

人工神经网络预报模型的过拟合研究*

金 龙

(广西壮族自治区气象减灾研究所, 南宁, 530022)

况雪源

(广西壮族自治区气候中心, 南宁, 530022)

黄海洪

(广西壮族自治区气象减灾研究所, 南宁, 530022)

覃志年

(广西壮族自治区气候中心, 南宁, 530022)

王业宏

(南京气象学院, 南京, 210044)

摘 要

针对神经网络方法在预报建模中存在的“过拟合”(overfitting)现象和提高泛化性能(generalization capability)问题,提出了采用主成分分析构造神经网络低维学习矩阵的预报建模方法。研究表明,这种新的神经网络预报建模方法,通过浓缩预报信息,降维去噪,使得神经网络的预报建模不需要进行适宜隐节点数的最优网络结构试验,没有“过拟合”现象,并且与传统的神经网络预报建模方法及逐步回归预报模型相比泛化能力有显著提高。

关键词: 神经网络, 泛化性能, 过拟合现象, 预报建模。

1 引 言

20 世纪 90 年代以来,国内外在气象学科中开展了很多有关神经网络预报建模和气候分析等应用研究^[1~5]。然而随着神经网络方法在大气科学领域应用研究的不断深入,发现人工神经网络方法在实际业务天气预报应用中存在一个很重要的“瓶颈”问题,即在利用神经网络方法进行气象预报建模时,对已知训练样本集的学习训练达到什么样的拟合精度,才能使预报模型对未知样本具有最好的预报能力。该问题的研究不仅关系到在大气科学中能否进一步深入开展有关神经网络方法的业务预报应用,并且也是目前人工神经网络应用理论研究中尚未得到很好解决的有关神经网络模型的过拟合和泛化性能的很关键问题^[6,7]。本文针对实际气象预报问题,尝试从如何构造一种能够全面反映预报因子与预报量关系的低维学习矩阵的“主动”性方法,来探讨神经网络的泛化性能和过拟合现象。

2 影响泛化性能的因子分析

2.1 网络隐层节点

人工神经网络在众多学科的应用研究中已在理论上证明,一个 3 层的前馈网络能够实现任意精度的连续函数映射。并且人们在神经网络的学习算法方面进行了大量的研究,提出了很多关于前馈网络的改进学习算法,使得大多数有关神经网络学习精度的问题都能达到要求。但是在神经网络的预报建模研究中很关键的问题是关于神经网络预报模型的泛化能力,具体在天气预报问题中,就是神经网络天气预报模型根据训练样本建立预报模型后,对未参加学习训练样本(即独立样本)的预报能力,这显然是神经网络预报模型好坏及该方法能否应用于实际业务预报的核心问题。因此,如何提高神经网络预报模型的泛化能力成为神经网络理论研究和实际应用研究最为重要的研究课题。该问题研究的主要内容包括神经网络的网络结构(主要是适宜隐节点数的确定)和网络结构的学习算法对泛化性能和过拟合现象的影响研究^[8~11]。而一般关于网络结构的影响研究认为网络结构增大,容易产生过拟合现象,使网络泛化性能下降。一般只要在网络收敛的条件下,小的网络结构泛化能力强,且训练速度快。并且认为,隐节点数应小于样本数,但是在隐节点数小于

* 初稿时间:2003 年 2 月 13 日;修改稿时间:2003 年 4 月 25 日。
资助课题:国家自然科学基金项目(40075021)。

样本数的情况下怎样确定适宜的隐节点数能使模型的泛化能力达到最大, 至今没有很好的结论^[6, 9, 10]。为了寻找泛化性能好的适宜网络结构, 不同的研究工作给出了一些不同的确定适宜隐节点数的原则性计算公式, 其中文献[12]分析认为, 隐节点的大致取值范围可由以下两个公式估计:

$$k < \sum_{i=0}^n C_i^h \quad (1)$$

$$h = \sqrt{n + m + a} \quad (2)$$

其中 k 为样本数, n 为输入层节点数, m 为输出层节点数, h 为隐层节点数, a 为 1~10 之间的常数, 如果 $i > h$, $C_i^h = 0$ 。而 Mirchandani 等^[13]认为, 隐节点数 h 与输入节点 n 的关系为:

