计值.

当所有的学习样本 $Z^k(k = 1, 2, \cdots, M)$顺序地输入感知器，并且根据(8)和(9)修正参数 W, b 和 S 以后,惩罚参数 r 按如下方式递增，

$$r = r + \Delta r \tag{10}$$

现在将动态稀疏化学习算法总结如下:

步骤 1：初始化. 设置初始权值 $w_i(t=0) = 0, b(t=0) = 0.1$ 和 $s_i(t=0) = 0.5$ (是 0 和 1 的中间值);参数 $\beta_1 = 0.1, \beta_2 = 0.05, \beta_3 = 0.1$；惩罚参数 $r = 0.0$；迭代次数 $t = 0$;

步骤 2：顺序地对每个学习样本 $Z^k(k = 1, 2, \cdots, M)$,按照(8)和(9)式修正连接权值 W, b 和变量 S;

步骤 3：如果目标函数 $E(W, S, b) < 10^{-4}$，或者迭代次数 t 大于某个预定值，则终止学习算法;否则,执行步骤 4;

步骤 4：迭代次数 $t = t + 1$;根据(10)式增加 r 值(实验中置 $\Delta r = 0.01$)，转移到步骤 2.

显然,学习算法同时估计连接权值 W，b 和稀疏化变量 S，与静态稀疏化有本质差别.

3 计算机实验结果

由于感知器只能解决线性可分问题，我们考察如下三组线性可分实验，相应的决策平面是 $f(x) = a_0x_0 + a_1x_1 + \cdots + a_Nx_N = 0$。学习样本 $X^k(k = 0,1,\cdots,M)$ 的每一维坐标值 x_i^k 在[−1,+1]内随机产生，如果 $f(X^k) > 0$，则 $t_k = 1$。否则 $t_k = 0$。同样产生M个非学习样本，用来测试感知器正确分类样本的比例。

在第一组实验中，$f(x) = x_1 + x_3 + 0.25$，样本个数 $M = 50$，输入维数 $N = 3$。显然，第二维坐标 x_2 与感知器分类无关。利用动态稀疏算法（$P = 3$），获得权向量W和稀疏化向量 S 分别是

$$W = [0.070,\ \ 0.071,\ \ 0.017,\ \ 0.083]$$

$$S = [1.000,\ \ 1.000,\ \ 0.000,\ \ 1.000]$$

权分量 w_2 被删除，与决策平面完全相同。正确分类非学习样本的比例是 98%。为了与静态稀疏化相比较，我们首先用自适应 HK 算法[9]训练感知器，获得的权向量是

