

# 基于 BP 和统计的混合法前馈型神经网络及其应用\*

姜天戟 袁曾任

(清华大学计算机系 北京 100084)

**摘要** 为避免 BP 算法本身易陷入局部极小值的缺陷,本文将具有新组合激活函数的 BP 法与传统的 BP 方法(标准和带动量项)分别与统计最优化方法相结合组成混合法,将他们分别应用于天气预报和贷款之中,并进行了仿真比较,在预报准确率和学习速度方面获得了比较满意的结果.本模拟程序在 Turbo-Pascal/6.0 环境下编制,在 IBM PC386 和 486 机器上调试通过并运行.

**关键词** 人工神经网络,最优组合函数, Madaline 网络,天气预报.

神经网络的研究在 80 年代掀起第 2 次高潮后,现在已进入相对平稳发展时期.近年来,许多专家、学者对神经网络计算做了广泛的研究并取得了各种各样丰硕的成果.特别是由 Rumelhart 和 McClelland 提出的 BP 算法,它是目前最具有发展前途的算法之一,而且对此领域的发展产生了巨大的推动作用.但是,这种算法本身存在着重大的缺陷:(a) 它不能确保收敛到神经网络整体误差函数的全局极小值.(b) 因为它是由最陡下降算法组成,此算法在目标函数搜索时使用的是固定参数法,所以它不能确保整体误差函数具有单调下降特性,因而它甚至不能确保神经网络整体误差函数收敛到全局极小值.

在许多文献中都提到,可以将随机最优化方法引入以解决这个问题.本文作者正是基于上述思想,提出了一个将 BP 算法和随机最优化方法相结合的混合法,此方法可保证神经网络整体误差函数在很少的步骤上收敛到全局极小值.

本文第 1 部分是混合法介绍;第 2 部分为 BP 和统计最优化方法结合的混合法在天气预报中的应用;第 3 部分为基于 BP 和统计的混合法在贷款中的应用;第 4 部分结论.

## 1 混合法介绍

本文提到的混合法是由 BP 法和随机最优化法组成, BP 算法是现今在神经网络领域

\* 本文研究得到国防科工委国防科技预研基金和航天基金的资助.作者姜天戟,1971 年生,现在美国佐治亚理工学院计算机科学系读博士,主要研究领域为智能控制及人工智能神经网络和应用.袁曾任,1934 年生,教授,主要研究领域为智能控制和人工神经元及其应用.

本文通讯联系人:袁曾任,北京 100084,清华大学计算机系

本文 1995-04-13 收到修改稿

中使用最广泛的算法之一,笔者认为大多数读者对此算法已经了解,故在此不再重述.而随机最优化法(即统计法)是新近提出的一种旨在解决 BP 算法易陷入局部极小值问题的方法,详细介绍如下.

### 1.1 随机最优化法(即统计法)<sup>[1,2]</sup>

为了防止算法在整体误差函数的局部极小点停止,我们将利用随机最优化方法的好的收敛特性,即目标函数在紧空间上可确保收敛到全局极小值.

算法如下:

(a) 在权矩阵的搜索空间任选一初始点  $w^{(0)}$ , 设  $M$  为训练步骤.

(b) 产生高斯(Gaussian)随机数  $\zeta^{(k)}$ , 如果  $w^{(k)} + \zeta^{(k)} \in W$ , 则转步骤(c); 否则, 设  $w^{(k+1)} = w^{(k)}$ , 转步骤(d).

(c) 如果  $E(w^{(k)} + \zeta^{(k)}) < E(w^{(k)})$ , 则设  $w^{(k+1)} = w^{(k)} + \zeta^{(k)}$ ,  $b^{(k+1)} = 0.2 \times b^{(k)} + 0.4 \times \zeta^{(k)}$ , 如果  $E(w^{(k)} + \zeta^{(k)}) \geq E(w^{(k)})$  且  $E(w^{(k)} - \zeta^{(k)}) < E(w^{(k)})$ , 则设  $w^{(k+1)} = w^{(k)} - \zeta^{(k)}$ ,  $b^{(k+1)} = b^{(k)} - 0.4 \times \zeta^{(k)}$  否则 设  $w^{(k+1)} = w^{(k)}$ ,  $b^{(k+1)} = 0.5 \times b^{(k)}$  ( $b^{(0)} = 0$ )

(d) 如果  $K = M$ , 停止整个计算. 如果  $K < M$ , 设  $K = K + 1$ , 转步骤(b).