$$h = \log_2 n \quad (3)$$

然而, 目前这种有关神经网络模型泛化性能的研究工作, 大多是一种“被动”方法, 即这些方法主要是在给定学习样本资料的条件下, 去寻找最小的适宜网络结构^[6]。其实质是寻找在最少隐节点数条件下网络模型具有最好的泛化性能。由式(1)~(3)可以看到, 适宜隐节点数的确定实际是与输入和输出节点有关, 在实际的单输出或多输出模型中, 输出节点数是确定的。因此, 可以看到, 隐节点数原则上是与输入节点有关, 即网络规模大小是直接和输入节点有关。而由于对同一个问题, 越小的网络结构, 只要在收敛条件下, 网络模型的泛化能力越强。由此可以推论对于同一个预报问题, 在构造神经网络的学习矩阵时, 输入节点越少可能使网络模型结构越小, 从而可能使网络模型的泛化能力越强。而问题的另一方面是, 在建立一般的气象预报模型时, 一般影响天气气候未来状况的因素较多, 因此在建立预报模型时必须考虑多种预报因子对预报量未来状况的影响。但是如果考虑太多的预报因子时, 不仅神经网络预报模型结构加大, 影响网络模型的泛化性能和模型的误差函数收敛速度。同时, 从统计预报的角度考虑, 众多预报因子之间又往往存在复共线性关系, 并且众多因子中各自所带的噪声信息叠加, 显然也会造成信息重叠、噪声增加, 而影响到预报模型的预报能力。因此可以看到, 如何构造一种既能包含影响未来天气气候变化的多种因素, 又能尽可能减小神经网络学习矩阵的输入节点数是改善神经网络预报模型性能的关键之一。

2.2 网络结构算法

在有关人工神经网络的泛化性能研究中, 为了

获得适宜的网络结构, 也采用了一些网络结构学习算法, 其主要分为神经网络的剪除型结构学习算法和增长型结构学习算法。在剪除型结构算法中, 采用的主要方法是最初设计一个比求解问题大的网络结构^[14, 15], 对预报模型作学习训练, 在学习训练过程中, 不断地去掉作用小的节点和权, 使这些对网络输出影响不大的权值不断被删除, 从而使网络预报模型的泛化能力提高。而增长型结构学习算法与剪除型结构学习算法正好相反, 它是从小的网络结构开始训练, 然后逐渐增大, 直到网络模型即能对训练样本有很好的识别, 又有较好的泛化能力。然而从这两类网络结构算法的学习训练过程可以看到, 其本质上仍属于比较“被动”的方式来研究提高网络模型的泛化性能。即它是对已给定的一个学习矩阵, 试验用剪除型结构学习算法或增长型结构学习算法对网络模型进行训练, 其结果对提高网络模型的训练误差收敛速度是明显的, 因为网络结构庞大, 容易产生冗余节点, 会使训练时间增加, 并且有时还会出现振荡, 不收敛情况。但是这些学习算法对预报模型的泛化能力改进是有限的, 因为对于一个预报系统, 只有构造出一个真正反映系统输入与输出关系的学习矩阵, 并且由第 2.1 节有关隐节点原则性确定公式看到, 如果学习矩阵的输入节点越少, 维数越低, 可以使网络结构的规模越小, 从而可以提高网络模型的泛化性能。基于以上分析, 本文针对实际的气象预报问题, 采用主成分分析方法, 对神经网络预报模型的学习矩阵进行降维去噪处理, 来研究神经网络预报模型的过拟合现象和提高预报模型的泛化性能。

3 预报建模方法

3.1 主成分分析方法

主成分分析方法(PCA)是系统降维和特征提取的一种基本方法, 该方法被广泛应用于大气科学的预报和分析研究工作中^[16, 17]。而主成分分析的基本思想是将原多维序列特征, 通过线性变换从原特征中得到一组数目与原特征相同的新特征。即对于大气科学中的统计预报建模问题, 就是对某一预报对象有 n 个预报因子, $X = (x_1, x_2 \dots x_n)^T$ 可以通过主成分分析方法构造出 n 个新的综合因子变量 $Z = (z_1, z_2 \dots z_n)^T$ 。而这种新的综合因子变量与原因子变量相比有如下特点:

(1) 每一个新的综合因子变量是原因子变量的

线性组合,即

$$z_i = v_{i1}x_1 + v_{i2}x_2 + \dots + v_{in}x_n \quad (i = 1, 2 \dots n) \quad (4)$$

其中 v_{ij} 为特征向量。

(2) 新的综合因子变量之间是相互正交的,即各个因子间的相关系数为零:

