

文章编号: 1000-2022(2005)03-0398-06

动态学习率神经网络预测气温的尝试

周 慧^{1,2}, 朱 彬¹, 陈万隆¹,
朱国强³, 黄赛群⁴

(1. 南京信息工程大学 应用气象学系, 江苏 南京 210044; 2 湖南省气象台, 湖南 长沙 410007;
3 南京信息工程大学 信息与通信系, 江苏 南京 210044; 4 沅江市气象局, 湖南 沅江 416000)

摘 要: 采用单站气象资料建立动态学习率的神经网络预测模型, 进行逐日气温模拟预测, 并与逐步回归预测模型和固定学习率的神经网络模型比较, 结果表明, 神经网络模型预测能力较好, 但训练时间长。采用动态学习率的网络模型在不损失预测精度的前提下大大减少了训练时间, 为神经网络在气象中的应用提供了一种方法。

关键词: 神经网络; BP算法; 气温预测; 逐步回归

中图分类号: P456.9 TP18 **文献标识码:** A

神经网络在气象上的应用往往是利用固定学习率的神经网络的 BP 算法建模, 多用于月平均气温^[1]、极端气温^[2]、暴雨^[3]等方面。而神经网络 BP 算法, 存在学习过程中易陷入局部极小, 学习收敛速度慢, 学习过程易出现振荡现象和网络存在冗余连接等缺陷^[4]。基于这些缺陷以及计算机硬件因素, 这些神经网络学习时间都很长。本文采用动态学习率的 BP 算法, 减少了网络的训练时间, 提高了预测的时效性, 为神经网络在气象业务中的应用提供了新的思路, 具有重要的应用价值。

1 模型及算法简介

1.1 固定学习率的 BP 算法

BP 网络可实现输入空间到输出空间的非线性映射, 实现从 n 维至 m 维欧氏空间的映射。它采用了优化算法中的梯度下降法, 把一组样本的 I/O 问题变为非线性优化问题, 经迭代运算求解权值, 使误差信号达到要求的程度。隐层的作用使优化问题的可调参数增加, 使解更精确^[5]。BP 神经网络模型一般由输入层、隐层、输出层构成 (图 1), 该方法的数学原理和推导方法见文献 [6]。网络学习矩阵的学习算法主要归结如下:

设网络的学习样本和期望输出 $A_k, C_k (k = 1, 2, \dots, m, \text{为学习矩阵的模式输入对个数})$ 。随机地给出从网络的输入层到隐层的初始连接权系数 V_{hi} , 同时也随机地给出一组隐层到输出层的连接权系数 W_{ij} 以及隐层单元和输出层单元的阈值 θ_i, γ_i , 然后对学习样本的输入和期望输

收稿日期: 2004-05-10 改回日期: 2004-09-23

基金项目: 国家自然科学基金资助项目 (49375248)

作者简介: 周 慧 (1973-), 女, 湖南益阳人, 工程师, 硕士, 研究方向: 大气物理与大气环境. E-mail: zhumo@sohu.com.

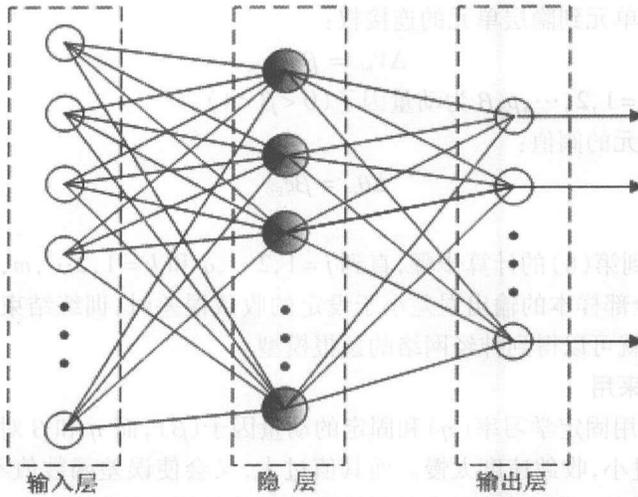


图 1 三层神经网络 BP 型模式

Fig 1 The three-level neural network BPM odel

出 $A_k, C_k (k= 1, 2 \dots, m)$ 进行下列计算^[7]:

(1) 根据连接权矩阵 (初始时刻为给定的一组随机小量) 和学习样本的输入, 计算隐层新的激活值:

$$b_i = f\left(\sum_{n=1}^{h=1} a_n V_{hi} + \theta_i\right) \tag{1}$$

其中 $i= 1, 2, \dots, p$ (p 为隐层节点数), $h= 1, 2, \dots, n$ (n 为输入层节点数)。节点的转移函数为 Sigmoid 函数:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

(2) 计算输出层单元的激活值:

$$C_j = f\left(\sum_{i=1}^p W_{ij} b_i + y_j\right) \tag{3}$$

其中 $j= 1, 2, \dots, q$ (q 为输出层节点数)。在初始时刻, 权系数 W_{ij} 为一组给定的随机小量。

(3) 计算输出层单元的一般化误差:

$$d_j = C_j(1 - C_j)(C_j^k - C_j) \tag{4}$$

其中 $j= 1, 2, \dots, q, C_j^k$ 为输出层单元 j 的期望输出。

(4) 计算隐含层单元相对于每个 d_j 的误差:

$$e_i = b_i(1 - b_i) \sum_{j=1}^q W_{ij} d_j \tag{5}$$

其中 $i= 1, 2, \dots, p$ 。

(5) 调整隐层单元到输出层单元的连接权:

$$\Delta W_{ij} = \eta b_i d_j \tag{6}$$

其中 $i= 1, 2, \dots, p, j= 1, 2, \dots, q, \eta$ 为学习率 ($0 < \eta < 1$)。

(6) 调整输入层单元的阈值:

$$\Delta y_j = \eta d_j \tag{7}$$

其中 $j= 1, 2, \dots, q$

(7)调整输入层单元到隐层单元的连接权:

$$\Delta V_{hi} = \beta u_h e_o \quad (8)$$

其中 $h = 1, 2, \dots, n, i = 1, 2, \dots, p, \beta$ 为动量因子 ($0 < \beta < 1$)。

(8)调整隐层单元的阈值:

$$\Delta \theta_i = \beta e_o \quad (9)$$

其中 $i = 1, 2, \dots, p_o$

(9)重复第(1)到第(8)的计算步骤,直到 $j = 1, 2, \dots, q$ 和 $k = 1, 2, \dots, m$, 计算实际输出与期望输出的误差。当全部样本的输出误差小于设定的收敛误差时,训练结束。根据这些确定的连接权系数和阈值,就可以得到神经网络的预报模型。

1.2 动态学习率的采用

上述 BP 算法采用固定学习率 (η) 和固定的动量因子 (β), 而 η 和 β 对算法成败起着关键作用。若它们的值过小,收敛速度太慢。而其值过大,又会使误差函数值不下降,导致算法不收敛^[8]。针对这一弊端,本文采用动态的学习率的具体处理方法如下:

(1)如果整个学习矩阵上的总误差在权值和阈值更新后增加了,且超过了某个设置的百分数 ζ (典型值为 1% ~ 5%), 则权值更新被取消,学习率 η 和动量因子 β 乘以一个因子 ρ ($0 < \rho < 1$)。

(2)如果总误差在权值和阈值更新后减少了,则权值和阈值更新被接受,学习率 η 和动量因子 β 乘以一个因子 λ ($\lambda > 1$)。

(3)如果总误差的增长率小于 ζ 则权值和阈值更新被接受,学习率 η 和动量因子 β 保持不变。

这一方案根据算法的性能,动态选择最合适的学习率和动量因子。这样既避免了因 η 和 β 值过大导致误差函数值不下降,又避免了 η 和 β 值过小而收敛过慢的问题。而且大大提高了算法的收敛速度。

