第5卷第4期
1994年4月

软 件 学 报
JOURNAL OF SOFTWARE

Vol. 5, No. 4
Apr. 1994

Neocognitron 学习算法分析*

洪家荣 李星原

(哈尔滨工业大学计算机系,哈尔滨 150006)

摘要 现有的对 Neocognitron 的分析都采用代数法,因而无法研究它的动态特性.本文把 Neocognitron 及其学习算法推广到连续时域,借助微分方程来研究 Neocognitron. 文中给出了无教师学习算法一般情况下,学习过程中 u_s 层神经元输出变化规律的微分方程,指出其增加的条件,并推出权 a、b 初始值选择的一个必要条件;进一步得出无教师算法代表稳定后和有教师学习情况下 u_s 变化的一种等效显式函数,指出此时学习过程是 u_s 层神经元输出向一个系数的逼近过程,且有教师学习过程的最后状态与可变权初值和学习率无关,并讨论了影响 u_s 终值和逼近速度的因素.

关 键 词 神经网络,学习算法,自组织,Neocognitron.

Fukushima 的 Neocognitron 是一种用于视觉模式识别的神经网络模型[2-5],已经取得了广泛的关注,成为国际上著名的神经网络模型之一.它松散地建立在哺乳动物的视觉系统特性之上,一般认为它具有容忍模式变形、位移的能力,已被广泛应用于图形、数字、汉字和朝鲜文字的识别[7-10].它为视觉模式识别提供了一种新的方法,但在应用中也暴露出该模型的一些问题. Menen 和 Heinemann 的工作[8]表明,在某些情况下模型不具有对位移的容忍能力,我们[9,10]也发现其用于实际手写文字识别时在识别率和时空开销方面较之 BP、LVQ 模型有待改进.另一方面,Neocognitron 的参数多,究竟可变权的初值和一些参数是如何影响模型的性能,仍然不大清楚,这给模型的设计带来了一些困难.为了更好地设计和改进 Neocognitron,对其算法进行数学上的分析是非常必要的. Fukushima 对模型做过一些分析[5],直观地解释了 u_s、u_c 单元的本质,讨论了参数 r 等对模型的影响,Barnard 和 Casasent[1]对模型的位移无关特性进行了分析,指出只有当参数选择合适时才可获得位移近似无关,并给出了一个可达到近似无关的条件.这些工作[1,5,6]为设计和改进 Neocognitron 提供了一个基础.然而,他们对 Neocognitron 的分析都是用代数方法进行的,因而不能研究学习过程中的动态特性.本文引进微积分的方法研究 Neocognitron,并导出了一些新的有用结果.

* 本文 1991 年 10 月 22 日收到,1992 年 1 月 27 日定稿

作者洪家荣,55 岁,教授,主要研究领域为机器学习,专家系统,计算几何,文字识别.李星原,30 岁,助研,主要研究领域为文字识别,神经网络,视觉计算,机器学习.

本文通讯联系人:洪家荣,哈尔滨 150006,哈尔滨工业大学计算机系

1 Neocognitron 的连续数学描述

Neocognitron 网络是一种前馈结构，由多级组成.起始级为输入层，接收二维输入图象，以后每一级由两层组成，第一层为 u_s 单元层，它实际完成特征抽取的任务，第二层为 u_c 单元层，它产生对位置变化的容忍能力.

u_c 层神经元的权 d 和 v_c 神经元的权 c(也可以看作 u_s 的权)是固定不变的，只有 u_s 的权 a、b 可学习，也就是说，学习过程只影响 u_s 单元.因此，u_s 的变化是学习的度量，我们可以通过 u_s 的暂态和稳态来研究它的特征选择和特征抽取，通过各参数对 u_s 暂态和稳态的影响来研究它们对学习快慢和特征抽取的影响.

下面我们给出 Neocognitron 的连续数学描述.采用与文[2]一致的符号，并省去 u_s、u_c、v_c、r、q、a、b 的下标 l 或 $l-1$.