$$r(z_i, z_j) = 0 \quad (i, j = 1, 2 \dots n, i \neq j) \quad (5)$$

(3) 各个新的综合因子变量的特征值明确表示了它对于原因子组的方差贡献大小。

显然,将主成分分析得到的新的综合因子变量用于构造神经网络学习矩阵,其优越性十分明显:①可以很直观地去掉特征值为零或接近于零的特征值对应的主分量,因为这些主分量几乎没有包含原变量的信息。②只保留特征值相对较大的且与预报量相关程度高的主分量,降维作用直观显著。③一般前几个主分量通常具有更大的方差贡献且与预报量相关较好。④由于主分量间是正交的,所以最后确定的几个特征值大且与预报量相关高的主分量所构成的学习矩阵,不会有多余的重复信息噪声影响。而一般直接根据原始因子构造预报模型学习矩阵时,舍弃的预报因子信息无法在最终的预报方程中体现,而保留的预报因子由于都与预报量(同一对象)存在较好的相关关系,所以这些保留的预报因子间往往存在较高的相关关系,造成有用信息的重复,直接造成学习矩阵的规模增大,对预报量无用的“噪声”也相应增加,直接影响神经网络模型的训练速度和泛化性能。由上述分析不难看出,利用主成分分析方法构造神经网络学习矩阵,能很好地保留原全部因子的有用信息,对原预报因子矩阵的降维去噪作用直观明显。

3.2 神经网络方法

目前,在多种人工神经网络模型中,大气科学中应用最为广泛的是前馈网络,其基本算法已有不少文献作过介绍,因此该方法的详细计算推导过程见文献[3]。本文主要讨论前馈网络模型的泛化性能和“过拟合”问题。而一般一个3层前馈网络计算过程可以简单概括为:

模型的最终预报计算公式:

$$\hat{y}_j = f\left(\sum_{i=1}^p b_i w_{ij} + y_j\right) \quad (6)$$

式中 b_i 是隐含层到输出层新的激活值, w_{ij} 是隐含层到输出层的连接权系数,初始时刻为一组随机数, y_j 为输出层单元阈值, f 取 Sigmoid 函数:

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (7)$$

前馈网络模型学习矩阵训练过程也可简单归结为:

(1) 随机给定输入层到隐含层和隐含层到输出层的连接权 w_{ij} , v_{hi} , 并设定总体收敛误差 ε 。

(2) 根据学习矩阵样本对网络进行有导师学习训练,计算网络的实际输出与期望输出的误差,并采用前馈网络的误差逆传播学习算法,调整输入层到隐含层和隐含层到输出层的连接权系数。

(3) 当模型输出的收敛误差大于 ε 时,转到步骤(2),否则训练结束,并根据连接权系数和阈值由式(1)计算预报值。

4 实例分析

本文对多个实际预报问题进行了计算分析,限于篇幅,并为了避免有意挑选好的预报个例,本文给出了6、7和8月降水量3个预报个例计算结果。其计算过程为:首先利用主成分分析构造神经网络的学习矩阵,然后进一步通过对学习矩阵的训练,建立预报模型,再分别对3个预报模型的泛化性能、过拟合现象作实例分析。

预报个例1:以广西8个代表站(南宁、柳州、桂林、资源、北海、百色、梧州和玉林)6月平均降水量作为预报对象,计算找出上年1~12月和当年1~4月500 hPa月平均高度场和月平均海温场达到0.05相关显著水平的预报因子,结果共得到32个预报因子,其中14个海温场因子,18个高度场因子。首先直接对32个因子按常规的逐步回归方法建立预报方程,并且为了统一的对比分析,取预报量1958~1996年39 a样本作为预报建模样本,1997~2002年样本作为独立样本对预报方程进行检验。当 F 值取6,从32个预报因子中筛选出6个预报因子,得到预报方程:

$$\hat{y} = -581.97 + 2.41x_6 + 1.02x_{15} + 0.84x_{17} + 5.66x_{26} - 5.93x_{29} - 1.89x_{31} \quad (8)$$

该方程的复相关系数为0.9177($n=39$),方程对历史样本的拟合平均相对误差为9.970%,平均绝对误差为24.68 mm,该预报方程对1997~2002年6 a独立样本预报的平均相对误差为29.03%,平均绝对误差为87.39 mm。

4.1 主分量学习矩阵模型的泛化性能

根据3.1节的主成分分析方法,对这同样的32

个预报因子作主成分分析, 共计算得到 32 个主分量系数。以主分量特征值相对较大及主分量与预报量相关系数高作为取舍标准, 以确定神经网络学习矩阵的输入。实际计算结果发现, 第 1 主分量与预报量的相关系数为 0.73, 相应的特征值也最大为 9.41, 第 2 主分量与预报量的相关系数为 -0.48, 特征值为 3.02, 而第 5 主分量与预报量的相关系数为 0.18, 其特征值也相对较大, 为 1.89, 其余 29 个主分量与预报量相关系数均小于等于 0.17, 且特征值也相对较小。因此本文以特征值相对较大, 与预报量相关系数最大的第 1, 2 和 5 共 3 个主分量来构造神经网络的学习矩阵。为了分析采用主分量构造神经网络学习矩阵预报模型的泛化性能和过拟合情况, 在表 1 中给出了对该学习矩阵从 5000 到 30000 次学习训练后神经网络预报模型的计算结果。其中隐节点数是根据式 (2) 确定, 并考虑到训练样本长度为 39, 而一般认为隐节点数应远小于训练样本数, 所以本文的研究只取式 (2) 中 a 值分别取 1 和 5。因而, 以 3 个主分量构造的神经网络预报模型其最小规模隐节点数为 3, 中等规模的隐节点数为 7。而为了统一比较, 神经网络预报模型的学习因子和动

