

复合对向-反向传播人工神经网络模型及其应用

张尊建 余书勤¹ 相秉仁 安登魁

(中国药科大学分析计算中心,¹药理学研究室,南京 210009)

摘要 组合 Kohonen 竞争学习和反向传播学习的优点,本文首次提出了复合对向-反向传播人工神经网络模型,该模型较好地体现了生物神经网络系统信息处理时的自适应、自组织、分布式存贮及并行处理等特点。它保留了反向传播网络的优点,同时较后者更易收敛,计算时间缩短,网络参数设置也更为自由。通过在临床精液检查结果分析中的成功应用,证明了该系统的有效性和可靠性。

关键词 人工神经网络; 学习算法; 模型

反向传播网络(BP)作为人工神经网络研究中的一个重要成果,在许多研究领域得到了成功的应用^[1,2]。应用于模式识别中,它不需要模式分布的假设,算法简单稳定,训练后的网络可对未知模式直接识别,不失为一个非常有效的算法。但 BP 方法也存在着明显的缺点:首先,从数学上看 BP 模型是一非线性优化问题,因而不可避免地存在局部极小问题,虽然可采用改变学习算法、加入随机噪声方法以图跳出局部极小,但并不是总能奏效;第二,算法收敛速度还嫌太慢,通常要迭代几千次或更多,虽然增大学习速率和增加隐含层单元数有时可改善之,但前者与 LMS 算法的原理相悖^[3],权值不稳定,太大还易造成振荡不收敛,后者虽迭代步骤减少,但计算时间降低并不明显,且有研究表明,隐含层单元数太多,网络的拟合能力提高,但预测能力却不然;第三,隐含层单元数的选取尚无理论上的指导,只能根据经验确定,而可选范围又太大。有鉴于此,本文通过组合 Kohonen 的自组织特征映射(SOM)学习^[4]和 BP 学习提出了复合对向-反向传播网络(combined counter-back propagation,CCBP),以期能对上述诸方面有所改善。

1 基本原理

正如人脑由各种相对分工不同的区域组成一样,当一个神经网络接受外界输入模式时,也将会分为不同的区域,各区域对输入模式具有不同的响应特征,且这一过程可自动完成。各神经元之间的连接权值具有一定的分布,最邻近的神经元互相激励,而较远的则互相抑制,更远一些的则又具有较弱的激励作用,在外界刺激最强的地方形成一个墨西哥帽(Bubble),如图 1 所示。在此 Bubble 区中,神经元权向量可自动调节直到与输入向量(刺激信号)的某一最大主分量方向相同为止。据此,可构造一多层网络,当加入输入矢

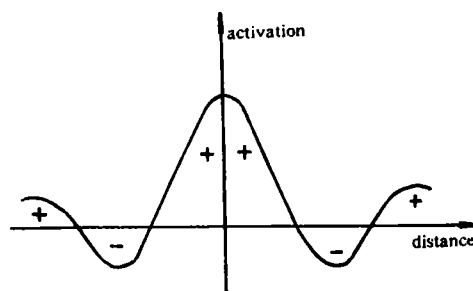


Fig 1. Relationship between the activation of neuron and distance

收稿日期 1996-09-02 江苏省计生委资助课题

量后,在中间层形成 Bubble 区,找出该区域的中心,以一定方式调节该中心邻域(也称胜域)内各神经元与输入层神经元之间的连接权,同时计算中间层各神经元的激活值,并通过与输出层神经元间的连接权传递至输出层产生输出,与期望输出相比较得到输出误差,据此误差以 BP 方法调整胜域内各神经元与输出层神经元之间的连接权,重复此过程,直到满足要求为止。图 2 为本文所用 CCBP 网络的拓扑结构。

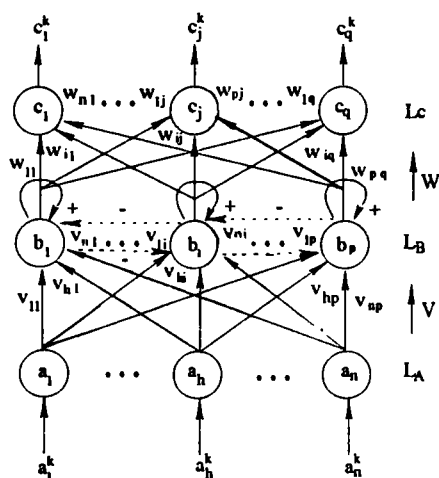


Fig 2. Topological structure of CCBP neural network

2 算法与实现

CCBP 网络采用竞争学习和 BP 学习相结合的方法进行训练。具体算法描述如下:

(1) 归一化所有输入模式 $A_k (k=1, 2, \dots, m)$

$$a_i^k = a_i^k / ||A_k||$$

式中, $k=1, 2, \dots, n$;

(2) 将 L_A 层单元到 L_B 层单元的连接权矩阵 V 及 L_B 层单元到 L_C 层单元的连接权矩阵 W 各元素随机地赋予 $[0, 1]$ 区间的值;