为把模型推广到连续时域，我们把学习过程中的 u_s、a、b 看做连续时间变量.此时，各类神经元的输出表达式仍同文[2]，此处给出学习规则的数学描述为：

$$a(k_{l-1},v,\hat{k}_l)=qc(v)\int u_c(k_{l-1},\hat{n}+v)dt \tag{1.1}$$

$$b(k_l)=q\int v_c(\hat{n})dt \tag{1.2}$$

其中，q 为正常数. $\hat{k}_l$、$\hat{n}$ 的选择如下：

1.采用无教师学习算法，则 $\hat{k}_l$、$\hat{n}$ 应满足：

先选候选(candidate) $u_s(\bar{k}_l,\bar{n})$：

$$u_s(\bar{k}_l,\bar{n})=max(\{u_s(k_l,n)\,|\,n\in s_c,k_l=1,\cdots,K_l\})\text{其中 } s_c \text{ 为预定义好的区域.} \tag{1.3}$$

再选代表(representative) $u_s(\hat{k}_l,\hat{n})$：

$$u_s(\hat{k}_l,\hat{n})=max(\{u_s(\bar{k}_l,\bar{n})\,|\,\bar{k}_l=\hat{k}_l\}) \tag{1.4}$$

2.采用有教师学习算法，则对每一个 k_l，都有一个预先选定的 u_c 和 $\hat{n}$，即 $\hat{k}_l$ 和 $\hat{n}$ 与时间无关.

由此可以看出，Neocognitron 的有教师学习算法可以看做其无教师学习算法的特殊情况.

2 学习过程中 u_s 的变化规律

2.1 无教师算法一般情况下的 u_s 变化规律

$$u_s=r\varphi[x/y-1] \tag{2.1}$$

其中，
$$x=1+\sum_{k_{l-1}}\sum_{v}a(k_{l-1},v,k_l)u_c(k_{l-1},n+v) \tag{2.2}$$

$$y=1+r/(1+r)b(k_l)v_c(n) \tag{2.3}$$

r 为大于 1 的常数，而 $\varphi(x)=\begin{cases}x,x\geq 0\\0,x<0\end{cases}$，为方便，以下省去 $\sum$ 下面的 k_{l-1} 和 v.

事实上，我们只要知道了 $u_s=r[x/y-1]$的变化规律，便也知道了 u_s 的变化规律.因此，我们

设

$$u_s=r(x/y-1) \tag{2.4}$$

于是，$du_s/dt=r(ydx/dt-xdy/dt)/y$ (2.5)

如果 u_c 不变化，$dx/dt=\sum\sum u_c(k_{l-1},n+v)da(k_l,v,k_{l-1})/dt$

用式(1.1)得：$dx/dt=q\sum\sum u_c(k_{l-1},n+v)c(v)u_c(k_{l-1},\hat{n}+v)$ (2.6)

同样可得，$dy/dt=q[r/(1+r)]v_c(n)v_c(\hat{n})$ (2.7)

用式(2.6)，(2.7)代入(2.5)，并用(2.4)消去 x，得：

$$\frac{du_s}{dt}=\frac{qr}{1+r}\frac{v_c(n)v_c(\hat{n})}{y}\left[(1+r)\frac{\sum\sum c(v)u_c(k_{l-1},n+v)u_c(k_{l-1},\hat{n}+v)}{v_c(n)v_c(\hat{n})}-r-u_s\right] \tag{2.8}$$

注意到 $\hat{n}$ 和 y 是与时间有关的，上式可以写成如下形式：

$$du_s/dt=f_1(t)[f_2(t)-u_s] \tag{2.9}$$

这样，如果 $f_1(t)$ 和 $f_2(t)$ 是初等函数，我们可以解出 u_s 来. 可惜的是，$f_1(t)$ 和 $f_2(t)$ 都不是时间 t 的初等函数(因 $\hat{n}$ 要满足 max 函数).

如果我们把 $u_c(k_{l-1},n+v)$ 和 $u_c(k_{l-1},\hat{n}+v)$ 按 k_{l-1} 和 v 所取的不同值分别看成向量 VA(n)和向量 $VA(\hat{n})$ 的全部分量，并定义向量 X=(x(1),x(2),…,x(V))和向量 Y=(y(1),y(2),…,y(V))的内积为：$(X,Y)=\sum c(v)x(v)y(v)$，向量 X 的模为：$\|X\|=(X,X)$

则：$\|VA(\hat{n})\|=\sqrt{\sum\sum c(v)u_c(k_{l-1},\hat{n}+v)}=v_c(\hat{n})$

$\|VA(n)\|=\sqrt{\sum\sum c(v)u_c(k_{l-1},n+v)}=v_c(n)$

而 $(VA(n),VA(\hat{n}))=\sum\sum c(v)u_c(k_{l-1},n+v)u_c(k_{l-1},\hat{n}+v)$

因此，(2.8)可写成：

$$\frac{du_s}{dt}=\frac{qr}{1+r}\frac{VA(\hat{n})VA(n)}{y}\left[(1+r)\frac{(VA(n),VA(\hat{n}))}{\|VA(n)\|\ \|VA(n)\|}-r-u_s\right] \tag{2.10}$$