量因子也统一取 0.9 和 0.7。由表 1 的计算结果可以清楚地看到, 在最小的网络模型结构 3-3-1 的情况下, 无论训练次数从 5000 到 30000 次如何变化, 不仅预报模型对历史样本的拟合误差基本保持不变(平均相对误差的最大值与最小值的差值, 即极差为 0.35%, 平均绝对误差的极差为 0.46 mm)。并且, 对 6 a 独立样本的预报误差也变化很小, 平均预报绝对误差的极差 0.48 mm, 平均预报相对误差的极差为 0.38%。并且, 由这 3 个主分量得出的主分量神经网络预报模型, 对 6 a 独立样本的预报平均相对误差和平均绝对误差分别为 14.50% 和 47.37 mm(见表 1), 该预报结果比前面同样根据 32 个预报因子采用逐步回归预报方程对 6 a 独立样本的预报平均相对误差 29.03% 和平均绝对误差 87.39 mm 提高了近一倍。进一步分析中等规模网络结构 (3-7-1) 模型的计算结果(见表 1) 同样可以看到, 随着训练次数从 5000 次到 30000 次的改变, 该预报模型对历史样本的拟合和 6 a 独立样本的预报结果与上述 (3-3-1) 网络结构模型预报结果非常相似, 没有出现“过拟合”现象, 独立样本的预报精度比回归方程也提高了近一倍, 同样有很好的泛化能力。

表 1 预报个例 1 的主分量学习矩阵模型训练和预报结果

Table 1 Prediction errors of the trainings and fitting on Principal component ANN model for sample I

网络结构	训练次数	收敛误差	拟合结果		预报结果	
			相对误差(%)	绝对误差(mm)	相对误差(%)	绝对误差(mm)
3-3-1	5000	0.004077	12.20	28.01	14.29	47.11
	10000	0.004050	12.09	27.86	14.40	47.25
	15000	0.004033	12.01	27.76	14.48	47.34
	20000	0.004019	11.95	27.69	14.55	47.42
	25000	0.004008	11.90	27.62	14.61	47.51
	30000	0.003999	11.85	27.55	14.67	47.59
平均值			12.00	27.75	14.50	47.37
极差			0.35	0.46	0.38	0.48
3-7-1	5000	0.004031	12.07	27.88	14.38	47.10
	10000	0.004012	12.97	27.75	14.52	47.31
	15000	0.003999	12.90	27.65	14.53	47.30
	20000	0.003988	11.85	27.53	14.46	47.15
	25000	0.003976	11.79	27.41	14.39	47.95
	30000	0.003962	11.74	27.30	14.33	47.76
平均值			11.89	27.59	14.44	47.10
极差			0.33	0.58	0.20	0.55

4.2 回归因子学习矩阵模型的泛化性能和过拟合现象

为了进一步对比分析, 再将由上述同样 32 个预报因子采用逐步回归方法筛选出的 6 个因子(即式 (8) 的 6 个预报因子) 作为神经网络学习矩阵输入,

来建立回归因子的神经网络预报模型, 并分析这种神经网络模型是否会出现“过拟合”现象。为了客观比较, 回归因子神经网络预报模型的学习因子、动量因子等参数、样本数与 4.1 节主分量神经网络预报模型完全一样, 并同样对学习矩阵作 5000 次到

30000 次训练。表 2 给出了这两种网络结构模型的计算结果,由表 2 看到,无论是最小的网络结构(6-4-1)和中等规模的网络结构(6-8-1),随着训练次数的增加(从 5000 到 30000 次),预报模型对训练样本的拟合精度明显提高,两种模型对历史样本的拟合平均绝对误差分别从 26.15 mm 下降到 17.09 mm 和从 24.98 mm 下降到 13.74 mm。然而与此同时,这两种预报模型对 6 a 独立样本的预报误差则出现完全相反的趋势,即随着训练次数的增加,拟合精度的提高,两种网络结构预报模型对独立样本预报平均绝对误差分别从 83.39 mm 增加到 105.44 mm 和从 88.19 mm 增加到 128.29 mm。独立样本的平均预报相对误差也是如此,出现了非常明显的“过拟合”现象。并且从计算结果看到,用回归因子构造的学习矩阵神经网络预报模型对历史样本的拟合精度可以远高于主分量神经网络预报模型,比如,回归因子学习矩阵网络结构为 6-8-1 的预报模型作 30000 次训练后,对历史样本的拟合平均相对误差精度达到 5.75% 几乎比主分量神经网络