2 模型的建立和分析

2.1 预测建模

本文采用湖南省沅江气象站 2002 年 1 月 1 日—2002 年 4 月 30 日共 120 d 的实际观测资料,前 100 d (1 月 1 日—4 月 10 日) 的实测资料用作建模的训练样本;后 20 d (4 月 10 日—4 月 30 日) 的资料作为独立样本,用于对模型预测能力的检验。具体应用中,按 02 时、08 时、14 时、20 时将资料分为 4 组,建立 4 个模型分别用来对 02 时、08 时、14 时、20 时的气温作逐日预报,根据这 4 组结果的综合以及当地预报经验作出公众预报。

在选择作为输入矩阵的预报因子时,既考虑影响气温的其他气象要素,也考虑逐日气温本身不同时间尺度的周期变化。逐日气温序列与各实测气象要素普查相关,考虑到模型的实际应用,因子选定前 1 d 同一时刻达到 0.05 相关显著性水平以上的气象要素,最终选定的有 6 个,即前 1 d 同一时刻的本站气压、水汽压、露点温度、5 cm 地温、10 cm 地温和 15 cm 地温。考虑到逐日气温序列本身不同尺度的周期变化,采用达到 0.01 相关显著性水平以上的延拓均生函数^[9]序列。延拓序列的具体计算方法见文献[9]。满足要求的延拓序列有 6 个。

基于上述预报因子的选择采用文献[6]的混合预报模型进行预测建模。将前 100 d 14 时气温序列(实际上只用后 99 个)作为学习矩阵的期望输出。选定的 12 个相关显著因子(6 个其他气象要素,6 个气温延拓序列)作为输入矩阵(为了满足 BP 网络节点函数的条件,先要将

训练样本的数据标准化在 0.1~0.9 之间), 即输入层节点数为 12 输出层节点为 1(相应时刻的待预测气温), 根据多次实验的经验, 取隐层节点数为 10 时效果最好, 初始学习率为 0.05, 动量因子取 0.6, ζ 取 2%; ρ 取 0.5, λ 取 1.5。总体误差取 0.001。将学习矩阵加载到网络输入端, 进行反复的学习训练, 当误差函数趋于稳定并达到规定限值, 训练结束, 得到相应的预测模型。

同样, 采用上述 12 个相关显著的因子建立逐步回归的预测模型以神经网络模型进行比较。

2.2 结果分析

本文以 14 时组数据所建模型和模拟预测结果作比较分析。关于神经网络预报方法跟其他方法如回归、平均和加权集成预报方法在降水量拟合及预测能力方面相比较的优势见文献 [7]。

用逐步回归建立预测模型, 在作者赛扬 2.0 G/256 M 的机器上耗时约 1 min。产生的模型对独立样本作预测的结果见表 1。

表 1 逐步回归模型 4 月 11 日—4 月 30 日逐日 14 时气温预测与实况比较

Table 1 Predicted (stepwise regression model) and observed
daily mean temperatures from Apr 11 to Apr 30

	11日	12日	13日	14日	15日	16日	17日	18日	19日	20日	21日	22日	23日	24日	25日	26日	27日	28日	29日	30日
实际	18.3	24.0	25.2	26.8	28.0	17.9	13.9	14.0	13.5	18.0	23.3	30.0	18.8	12.6	10.8	10.9	12.2	14.8	14.1	15.4
预测	18.8	26.2	28.0	27.7	25.8	18.2	16.1	16.5	15.6	14.4	20.2	28.7	18.2	11.1	14.9	11.3	14.1	17.4	17.2	16.3
误差	0.5	2.2	2.8	0.9	2.2	0.3	2.2	2.5	2.1	3.6	3.1	1.3	0.6	1.5	4.1	0.4	1.9	2.6	3.1	0.9

从表 1 可看出, 逐步回归模型的预测值与实际值拟合得较理想。趋势预报基本正确。预测值的绝对误差 (用实际值减去预测值后取绝对值) 最大为 4.1℃, 最小为 0.2℃。采用文献 [7] 中的 3 种统计评价指标, 分别为: 平均绝对百分比误差 (MAPE) 0.12%, 均方误差 (MSE) 2.22, 平均绝对误差 (MAE) 1.94。