这样，尽管我们不能将 u_s 解出来，也可以得到下面的结论：

定理. 如果学习时选择代表为 $u_s(\hat{k}_l,\hat{n})$，则学习后：

① $u_s(\hat{k}_l,n)$ 增加的条件是 $u_s(\hat{k}_l,n)<(1+r)(VA(n),VA(\hat{n}))/(\|VA(n)\|\ \|VA(\hat{n})\|)-r$

② $u_s(\hat{k}_l,\hat{n})$ 增加的条件是 $u_s(\hat{k}_l,\hat{n})<1$

证明：由(2.10)并注意到 $(X,X)/(\|X\|\ \|X\|)=1$，可得.

Neocognitron 的无教师学习算法是一种 winner－take－all 算法，因而 $u_s(\hat{k}_l,\hat{n})$ 在学习后应是增加的，所以 $u_s(\hat{k}_l,\hat{n})$ 应小于 1. 又因初始时任一 n 都可能选作代表，则有：

推论. 初始权 a、b 应使得 $u_s(k_l,n)<1$

由(1.2)和(2.3)得：$y=1+q[r/(1+r)]v_c(n)\int v_c(\hat{n})dt$，代入(2.9)，可知 q 与 $\triangle u_s$ 的关系，q 越大，则 $\triangle u_s$ 越大，但二者不成正比，当 q 足够大时，$\triangle u_s$ 几乎与 q 无关了. 作者在文[10]中曾提到 q 值影响学习收敛速度的经验，与此是一致的.

从式(2.9)可看出，当 $u_s>f_2(t)$ 时，u_s 减小，否则增大. 由此可否得到一种推测，u_s 的学习过程以 $u_s=f_2(t)$ 为渐近曲线.

2.2 $\hat{n}$ 不随时间变化时——采用有教师算法或无教师算法代表 $\hat{n}$ 稳定后 u_s 的变化规律

此时,式(2.9)可写成:$\frac{du_s}{dt}=\frac{K1}{y}(K2-u_s)$ (2.11)

其中,y 随时间变化,$K1=f_1(t)y$、$K2=f_2(t)$与时间无关

以 $\hat{n}$ 不随时间改变的时刻(对有教师学习就是起始时刻)为 t 的起点,由(1.2)和(2.3),得:$y=Y0+K3t$,其中 Y0 和 K3 与 t 无关,且 $K3=q[r/(1+r)]v_c(n)v_c(n)=K1$.

因此,$\frac{du_s}{dt}=\frac{K1}{Y0+K3t}(K2-u_s)$ (2.12)

解之可得:

$$u_s=K2-C(Y0+K3t)^{-K1/K3}$$

$$=K2-C\left[Y0+q\frac{r}{1+r}v_c(n)v_c(\hat{n})t\right]^{-1}\text{,其中 } C \text{ 为一积分常数} \quad (2.13)$$

现在,我们找到了学习过程中 u_s 层的一种等效显式函数. 并且,

$$\lim_{t\to\infty} u_s=K2$$

学习过程实际上是对 K2 的逼近过程. q 越大,逼近越快. 如果学习时间足够长,u_s 近似等于 K2.

考虑到式(2.4)对(2.1)的简化,u_s 终值$=\varphi(K2)=\varphi\left[(1+r)\frac{(VA(n),VA(\hat{n}))}{\|VA(n)\|\ \|VA(\hat{n})\|}-r\right]$,$u_s$ 终值大于零或者说不等于零的条件是$\frac{VA(n),VA(\hat{n})}{\|VA(n)\|\ \|VA(\hat{n})\|}>\frac{r}{1+r}=\frac{1}{1+1/r}$,r 越大,$u_s$ 对于模式变形容忍性越小,选择性越强,因而可以通过调整 r 来控制 u_s 的模式选择性. 这一点 Fukushima 用代数方法讨论过. 与此是一致的.

u_s 的终值 K2 与 $\hat{n}$、c(v)、r 有关. 采用无教师学习算法时,除 r、c(v)之外,t=0 前的历史,如 a、b 的初始值和学习时各模式的提供顺序,甚至 q 值,都有可能影响到 k_l 平面内 $\hat{n}$ 和 u_c 的选择,也就是各个 u_s 平面所抽取特征的选择,究竟如何影响,或者说是否影响,有待于进一步探讨. 当然这也是应用其它模型时存在的问题,即所谓的初值不稳健性(Mon-robustness)和参数不稳健性问题.