在一紧空间上,此算法可确保目标函数以概率 1 收敛到全局极小值.

### 1.2 混合法理论简介

假设神经网络设计者的目标是使整体误差函数的值减小到一个小数  $\epsilon$ , 则混合法如下所述. 在混合法训练时, 首先利用 BP 算法进行参数训练. 当整体误差函数  $E(w)$  的值的减小量比给定值  $\epsilon \vee \epsilon$  小时, 为了阻止  $E(w)$  陷入局部极小值, 我们将整体下降算法从 BP 法转为随机最优化法. 如果整体误差函数  $E(w)$  的值的减小量比  $E(w^{(k)})G \vee \epsilon$  ( $0 < G < 1$ ) 大时, 我们将整体下降算法从随机最优化法转为 BP 算法. 此变化是为了加速整体极小化. 同样的过程重复若干次. 当整个步数超过给定数  $M$  时, 计算停止. 此算法有很好的收敛特性, 即它能确保神经网络整体误差函数收敛到全局极小值.

注: (1)  $w^{(k)}$  表示神经网络在第  $K$  步的权矩阵. (2)  $a \vee b$  表示  $Max(a, b)$ . (3) 整体误差函数的减小量比  $E(w^{(k)})G \vee \epsilon$  大, 意味着当前权向量  $w^{(k)}$  已从局部极小点跳到  $E(w)$  的新的峰谷点.

## 2 基于 BP 和统计的混合法在天气预报中的应用

在现代社会中, 天气预报对一个国家的军事和民用以及工农业等部门的管理、经营及决策都有重大的影响, 它直接关系到经济和社会效益. 因此, 希望气象专家能提供一个精确的天气预报. 但是由于周围国家和地区提供天气情况的准确度、用建立数学模型求解的方法遇到精确建模和求解 2 方面的困难, 再加上现有串行计算机的速度和存储能力等问题, 从而造成天气预报不精确. 本文作者试图探索不用精确建立数学模型和求解方程的方法——利用神经网络进行天气预报.

神经网络可以仅通过外界提供的训练样本——过去一段时间内的天气状况及参数来透过表面抽取隐藏在数据后面的本质. 国外学者早已从事这方面的研究工作. 例如, 美国 Neural Ware 公司在 1987 年已开发设计了一个人工神经网络的天气预报系统的模拟程序.

但其采用的是 Madaline 网络,预报的准确率及速率均不尽人意<sup>[3]</sup>,故作者提出将 3 种不同激活函数组成的神经网络应用于天气预报中,以便提高预测的速率及精确度。

## 2.1 Madaline 网络简介

Adaline 网络是自适应线性神经元(Adaptive Linear Neuron)网络的简称,是一个 2 层前馈感知机网络, Madaline 则是由几个 Adaline 组成,它是一个异联想最邻近模式匹配器。其使用 LMS 算法离散学习,按离散时间方式运行。Madaline 通过使输出层各单元的实际输出与期望输出的均方误差最小化来学习。

## 2.2 算法<sup>[4,6]</sup>

本章所提出的天气预报方法实质上为一个神经网络模拟程序,可用于预报未来 3 天(今天、明天和后天)的天气情况。图 1 为此仿真方法中所用的神经网络模型——BP 网络。这个网络有 3 层:输入层、中间层和输出层。输入层有 4 组单元,各自对应于当天绝对气压、当天气压与前一天同时间气压差、此气压差符号变化(正或负)以及当天的风向这些气象要素。输出层有 3 个单元,分别对

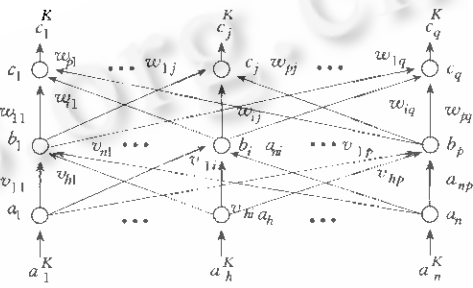


图1

应于当天、明天和后天有无雨的预报输出。中间层的神经元数目可 on-line 的由用户输入(一般 $\leq 10$ )。网络中各神经元可选取 2 种激活函数: Sigmoid 函数或最优组合函数。<sup>[4]</sup>网络的输入均为 0 或 1,输出可为 $[0, 1]$ 之间的任何值。对于一给定的输入模式,如果网络的实际输出匹配于网络的希望输出, BP 网络的连接权值不发生变化;反之,若不匹配,就要采用 LMS 算法,使输入连接权向着减小网络输出误差的方向变化,对于其它输入模式(样本),也同样使用上述过程直至网络收敛。在仿真中,作者共提供了 31 个样本,误差取为均方误差。综上所述,可得算法如下:

(1) 随机地设置初始权值  $W^0$ , 并计算误差函数  $E(w^{(0)})$ , 令  $E1 = E(w^{(0)})$ ;

设置  $M, G, \epsilon, \epsilon_1$  的初值, 设  $K = 0$ ; 选定激活函数(Sigmoid 函数或最优组合函数), 并设定收敛误差项为  $term. error$ 。

(2)  $E2 = E1, K = K + 1$ ; 用 BP 算法修改权矩阵, 设  $w^{(k)}$  为修改后的矩阵,  $E = L(w^{(k)}); E1 = E$ 。

(3) 如果  $E1 < \epsilon$  或  $K = M$ , 则转(12)。

(4) 如果  $|E2 - E1| < \epsilon_1 \vee \epsilon$ , 则转至(2)。

(5)  $E2 = E1, K = K + 1$ ; 用随机最优算法修改权矩阵, 设  $w^{(k)}$  为修改后的矩阵,  $E = E(w^{(k)})$ 。

(6) 如果  $E \leq E2$ , 则令  $E1 = E$ , 转至(8)。

(7) 如果  $K = M$ , 则转至(12); 否则转至(5)。

(8) 如果  $E1 < \epsilon$ , 则转至(12)。

(9) 如果  $|E2 - E1| \leq (G \times E2) \vee \epsilon$ , 转至(7)。

(10) 如果  $K = M$ , 则转至(12)。

(11) 跳转(2)。

(12) 结束.

### 2.3 仿真结果及分析

本文曾提到的 Neural Ware 公司用 Madaline 网络开发设计的天气预报系统,其预报性能如下,在训练周期为 9 233 epoch 后,对当天的预报精确率为 86.4%,对明天和后天的预报精确率分别为 76.3%和 88.1%.而作者采用标准 BP+统计、动量项 BP+统计及最优组合 BP+统计 3 种算法分别构成的网络,在样本数为 31 个,所有样本的值均为 0 或 1 编码及隐含层节点数=8 的情况下,进行仿真.通过仿真结果我们可以看到,预报的准确率还是比较高的,尤其是今天的预报精确率在 95%以上(误差取为 0.1),对明天及后天的预报精确率也为 95%及 90%左右(误差亦为 0.1).再与 Madaline 的结果相比较,我们发现,BP 网络对天气进行预报,从精确率上看比 Madaline 要精确,其结果是令人满意的.尤其值得一提的是,在作者将基于 BP 和统计的混合法应用到天气预报后,所做的仿真计算几乎都很好地避免了 BP 算法易陷入局部极小值的缺点.仿真结果如图 2,样本数为 31 个.

算法		误差			
		0.3	0.2	0.1	
标准 BP+统计	准确率	今天	87.10%	90.32%	96.77%
		明天	83.87%	87.10%	93.55%
		后天	83.87%	83.87%	93.55%
	Epoch	586	2 190	10 273	
动量项 BP+统计	准确率	今天	87.10%	96.77%	96.77%
		明天	87.10%	93.55%	90.32%
		后天	83.87%	83.87%	90.32%
	Epoch	2 019	12 092	21 763	
最优组合 BP+统计	准确率	今天	87.10%	90.32%	93.54%
		明天	83.87%	93.54%	93.54%
		后天	83.87%	83.87%	87.10%
	Epoch	104	3 328	8 109	

图2

$$\alpha=0.1; \quad \beta= \begin{cases} 0.1 & \text{动量项 BP+统计} \\ 0.0 & \text{最优组合 BP+统计} \end{cases}$$

隐含层节点数=8 输入=12 输出=3  
 $\epsilon=\epsilon_1=0.1, G=0.1$

作者在用图 2 中 31 个样本对网络训练完成后,另外取一些样本输入,对天气预测,其结果也是相当满意的.

### 3 基于 BP 和统计的混合法在贷款中的应用

同天气预报相似,“贷款”计算的本质也是一种预估,因此,我们可以设想人工神经网络 ANN(artificial neural networks)将在此方面有所作为.