预报模型的 11.74% 提高了一倍,但是它对 6 a 独立样本的预报平均相对误差(45.01%)比主分量神经网络预报模型(14.33%)增大了 2 倍。由此可以看出,回归因子神经网络预报模型的这种“过拟合”现象,确实难以使神经网络方法在大气科学中进行业务预报的应用,因为这时候事先无法知道训练次数达到多少,训练误差的精度达到什么程度,模型具有最好的预报能力。这也是目前神经网络应用研究理论最为关心和尚未解决的关键问题^[6~9]。但是从本文的计算结果看到,由本文构造的主分量神经网络学习矩阵预报模型,不仅没有随着训练次数变化和隐节点变化产生“过拟合”现象,并且泛化能力比逐步回归神经网络预报模型和逐步回归预报方法显著提高。从实际应用角度考虑,这 3 种模型都是在完全相同的 32 个预报因子基础上建立的,显然这样的结果对于改进实际业务预报的准确性具有重要意义。并且可以为推动神经网络方法的实际业务天气预报应用提供重要依据和方法。

表 2 预报个例 1 的回归因子学习矩阵模型训练和预报结果

Table 2 Prediction errors of the trainings and fitting on regression ANN prediction model for sample I

网络结构	训练次数	收敛误差	拟合结果		预报结果	
			相对误差(%)	绝对误差(mm)	相对误差(%)	绝对误差(mm)
6-4-1	5000	0.003383	11.06	26.15	29.65	84.54
	10000	0.003296	10.74	25.62	29.11	83.39
	15000	0.002976	9.95	24.13	29.39	84.11
	20000	0.002331	8.44	20.88	31.21	88.13
	25000	0.002049	8.11	19.13	33.24	98.69
	30000	0.001727	7.35	17.09	34.81	105.44
平均值			9.28	22.17	31.24	90.72
极差			3.71	9.06	5.70	22.05
6-8-1	5000	0.003068	10.59	24.98	30.83	88.19
	10000	0.002414	9.03	21.52	34.86	99.11
	15000	0.001919	8.12	18.71	38.86	115.15
	20000	0.001451	6.83	16.06	41.95	122.61
	25000	0.001252	6.38	15.10	44.01	126.61
	30000	0.001038	5.75	13.74	45.01	128.29
平均值			7.78	18.35	39.25	113.33
极差			4.84	11.24	14.18	40.10

4.3 预报个例 2 和 3 的计算分析

预报个例 2 是以广西 8 个站(站点同预报个例 1)7 月份平均降水量作为预报量,并计算得到前期月平均海温场、500 hPa 高度场共 26 个预报因子(相关显著水平达到 0.05 以上)。同样对这 26 个预报因子,先用逐步回归方法, F 值取 3, 可以自动筛选得到 6 个因子的回归预报方程:

$$\hat{y} = -1568.88 + 1.10x_3 + 2.49x_7 - 1.47x_{14} - 4.42x_{17} + 2.08x_{18} + 1.28x_{24} \quad (9)$$

该预报方程的复相关系数为 0.8772, 方程对历史样本的拟合平均相对误差、平均绝对误差分别为 18.02% 和 36.69 mm, 而方程对 6 a 独立样本的预报平均相对误差和平均绝对误差为 22.35% 和 57.49 mm。同样再采用与 4.1 节完全相同的方法,

进行主成分神经网络和回归因子神经网络的预报建模分析。表 3 给出了这两种神经网络预报模型的计算结果和独立样本的预报结果, 其主要结论与预报个例 1 完全相似:

① 以主分量神经网络预报模型的泛化能力最好, 6 a 独立样本的预报平均相对误差为 14.8%, 平均绝对误差为 41.0 mm。并且该模型随着网络结构规模的增大或训练次数的增加没有出现“过拟合”

现象。

② 6 个因子的逐步回归神经网络预报模型, 出现了明显的“过拟合”现象, 即随着训练次数的增加或网络结构规模的加大, 预报模型的拟合精度明显提高, 拟合精度最好的平均相对误差、平均绝对误差分别为 12.2% 和 25.1 mm, 明显低于主分量神经网络模型(16.0% 和 34.0 mm), 但是模型的泛化能力正好相反, 主成分神经网络预报模型 6a 独立样本

表 3 预报个例 2 的主分量、回归因子学习矩阵模型训练和预报结果比较

Table 3 Comparison of the Principal component and regression ANN prediction models for sample II