对于有教师算法,$u_s(k_l,n)$的 $\hat{n}$ 和 $u_c(k_{l-1},\hat{n}+v)$是预先选定的,u_s 的终值只受 c(v)和 r 的影响,而与 a、b 的初值及 q 值无关,这一点具有重要意义,它使我们在设计和改进有教师 Neocognitron 时不必考虑 a、b 的初值和 q 值.

这里可以看出,c(v)对于 u_s 终值的重要性. 每个 u_s 平面抽取一种特征,而 Neocognitron 的 c(v)和 r 是固定的,即对不同的特征采用了相同的 c(v)和灵敏度 r. 人在识别视觉模式时对不同的特征是用了不同的加权. 因此,如果对多个模式中每种特征的变形,用某种学习算法(如最小均方误差准则)对 c(v)和 r 进行学习,必然会获得更准确的特征抽取能力. 当用有教师学习算法时,另一种可能的改进措施是提供某种选择 $\hat{n}$ 和 $u_c(k_l-1,\hat{n}+v)$的方法(例如聚类分析). 对于无教师学习算法,还应探讨 $\hat{n}$ 稳定的条件及 q 和可变权初值对 n 的影响. 当然,改进 Neocognitron 是一件大的工作,本文导出的结论只是关于 Neocognitron 权初值和 q、r、c(v)对特征抽取的影响,其它参数的设置尚有待于进一步探讨. 此外,Neocognitron 的改进还可能包括神经元特性、网络结构方面.

结 语:我们先将 Neocognitron 及其学习算法推广到连续时域,用微分法对 Neocognitron 进行了分析,获得了一些结论.式(2.10)给出了 u_s 变化一般规律的微分方程,并推出 u_s 增加的条件和权初始值要满足的条件.式(2.13)给出了有教师算法和无教师算法代表稳定后 u_s 变化的显式函数,讨论了影响 u_s 终值和变化速度的因素.这些结论对于设计和改进 Neocognitron 具有重要意义.同时,我们可以说,微积分作为当前分析神经网络的主要数学方法,对于 Neocognitron 也是行之有效的.

参考文献

1 Barnard E, Casasent D. Shift invariance and the Neocognitron. Neural Networks, 1990;**3**:403—410.

2 Fukushima K. Neocognitron: a self—organizing neural network model for amechanism of pattern recognition unaffected by a shift in position. Biological Cybernetics, 1980;**36**:193—202.

3 Fukushima K, Miyake S. Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recognition, 1982;**15**:455—469.

4 Fukushima K. Analysis of process of vision pattern recognition by the Neocognition. Neural Networks, 1989;**2**:413—420.

5 Fukushima K. A hierarchical neural network capable of visual pattern recognition. Neural Networks, 1988;**1**:119—130.

6 Johnson K, Daniell C, Burman J. Feature extraction in the Neocognitron. ICNN 1988, ppII—117—II—126, San Diego.

7 Lee Y *et al*. Hangul recognition using Neocognitron. INNS Conf, 1990:416—419.

8 Menon M, Heinemann K G. Classification of patterns using a self—organizing neural network. Neural Networks, 1988;**1**:201—215.

9 李星原,舒文豪,洪家荣. Neocognitron 同 BP,LVQ 在手写数字识别中应用的比较.中国首届神经网络大会论文集,1990:808—811.

10 李星原.神经网络在文字识别中的应用.硕士论文,哈尔滨工业大学,1990.

ON THE LEARNING PROCESS OF THE NEOCOGNITRON

Hong Jiarong and Li Xingyuan

(*Department of Computer Science, Harbin Institute of Technology, Harbin* 150006)

Abstract Existing analyses of the Neocognitron failed to discuss the dynamic characteristic during learning because they were all confined to algebraic method. This paper intruduces differentiation to analyse the Neocognitron. For the unsupervised learning, this paper derives a defferential equation to describe u_s, a condition of u_s increasing, and an inequality that the initial value of variable weight of a, b must satisfy. For fixed representation or the supervised learning, we obtain an u_s function of time. We show that, in this case, learning is a process in which u_s approximates a coefficience independence of q and the initial value of weight a, b.

Key words Neural networks, learning rule, self—organizing, Neocognitron.