#### 3.1 问题的提出

我们应设计出一个人工神经网络(ANN),能根据借贷申请人的月收入,生活费用支出,房租水、电、交通费用支出及其它费用支出,来实时地分析这一借贷申请是否合格.如合格,则批准申请给予贷款,否则给予拒绝.

#### 3.2 算法概述

本算法训练时所用的数据实际来自几个贷款信用部门,它们包括成功的贷款例和失败的贷款例.同时,根据历史情况得到一个“准则”,即申请人的生活费用支出近似等于或少于总收入的 30%,同时 3 项总支出之和近似等于或少于总收入的 35%,那样贷款人的申请就认为是合格的,否则不合格.我们在训练前将实际数据变换到 0~1 范围内,这是因为网络的转移函数是 S 型压缩函数.

我们采用 3 层 BP 网络,即一个输入层、一个隐含层、一个输出层,网络尺寸为 4×8×1.输入层转移函数是 S 型压缩函数,结点数为 4 加上偏置结点;隐含层的转移函数是 S 型压缩

函数,结点数是8;输出层转移函数是S型压缩函数,结点数为1.全网采用BP算法+统计训练法.

### 3.3 仿真结果

如下表所示,网络在训练后将样本所给出的信息全部学会,并储存在相应的结点中.当对新数据进行测试时,网络准确给出所需结果.

	标准BP+统计	动量BP+统计	最优组合BP+统计
Epoch	1 720	1 641	1 527

$$\alpha = 0.1; \quad \beta = \begin{cases} 0.1 & \text{动量项 BP+统计} \\ 0.0 & \text{最优组合 BP+统计} \end{cases}$$

$$\epsilon = \epsilon_1 = 0.1 \quad G = 0.1 \quad \text{Error} = 0.01$$

$$\text{隐含层节点数} = 8 \quad \text{输入} = 4 \quad \text{输出} = 1$$

## 4 结 论

在本文中,作者将3种不同激活函数组成的BP和统计的混合算法分别应用于天气预报和贷款之中,其训练结果都是相当令人满意的,有效地避免了BP算法本身易陷入局部极小值的缺陷.但同时我们也应当看出,基于BP和统计的混合法在训练速度上是相当缓慢的,这是由于训练过程中使用了随机数,网络权值是随机改变的缘故.因此,我们还可以从此处着手,对此算法作改进.

### 参考文献

- 1 BABA Nario, KOZAKI Motokazu. An intelligent forecasting system of stock price using neural networks. IEEE, 1-371-377, 1992.
- 2 Solis Francisco J, J B Wets Roger. Minization by random search techniques. Mathematics of Operations Research, 1981, 6: 19~30.
- 3 靳蕃, 范俊波, 谭永东. 神经网络与神经计算机·应用. 成都: 西南交通大学出版社, 1993.
- 4 Yuan Zengren, Shen Xiaohui. A new method for faster backpropagation learning. Submitted to International Conference on Neural Information Processing ICONIP'94 - Seoul. Advances in Modelling & Analysis, A, AMSE Press, 1995, 24(1): 57~64.
- 5 郑君里, 杨行峻. 人工神经网络. 北京: 高等教育出版社, 1992.
- 6 Jiang Tianji, Yuan Zengren. Forward neural network with new activation function and application on the weather forecast. International Conference 'Information Processing' New Orleans (U. S. A.), Nov., 1994. 9~11.

## FORWARD NEURAL NETWORK BASED ON HYBRID ALGORITHM OF BP WITH NEW ACTIVATION FUNCTION AND STATISTICS AND THE APPLICATIONS

Jiang Tianji Yuan Zengren

(Department of Computer Science Tsinghua University Beijing 100084)

**Abstract** In order to avoid the disadvantage of BP algorithm, which is to trap into the local minimum, this paper provides Hybrid Algorithm consisted of BP with new combination activation function and statistics; then this new algorithm is applied to weather forecast and loan and the results are compared with Hybrid Algorithm based on backpropagation (standard and momentum-term) and statistics. Simulation results show that the accuracy and learning rate of prediction are both satisfactory. This program is written in TURBO-PASCAL 6.0 and run on IBM PC/386 and 486.

**Key words** ANN (artificial neural networks), OCAF (optimal combination activation function), MNN (madaline neural networks), weather forecasting.