主分量因子	训练次数	收敛误差	拟合结果		预报结果	
			相对误差(%)	绝对误差(mm)	相对误差(%)	绝对误差(mm)
3- 3- 1	5000	0.003996	17.04	35.05	14.76	40.96
	10000	0.003970	17.01	34.99	14.87	41.26
	15000	0.003953	16.96	34.91	14.89	41.30
	20000	0.003939	16.93	34.86	14.91	41.35
	25000	0.003924	16.92	34.83	14.95	41.44
	30000	0.003898	16.93	34.81	14.99	41.49
	平均值			16.97	34.91	14.90
最大值(A)			17.04	35.05	14.99	41.49
最小值(B)			16.92	34.81	14.76	40.96
A - B			0.12	0.24	0.23	0.53
3- 7- 1	5000	0.003944	16.28	34.27	14.07	39.13
	10000	0.003877	16.51	34.35	14.56	40.35
	15000	0.003830	16.63	34.39	14.81	40.94
	20000	0.003788	16.67	34.32	14.96	41.25
	25000	0.003746	16.67	34.21	15.06	41.40
	30000	0.003702	16.64	34.07	15.12	41.42
	平均值			16.57	34.27	14.76
最大值(A)			16.67	34.39	15.12	41.42
最小值(B)			16.28	34.07	14.07	39.13
A - B			0.39	0.32	1.05	2.29
回归因子	训练次数	收敛误差	拟合结果		预报结果	
			相对误差(%)	绝对误差(mm)	相对误差(%)	绝对误差(mm)
6- 4- 1	5000	0.003829	17.20	35.17	21.65	56.42
	10000	0.003415	16.61	32.94	20.67	53.61
	15000	0.003227	15.92	31.11	20.01	51.80
	20000	0.003110	15.58	30.27	19.91	51.37
	25000	0.002985	15.23	29.43	19.78	50.67
	30000	0.002825	14.87	28.65	19.92	50.18
	平均值			15.90	31.26	20.32
最大值(A)			17.20	35.17	21.65	56.42
最小值(B)			14.87	28.65	19.78	50.18
A - B			2.33	6.52	1.87	6.24
6- 8- 1	5000	0.003866	17.07	35.08	21.26	55.65
	10000	0.003416	16.46	32.86	20.60	53.07
	15000	0.003084	15.61	30.85	20.20	52.19
	20000	0.003109	14.19	29.88	19.37	47.80
	25000	0.002853	13.27	27.93	20.83	51.47
	30000	0.002474	12.27	25.19	20.96	52.85
	平均值			14.81	30.29	20.54
最大值(A)			17.07	35.08	21.26	55.65
最小值(B)			12.27	25.19	19.37	47.80
A - B			4.80	9.89	1.89	7.85

的预报平均相对误差和绝对误差为 14.8% 和 41.0 mm, 而 6 因子的逐步回归神经网络模型中对 6 a 独立样本预报精度最高的预报模型预报平均相对误差和预报平均绝对误差为 19.8% 和 51.3 mm。

预报个例 3, 以广西 8 个站(站点同预报个例 1, 2) 8 月平均降水量作为预报对象, 采用与以上预报个例 1 和 2 完全一样的计算分析方法, 进行了相同的计算对比分析, 所得结果与预报个例 1 和 2 完全

相似(见表 4), 限于篇幅, 不再赘述。

从本文的 3 个预报个例计算结果看到, 利用相同的预报因子, 分别建立逐步回归方程和主成分神经网络模型的主要差别在于: 回归方程对历史样本的拟合与主分量神经网络模型相差不大, 但是对于独立样本的预报精度则主分量神经网络模型的预报精度明显高于逐步回归方程。其原因主要是因为, 回归方程从众多因子中筛选出的若干个预报因子建

表 4 预报个例 3 的主分量、回归因子学习矩阵模型训练和预报结果比较

Table 4 Comparison of the Principal component and regression ANN prediction models for sample III