(3) 将 L_B 层中各单元的连接权向量 $V_i (i=1, 2, \dots, p)$ 归一化

$$v_i = v_i / ||V_i||$$

式中, $k=1, 2, \dots, n, i=1, 2, \dots, p$;

(4) 对于每一输入模式对 $(A_k, C_k) (k=1, 2, \dots, m)$, 进行如下操作:

①求 $V_i (i=1, 2, \dots, p)$ 中与 A_k 距离最近的向量 V_g ,

$$||A_k - V_g|| = \min_{i=1}^p ||A_k - V_i||$$

即

$$A_k \cdot V_g = \max_{i=1}^p (A_k \cdot V_i) = \max_{i=1}^p (\sum_{h=1}^n a_i^k v_{ih})$$

②将权向量 V_g 向输入向量 A_k 靠拢, 即

$$v_{hg}^{\text{新}} = v_{hg}^{\text{旧}} + G(t)(a_i^k - v_{hg}^{\text{旧}})$$

其中, $h=1, 2, \dots, n, G(t)$ 为网络学习时间 t 的函数, 由下式定义

$$G(t) = 0.2(1 - t/S)$$

式中 S 为大于 t 的可变正整数;

③重新归一化向量 V_g ,

$$V_{hg} = v_{hg} / ||V_g||$$

式中, $h=1, 2, \dots, n$;

④以 L_B 层胜域 N_g 内各单元到 L_C 层单元的连接权 $w_{ij} (i \in N_g, j=1, 2, \dots, q)$ 将其激活值送到 L_C 层单元, 并计算 L_C 层单元的激活值

$$c_j = f(\sum_{i \in N_g} w_{ij} b_i)$$

式中, $j=1, 2, \dots, q, f$ 为作用函数, 取为 S 型函数, b_i 为胜域内各单元的激活值

$$b_i = \sum_{k=1}^n a_i^k v_{ki}$$

其中 $i \in N_g$;

⑤计算 L_C 层单元的输出误差

$$d_j = c_j(1 - c_j)(c_j^* - c_j)$$

式中, $j=1, 2, \dots, q, c_j^*$ 为希望输出;

⑥调整胜域内各单元到 L_C 层各单元的连接权

$$w_{ij}^{\text{新}} = w_{ij}^{\text{旧}} + \alpha b_i d_j$$

式中, $i \in N_g, j=1, 2, \dots, q, \alpha$ 为修正速率, 通常 $0 < \alpha < 1$;

(5) 对于不同的 $t=1, 2, \dots, T (500 \leq T \leq 10000)$, 重复步骤(4)或直到 $d_j (j=1, 2, \dots, q)$ 且 $k=1, 2, \dots, m$ 足够小。

所用 CCBP 网络方法的源程序以 Turbo

BASIC 语言编写,在 COMPAQ 486/33M 计算机上运行。

3 应用

为考察模型的有效性和可靠性,本文用所建立的 CCBP 方法对前文^[6]数据进行了分析。将全部 134 个样本一半作为训练集,一半作为试验集输入网络进行计算,分别记录迭代开始时、迭代 500 次、1000 次及 3000 次时的网络参数及各模式的输出;将训练集与试验集交换,重复计算。结果表明:当网络训练 500 次后,其分类正确率为 99.3%(N5 号样本错分),预示正确率亦为 99.3%(A17 号样本错分),增加训练次数至 1000 次,判别正确率没有变化,只是输出误差略有减小,说明拟合更精确。当训练次数为 3000 次及以上时,预示正确率依然为 99.3%,而 N5 号样本及 A17 号样本在训练集中则被同时错分,分类正确率降为 98.5%。由此可知,训练 1000 次后的网络已经具有了两类模式判别分类的能力。此结果与用 BP 方法所得结果几乎完全一致^[5]。

4 讨论

Kohonen 层节点数 p 的影响:在 BP 网络中,隐含层节点数的设置不仅影响分类结果,而且对计算时间(收敛速度)有明显影响,有时甚至不收敛。在 SOM 网络和对向传播网络^[6],也都要求 Kohonen 层有足够多的单元以保证能够自适应形成分类边界。本文考察了节点数分别为 4,8,12,16,20,24 时网络的分类能力,结果表明,除 $p=4$ 时预示正确率为 95.5%略低外,其余均为 97.0%,且错误分类几乎都是发生在 N5 和 A17 两个样本上。当节点数由 4 增加到 24 时,迭代相同次数的时间仅增加 1 倍,而不象 BP 网络中要增加近 5 倍。因而 CCBP 网络中中间层节点数的设置并不十分严格,可在 10~20 之间,对分类结果和收敛速度均无明显影响,本文中