因为神经元模型与初始的权连接系数 (给定的随机小量) 有一定的关系, 所以有必要建立多个预测模型并选取其中预测效果适中的例子进行分析。先用 14 时组的学习矩阵以固定学习率的 BP 算法建立 10 个预测模型, 并分别对 2002 年 4 月 11 日—30 日 14 时的气温作模拟预报。该 10 个模型的建立在作者赛扬 2.0 G/256 M 的机器上平均约经过 2 210 000 次学习训练, 平均每个模型耗时 32 min。选取其中模拟预测效果适中的一例, 预测值与实际值的比较见表 2。

表 2 固定学习率模型 4 月 11 日—4 月 30 日逐日 14 时气温预测与实况比较

Table 2 Predicted (fixed learning rate model) and observed
daily mean temperatures from Apr 11 to Apr 30

	11日	12日	13日	14日	15日	16日	17日	18日	19日	20日	21日	22日	23日	24日	25日	26日	27日	28日	29日	30日
实际	18.3	24.0	25.2	26.8	28.0	17.9	13.9	14.0	13.5	18.0	23.3	30.0	18.8	12.6	10.8	10.9	12.2	14.8	14.1	15.4
预测	15.5	23.4	23.9	27.6	27.5	20.8	13.7	12.6	14.6	15.2	20.8	32.4	20.2	9.9	10.9	11.3	11.7	12.5	11.6	13.6
误差	2.8	0.6	1.3	0.8	0.5	2.9	0.2	1.4	1.1	2.8	2.5	2.4	1.4	2.7	0.1	0.4	0.5	2.3	2.5	1.8

从表 2 可看出, 固定学习率模型的预测值和实际值拟合得较好, 趋势预报基本正确。预测

值的绝对误差最大为 2.8°C , 最小为 0.1°C 。3种统计评价指标分别为: 平均绝对百分比误差 0.090、均方误差 1.83、平均绝对误差 1.55。神经网络的预测能力明显优于相应的回归方法。

然后用 14时组的学习矩阵以动态学习率的 BP算法同样建立 10个预测模型, 并分别对 2002年 4月 11日—30日 14时的气温作模拟预报。该 10个模型的建立在作者的机器上平均约经过 1 380 000次学习训练, 平均耗时 20 min。选取其中模拟预测效果适中的一例, 预测值与实际值的比较见表 3。

表 3 动态学习率模型 4月 11日—4月 30日逐日 14时气温预测与实况比较

Table 3 Predicted (dynamic learning rate model) and observed daily mean temperatures from Apr 11 to Apr 30 $^{\circ}\text{C}$

	11日	12日	13日	14日	15日	16日	17日	18日	19日	20日	21日	22日	23日	24日	25日	26日	27日	28日	29日	30日
实际	18.3	24.0	25.2	26.8	28.0	17.9	13.9	14.0	13.5	18	23.3	30.0	18.8	12.6	10.8	10.9	12.2	14.8	14.1	15.4
预测	20.5	25.5	22.6	28.3	29.3	17.3	15.2	15.7	16.0	18.3	20.5	27.6	20.5	14.5	12.4	12.0	12.0	16.1	12.5	14.6
误差	2.2	1.5	2.6	1.5	1.3	0.6	1.3	1.7	2.5	0.3	2.8	2.4	1.7	1.9	1.6	1.1	0.2	1.3	1.6	0.8

从表 3 可看出, 动态学习率模型预测值和实际值拟合得比固定学习率模型略好, 趋势预报基本正确。预测值的绝对误差最大为 2.8°C , 最小为 0.2°C 。3种统计评价指标分别为: 平均绝对百分比误差 0.089、均方误差 1.70、平均绝对误差 1.54。各评价指标均跟固定学习率模型相差甚微。

上述 3种模型评价指标总结如表 4所示。

表 4 3种预测模型评价指标比较

Table 4 Comparison in the fitting accuracy of the three prediction models

特征指标	平均建模时间 /min	平均绝对百分比误差	均方误差	平均绝对误差
逐步回归预测模型	1	0.120	2.22	1.94
固定学习率网络模型	32	0.090	1.83	1.55
动态学习率网络模型	20	0.089	1.70	1.54