主分量因子	训练次数	收敛误差	拟合结果		预报结果	
			相对误差(%)	绝对误差(mm)	相对误差(%)	绝对误差(mm)
3- 3- 1	5000	0.0024335	14.02	23.18	14.70	26.92
	10000	0.0024276	13.94	23.06	15.09	27.30
	15000	0.0024222	13.86	22.96	15.43	27.65
	20000	0.0024158	13.80	22.87	15.54	27.75
	25000	0.0024058	13.74	22.78	15.51	27.66
	30000	0.0023946	13.67	22.69	15.40	27.46
	平均值			13.84	22.92	15.28
最大值(A)			14.02	23.18	15.54	27.75
最小值(B)			13.67	22.69	14.70	26.92
A - B			0.35	0.49	0.84	0.83
3- 7- 1	5000	0.0024289	13.90	23.29	14.71	26.79
	10000	0.0024191	13.87	23.14	15.14	27.24
	15000	0.0024004	13.83	23.01	15.54	27.63
	20000	0.0023606	13.77	22.79	15.62	27.44
	25000	0.0022998	13.70	22.44	14.95	25.88
	30000	0.0022451	13.68	22.16	14.96	25.49
	平均值			13.79	22.81	15.15
最大值(A)			13.90	23.29	15.62	27.63
最小值(B)			13.68	22.16	14.71	25.49
A - B			0.22	1.13	0.91	2.14
回归因子	训练次数	收敛误差	拟合结果		预报结果	
			相对误差(%)	绝对误差(mm)	相对误差(%)	绝对误差(mm)
6- 4- 1	5000	0.0029440	16.80	25.27	28.08	46.99
	10000	0.0027887	16.56	24.72	27.00	44.14
	15000	0.0026915	16.28	24.07	27.04	44.50
	20000	0.0025728	15.88	23.29	27.84	47.03
	25000	0.0024624	15.54	22.57	28.41	49.09
	30000	0.0023293	15.26	21.96	28.32	49.68
	平均值			16.05	23.65	27.78
最大值(A)			16.80	25.27	28.41	49.68
最小值(B)			15.26	21.96	27.00	44.14
A - B			1.54	3.31	1.41	5.54
6- 8- 1	5000	0.0029265	16.54	25.17	28.33	47.23
	10000	0.0027937	16.38	24.71	28.18	46.62
	15000	0.0025467	15.27	24.32	30.15	49.66
	20000	0.0018686	13.71	21.37	34.34	59.70
	25000	0.0014712	12.27	18.52	36.73	66.50
	30000	0.0011961	11.17	16.68	38.71	70.91
	平均值			14.22	21.79	32.72
最大值(A)			16.54	25.17	38.71	70.91
最小值(B)			11.27	16.68	28.18	46.62
A - B			5.37	8.49	10.53	24.29

立的方程, 其各预报因子间仍存在较高的相关性导致的复共线性问题, 引起噪声重叠, 而主分量神经网络模型的主分量之间是正交的, 相关为零, 没有复共线性关系。另外, 回归方程筛选因子时, 没有进入方程的预报因子中的有用预报信息全部被丢弃, 而主分量神经网络预报模型舍弃的原则是使得最后的主分量神经网络模型是由几个方差贡献大, 与预报量相关高且它们之间相关为零、没有复共线性关系的主分量组成, 所以它比回归方程预报精度高, 稳定且没有明显的“过拟合”现象。

另外, 本文在进一步对比回归因子建立的神经网络模型与主分量神经网络模型发现, 其最重要的差异在于主分量神经网络模型没有明显的“过拟合”现象, 而回归因子的神经网络预报模型的确出现了明显的“过拟合”现象, 即如果按这种一般的神经网络预报建模方法建立预报模型会产生拟合精度越高, 预报精度越差, 而使得在实际预报中无法确定最终的有效预报模型。

5 结 论

(1) 采用主动性方法构造能真实反映预报系统输入到输出之间内在联系的低维主分量神经网络学习矩阵, 是提高神经网络预报模型泛化性能, 防止和减轻过拟合现象的一种新途径。不仅为神经网络方

法在大气科学的实际预报应用提供了重要依据, 也适用于其他学科神经网络预报建模的应用和研究。

(2) 在构造神经网络学习矩阵的方法上, 采用主成分提取方法是气象统计预报的神经网络预报建模新方法。它可以浓缩众多预报因子的有用信息, 起到显著的降维作用, 直接使得网络结构规模变小。同时又由于各主分量的正交性可以使学习矩阵的输入节点之间减少重复信息和噪声。

(3) 实例计算结果表明, 在依据相同预报因子条件下, 主成分学习矩阵神经网络预报模型的预报精度显著高于回归学习矩阵神经网络预报模型和传统的回归预报方法。并且主成分学习矩阵模型在实际预报中不需要进行最佳网络结构试验和确定训练精度达到多少的问题, 比以往的适宜隐节点确定方法和网络结构学习算法有显著的优越性。

(4) 有关神经网络预报建模的最新研究认为^[6], 正则化方法对提高神经网络泛化能力有很好的效果, 但是该项研究是在假设隐含层节点数确定的情况下得到的结果, 并指出如何确定结构问题, 同样是泛化问题需要进一步研究的问题。然而本文的研究方法则不需要再作进一步的网络结构问题研究, 并且泛化能力显著提高, 是一种客观实用的神经网络预报建模方法。