$$W = [0.008,\ \ 0.077,\ \ 0.011,\ \ 0.089]$$

然后删除其中的任意一个连接权，正确分类非学习样本的比例如表 1 所示，其性能明显低于动态稀疏方法。例如，当删除权值 w_1 时正确分类比例是 78%。

表 1 静态稀疏化的正确分类比例

删除权值标号	0	1	2	3
正确分类比例	85%	78%	92%	61%

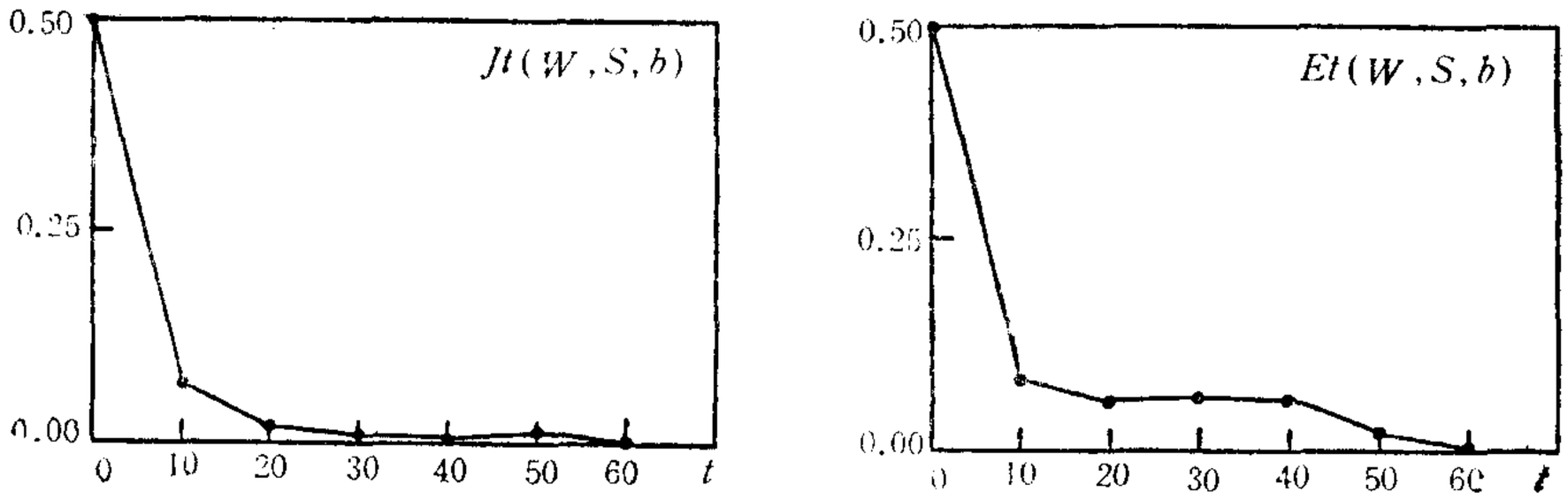


图 1 总体误差的变化情况

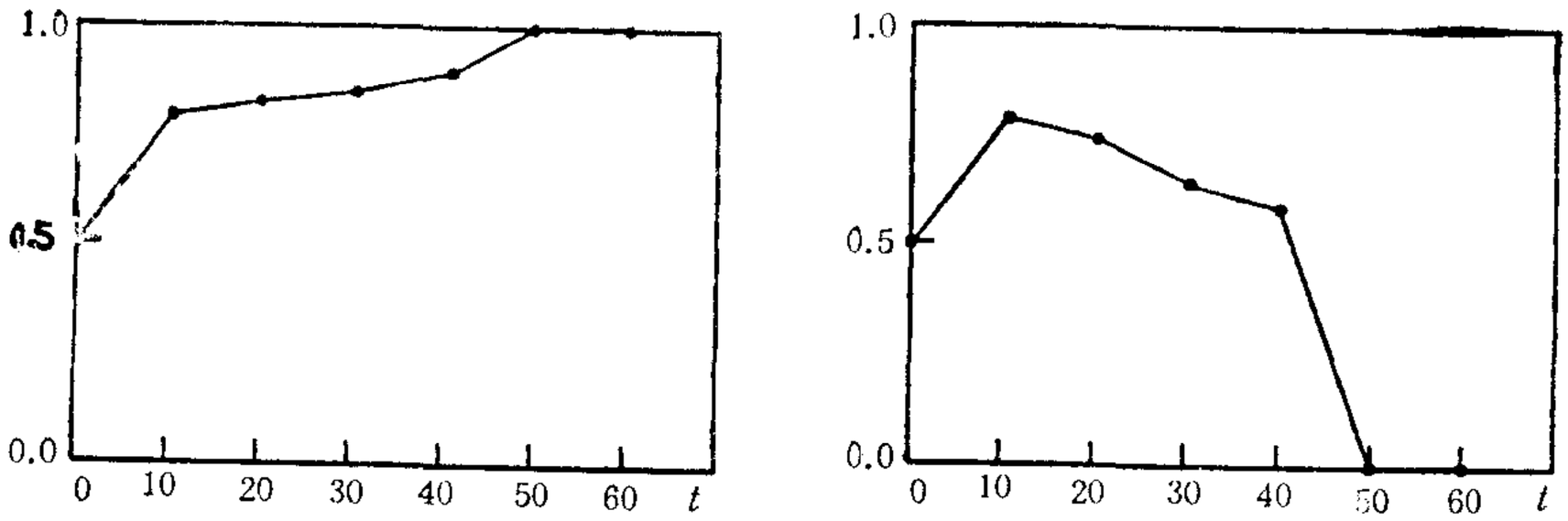
图 1 是动态稀疏化中总体误差 $J_t(W,S,b)$ 和 $E_t(W,S,b)$ 随迭代次数 t 变化的情况，其中

$$J_t(W,S,b) = \sum_k(\langle Z^k, Y\rangle - b_k)^2$$

$$E_t(W,S,b) = J_t(W,S,b) + r[\sum_i s_i^2(1 - s_i)^2 + (\sum_i s_i - P)^2]$$

稀疏变量 s_1 和 s_2 随迭代次数 t 的变化，如图 2 所示。

在第二组实验中，$f(x) = x_1$，样本个数 $M = 100$，输入维数 $N = 3$。显然，只有第



(a) s_1　　(b) s_2

图 2　稀疏变量的变化情况

一维坐标 x_1 与感知器分类有关．对于不同的稀疏参数 $P(=3,2,1)$，被删除的连接权分布和正确分类比例如表 2 所示．动态稀疏算法每次都能保留必要的连接权 W_1．