取 $p=12$ 。

输出层节点数 q 的设置:考察了两种情况,即 $q=1$ 和 $q=2$ 。结果表明,在其它条件相同时,网络给出的判别结果完全相同,只是每个样本的具体输出有些差异,即拟合的逼近程度不尽相同。本文中取 $q=2$ 。

胜域 N_i 的大小:本文尝试在输出节点的二维阵列分布中,以“半径” R 来确定胜域的大小。当 R 分别等于 1、2 和 3 时,迭代至收敛(以输出误差小于 0.001 判断)分别需要 501 次、3620 次及 1623 次,预示正确率分别为 97.0%,97.0%和 85.1%,且运算时间随着 R 的增加而增加。由此可见, R 的选取对结果有着一定的影响,本文中取 $R=1$ 。另外,随着 Kohonen 层单元数 p 的不同,Bubble 中心 g 的位置也有所变化,但对于相同的计算条件, g 是一定的,并且多能够在数十次迭代之内确定,配合时间函数 $G(t)$ 在迭代初始阶段的快速下降,可以很快地捕捉到输入向量的大致概率结构,加之随后的渐缓调整,使迭代迅速收敛。这也是 CCBP 网络较 BP 网络更快更易收敛的原因。

修正速率 α 的影响:在 BP 网络中,学习速率 α (及 β)的大小对收敛速度及分类结果均有较大影响, $\alpha(\beta)$ 太小,收敛慢; $\alpha(\beta)$ 太大,易造成振荡不收敛。CCBP 网络中的修正速率 α 用于控制 Kohonen 层单元至输出层单元间连接权调整的速度,与 BP 中的 $\alpha(\beta)$ 意义类似。本文考察了修正速率 α 的大小对网络性能的影响。当 α 分别等于 0.01,0.05,0.1,0.3,0.5 及 0.8 时,迭代至收敛所需的步骤分别为 5481 次,1113 次,556 次,185 次,111 次及 70 次,在 $\alpha \leq 0.1$ 时,预示正确率都是 98.5%(A17 样本错分),在 $\alpha \geq 0.3$ 时,预示正确率均为 97.0%(A6 及 A17 样本错分),可能是因为 α 较大,权值调整欠精细所致。综合权衡收敛速度及分类正确性,本文取 $\alpha=0.1$ 。而在 BP 方法中, $\alpha > 0.05$ 即不收敛,相对而言,在 CCBP 网络中, α 的选择范围明显

扩展,且收敛速度大大加快(CCBP 迭代 556 次至收敛耗时约 4 min,相同的数据用 BP 网络需训练 6460 次,耗时约 22 min)。

CCBP 网络借鉴了 BP 网络中以隐含层单元对输入模式实现“内部表示”的处理方法,本质上是相同的。但在 CCBP 网络中,有效的“内部表示”仅限于胜域内各神经元,所以能较 BP 方法更快地捕捉到输入向量的大致概率结构,对于不同类别的样本以不同的胜域表示,权值调整更加容易,因而收敛速度也加快。在计算条件的选取上,除胜域大小(R 通常在 1~2 之间)外,Kohonen 层节点数、输出层节点数、修正速率等参数的设置均不苛刻,计算条件建立更加容易。本文中的成功应用,也为 CCBP 网络引入其它研究领域展示了希望。

参考文献

- 1 范俊波,靳 蕾,史 燕. 边缘检测的一种神经网络方法. 全国首届神经网络及其应用讨论会论文集. 北京. 1989. 56~59
- 2 郭爱克,潘 泓. 神经网络研究概述. 科学,1990,42 (3):176
- 3 靳 蕾,范俊波,谭永东. 神经网络与神经计算机. 原理·应用. 成都:西南交通大学出版社,1991. 146~157
- 4 Kohonen T. Automatic formation of topological maps in self-organizing system. *Proceedings of the 2nd Scandinavian Conf. on Image Analysis*, 1981. 214~220
- 5 张尊建,余书勤,相秉仁等. 人工神经网络在辅助男性生育力检查评价中的应用. 中国药科大学学报,1996,27 (10):598
- 6 Widrow B, Hoff W. Adaptive switching circuits. 1960 *WESCON Convention Record; Part 4*. 1960. 96~104

A New Artificial Neural Network Model: Combined Counter-Back Propagation and its Application

Zhang Zunjian, Yu Shuqin, Xiang Bingren, An Dengkui

Analysis and Computer Center, China Pharmaceutical University, Nanjing 210009

Abstract A new artificial neural network model, combined counter-back propagation(CCBP), was developed by combining all virtues of competitive learning and back propagation learning. The characteristics of CCBP model were studied throughly, and it was showed that CCBP represented well adaptation, self-organization and paralld distributed processing of biological neural network in information processing. In addition to its possessing all the advantages of BP, it was faster to converge and easier to choose the network parameters for CCBP than for BP. The validity of the system was confirmed through its successful application to the analysis of the data from clinical human semen test. Hence, the CCBP may be identified as one of the useful artificial neural network models.

Key words Artificial neural networks; Learning strategy; Model