参考文献

- 1 Dean, Andrew R, Brian H Fiedler. Forecasting warm season burr off low clouds at the San Francisco international airport using linear regression and a neural network. *J Appl Meteor*, 2002, 41(6): 629~ 639
- 2 Hsieh William W. Nonlinear canonical correlation analysis of the tropical Pacific climate variability using neural network approach. *J Climate*, 2001, 14(12): 2528~ 2539
- 3 Jin Long, Ju Weimin, Miao Qilong. Study on Anmr based multi step prediction model of short term climatic variation. *Adv Atmos Sci*, 2000, 17(1): 157~ 164
- 4 胡江林, 张礼平, 宇如聪. 神经网络模型预报湖北汛期降水量的应用研究. *气象学报*, 2001, 59(6): 776~ 783
- 5 金龙, 苗春生, 陈宁. 定性和定量长期预报模型的综合分析. *气象学报*, 2000, 58(4): 479~ 484
- 6 Wu Yan, Zhang Liming. A new regularization learning method for improving generalization capability of neural network, proceedings of 4th world congress on intelligent control and automation, Shanghai, 2002, 3: 2011~ 2015
- 7 Saito K, Nakano R. Second order learning algorithm with squared penalty term. *Neural Computation*, 2000, 12(3): 709~ 729
- 8 Vapnik V, Levin E, Lecun Y. Measuring the VC-dimension of a learning machine. *Neural Computation*, 1994, 6: 851~ 876
- 9 江学军, 唐焕文. 前馈网络泛化性能的系统分析. *系统工程理论与实践*, 2000, 20(8): 36~ 40
- 10 彭汉川, 甘强, 韦钰. 提高前馈神经网络推广能力的若干实际方法. *电子学报*, 1998, 26(4): 116~ 119
- 11 Atiya A, C Ji. How initial conditions affect generalization performance in large networks. *IEEE Trans on neural network*, 1997, 8(2): 448~ 451
- 12 袁曾任. *人工神经网络及其应用*. 北京: 清华大学出版社, 1999. 118~ 131
- 13 Mirchandani Gagan, Gao Wei. On hidden nodes for neural nets. *IEEE Transactions on circuits and systems*, 1989, 36(5): 661~ 664
- 14 方剑, 席裕庚. 神经网络构造设计的准则和方法. *信息与控制*, 1996, 25(3): 156~ 164
- 15 王科俊, 王克成. *神经网络建模、预报与控制*. 哈尔滨: 哈尔滨工程大学出版社, 1996. 90~ 101

- 16 张邦林, 丑纪范. 经验正交函数的稳定性研究. 气象学报, 1992, 50(3): 342~ 245
17 丁裕国. 经验正交函数展开气象场收敛性的研究. 热带气象, 1998, 4(4): 316~ 325

STUDY ON THE OVERFITTING OF THE ARTIFICIAL NEURAL NETWORK FORECASTING MODEL

Jin Long

(*Guangxi Research Institute of Meteorological Disasters Mitigation, Nanning 530022*)

Kuang Xueyuan

(*Guangxi Center of Climate, Nanning 530022*)

Wang Haihong

(*Guangxi Research Institute of Meteorological Disasters Mitigation, Nanning 530022*)

Qin Zhinian

(*Guangxi Center of Climate, Nanning 530022*)

Wang Yehong

(*Nanjing Institute of Meteorology, Nanjing 210044*)

Abstract

With the application of the artificial neural network (ANN) in the field of Atmospheric Science, a “bottle – neck” was found while the artificial neural network model was applied for weather forecast: the fitting precision of training sample could not be definitely determined to make the model showing its best forecasting capability. It was a key problem to be solved on the overfitting and generation capability of the ANN application theory area. Study on this problem is necessary for the further operating application of ANN in the field of Atmospheric Science. A new forecasting model has been proposed for model establishment by means of making a low – dimension ANN learning matrix through principal component analysis (PCA – ANN). The monthly rainfall of June, July and August were forecasted by using PCA – ANN, R – ANN (regression artificial neural network model) and SR (stepwise regression model) respectively. The results showed as follows:

(1) The matrix built with active method, which could really reflect the internal relationship between the input and output, is a new method to develop the quality of generation capability and avoid or decrease the overfitting in ANN model. It also could be applied in other fields.

(2) The PCA – ANN model could distill the main information of many factorials and remarkably decrease the dimension size of matrix. The noisy and overlap information of the inputting factorials were also reduced because of the orthogonality of each principal component.

(3) The forecasting accuracy of PCA – ANN model was obviously higher than that of R – ANN model and SR model. Moreover, the PCA – ANN model needs not to test the best network structure and training precision. So it showed remarkable superiority compare with the method of finding suitable hidden nodes and the training course of network structure.

Key words: Artificial neural network, Generalization capability, Overfitting, Establishment of a forecasting model.