由表 4 中 3种模型的评价指标比较可知, 逐步回归模型建立最快, 预测精度比较低。神经网络模型建立时间比较长, 但预测精度比较高。采用动态学习率的算法, 对预测精度没有显著的提高。但是采用动态学习率后, 算法的收敛速度比固定学习率的算法快, 前者所需要的时间比后者少 $\frac{1}{3}$ 以上。这对神经网络模型参数的调整及进一步研究模型的应用有重要的意义。

3 结 语

本文分析表明, 神经网络 BP多层模式是一种很有效的算法。它对网络模型的学习矩阵采用梯度下降法来进行学习训练, 具有较好的自适应学习能力和非线性映射能力。但 BP神经网络也表现出预测结果不够稳定, 可调参数较多, 误差函数出现振荡等问题。采用动态学习率, 能大大提高算法的收敛速度, 从而对调整网络结构及进一步研究该类模型有很重要的意义。这种动态学习率的网络模型在湖南省沅江市气象局逐日气温预测的实践中进行的客观预报对实际预报起了重要的作用, 是常规预报的补充。同时, 它也为神经网络在其他气象要素客观预报的应用方面提供了一种的方法。

致谢: 感谢南京信息工程大学信息工程系赵远东副教授提供关于神经元隐层节点个数选取和 BP 网络学习方面的指导。

参考文献:

- [1] 严绍瑾, 彭永清, 郭 光. 月平均气温预报的神经网络 BP 型多层映射模式 [J]. 南京气象学院学报, 1995, 18(3): 344-348
- [2] 袁美英, 周秀杰. 神经网络试报极端气温 [J]. 黑龙江气象, 1998, 15(1): 9-11
- [3] 陈百炼. 贵州省暴雨 BP 神经网络预报方法 [J]. 贵州气象, 2001, 25(1): 3-5
- [4] 焦季成. 神经网络系统理论 [M]. 西安: 西安电子科技大学出版社, 1990: 34-36.
- [5] 徐丽娜. 神经网络控制 [M]. 北京: 电子工业出版社, 2003: 18-23
- [6] Jin Long, Luo Ying, Guo Guang et al Study on mixed model of neural network for farmland flood/drought prediction [J]. Acta Meteorologica Sinica, 1997, 11(3): 364-373
- [7] 金 龙, 陈 宁, 林振山. 基于神经网络的集成预报方法研究和比较 [J]. 气象学报, 1999, 57(2): 198-207
- [8] 杨安华, 彭清娥, 刘光中. BP 算法固定学习率不收敛原因分析及对策 [J]. 系统工程理论与实践, 2002, 22(12): 22-25
- [9] 魏凤英, 曹鸿兴. 长期预测的数学模型及应用 [M]. 北京: 气象出版社, 1990: 9-14

Daily Air Temperature Predication Using a Dynamic Learning Rate Neural Network System

ZHOU Hui^{1, 2}, ZHU Bin¹, CHEN Wan-long¹,
ZHU Guo-qiang³, HUANG Sai-qun⁴

(1. Department of Applied Meteorology, NU IST, Nanjing 210044, China

2. Hunan Meteorological Observatory, Changsha 410007, China

3. Department of Information and Communication, NU IST, Nanjing 210044, China

4. Yuanjiang Meteorological Bureau, Yuanjiang 416000, China)

Abstract This paper discusses the construction and application of BP-type multilevel mapping model with dynamic learning rate η . Based on the 2002 observed temperature data of Yuanjiang Meteorological Station in Hunan Province, daily mean temperature prediction is made by using a dynamic learning rate model, a fixed learning rate model and a stepwise regression model. Results suggest that the BP neural network model is superior to the regression model in predictions, and also swift in learning if the dynamic learning rate is adopted.

Key words artificial neural networks; BP algorithm; air temperature prediction; stepwise regression