表 2　被删除的连接权分布和正确分类比例

稀疏参数 P	3	2	1
删除权的分布	2	2,3	0,2,3
正确分类比例	100%	100%	100%

在最后一组实验中，样本个数 $M=100$，输入维数 $N=3$．决策平面的参数 a_i 在[−5.0,+5.0]内随机产生，然后随机地置其中的一个参数 $a_i=0$．对于稀疏参数 $P=3$，在 100 次实验中，动态稀疏学习每次都能将对应于 a_i 的连接权删除，达到既减少连接权数目，又保护感知器性能的目的．

参　考　文　献

[1] Lippmann R P, An Introduction to Computing with Neural Nets, *IEEE ASSP Magazine*, 1987, 4—22.

[2] Muller B and Reinhardt J, Neural Networks: An Introduction, Springer-Verlag Berlin Heidelberg, 1990.

[3] Sompolinski H, Neural Networks with Nonlinear Synapses and A Static Noise, Phys. Rev. A, 1986, 34: 2571—2575.

[4] Chua L O and Yang L, Cellular Neural Networks: Applications, *IEEE Trans. on Circuits and Systems*, 1988, 35: 1273—1290.

[5] 卢科学，赵树芗．二维局域连接神经网络．计算机学报，1992，(7)：541—545.

[6] Ho Y C and Kashyap R L, An Algorithm for Linear Inequalities and Its Applications, *IEEE Trans. on Electron. Comput.*, 1965, 14: 683—688.

[7] Hassoun M H. Dynamic Heteroassociative Neural Memories. Neural Networks, 1989, 2: 275—287.

[8] Hassoun M H and Song J. Adaptive Ho-Kashyap Rules for Perceptron Training. IEEE Trans. on Neural Networks, 1992, 3: 51—61.

[9] Rao S S. Optimization Theory and Applications. New Delhi: Wiley Eastern, 1987.

DYNAMIC DILUTION FOR PERCEPTRON TRAINING

WANG TAO XING XIAOLIANG

(*Department of Computer Science & Engineering Zhejiang University Hangzhou* 310027)

ABSTRACT

In this paper, a dynamic dilution concept for perceptrons is proposed, which estimates the weights and reduces the number of connections at the same time. The dynamic dilution overcomes the weakness of the static dilution. Computer simulations are conducted to show its advantages.

Key words: Perceptron, learning algorithm, dynamic dilution, static dilution.

95'《中国控制会议》征文通知

95'《中国控制会议》拟定于一九九五年第三季度在安徽黄山举行. 会议由中国自动化学会控制理论专业委员会主办,由 IEEE 北京分部及旅英自动化协会协办,并由中国科学技术大学自动化系承办. 具体事宜如下:

一、**征文范围**: 控制理论及其应用未发表的论文,内容包括下列领域的理论与应用:

线性系统	非线性系统	随机控制系统	计算机集成制造系统
专家系统	分布参数系统	离散事件系统	社会经济系统
大系统	H_∞ 控制	适应控制	生态环境系统
鲁棒控制	预测控制	智能控制	机器人控制
模糊控制	神经网络	容错控制	系统辨识与建模
模型降阶	稳定性分析	最优估计	计算机辅助设计
工业控制			

二、**截止日期**: 收稿截止日期为 1995 年 3 月 31 日.

三、**会议请奖**: 凡申请«中国控制会议»第二届«关肇直奖»的论文,需在投稿时注明,交论文一式九份,并附工作证(或学生证)和身份证复印件,及至少一份同行教授级专家推荐意见(«关肇直奖»条例请向联系人索取).

四、**说明**:

1.会议录取的文章,将于五月初通知作者.

2.论文集将由正式出版社出版.

3.请作者自留底稿,无论是否录取,一律不退稿.

五、**联系人及地址**:

联系人: 张月田

通讯地址: 中国科学院系统科学研究所(北京中关村 100080)

电话: (01)2553063 传真: (01)2568364 电子信箱: jif@iss03.iss.ac.cn

中国自动化学会

控制理论专业委员会

一九九四年九月