

# 交叉熵神经网络及其在闽北大雨以上降水预报中的应用

吴木贵<sup>1</sup> 江彩英<sup>2</sup> 张信华<sup>2</sup> 赖荣钦<sup>1</sup>

## 摘要

基于误差平方和最小化准则的 BP 神经网络(ANN-MSE)并不适合解决小概率天气事件的预报问题,引进一种改进的以交叉熵函数为目标函数的神经网络方法(ANN-CE),该法是一个三层反向传播神经网络,其输出层只用一个节点.利用 2003—2008 年的 ECMWF 预报场资料,把该法用于福建省南平市 4—6 月部分大雨或以上降水 96 h 预报中,分别用原始因子和 PCA 降维后的主因子建立了 ANN-CE 预报模型和 ANN-MSE 预报模型,用这些模型对 2009—2010 年独立样本进行了试报.测试结果显示主因子预报模型 TS 评分比原始因子预报模型高且漏报次数少,其中,主因子 ANN-CE 预报模型的 TS 评分和漏报率分别是 0.51 和 0.17,其性能是所有模型中最好且最为稳定的,是一种适合于小概率事件预报的方法.

## 关键词

BP 神经网络;交叉熵;分类预报;小概率事件

中图分类号 TH71;TG803

文献标志码 A

收稿日期 2010-10-08

资助项目 福建省自然科学基金(2008J0241)

作者简介

吴木贵,男,高级工程师,从事天气预报工作. wmg-0828@163.com

## 0 引言

强降水具有局地性、突发性和活动规律多变的特点,其形成机制迄今尚未揭示清楚,其预报难度更大.目前以动力数值预报与统计预报相结合的 MOS 方法在强降水预报中起日益重要的作用,该方法的重要特点是能自动考虑数值预报的系统误差及局地气候学,其主要难点在于:一是强降水属小概率事件,诸如回归及传统 BP 神经网络等统计方法用于小概率事件的预报其效果不佳;二是强降水(或暴雨)的发生虽与大尺度背景有关,但往往是次天气尺度和中尺度系统直接造成的,而目前的数值预报产品,尚不能反映次天气尺度以下的天气系统.克服这些难点的常见办法<sup>[1]</sup>有:在建模前进行消空处理,进行天气分型,引入组合因子或比较因子(指时间或空间上的变化量)在预报对象的处理上选用若干个最大值的平均等.此外,对变量进行正态化<sup>[2]</sup>也是一种常见的预处理技术.

BP 神经网络作为一种非线性建模方法,在暴雨、冰雹等小概率事件的分类预报中已很常见<sup>[3-5]</sup>,该法核心在于提高其泛化能力,而泛化性能的好坏和建模过程中的许多环节都有关.在数据预处理上,文献[6]认为通常要把训练样本随机化,这尤其适用于抽样有中断(缺测)的情形,金龙等<sup>[7]</sup>曾提出用主成份分析技术(PCA)可压缩输入规模,排除因子之间相关性,可显著提高泛化性能;在网络结构选择上,在输入输出单元数确定情况下,相同精度条件下隐单元数应尽量少,这相当于使样本点的偏差在允许范围内用最平滑的函数(或参数最少的模型)去逼近未知的非线性映射;在网络训练过程中,用交叉验证办法确定训练结束时机是防止过度拟合训练数据的一个最成功方法.

传统 BP 神经网络的目标函数(误差函数)通常是用误差平方和,这种以误差平方和最小化为准则的 BP 神经网络(ANN-MSE)是在目标输出满足正态分布条件下导出的,因此它和回归方法有同样的局限性.事实上, BP 神经网络的目标函数还可用交叉熵<sup>[8-9]</sup>,其在推导中不含有正态分布假定,因此其适用于小概率事件的预报.有人曾把交叉熵神经网络用于 DNA 结合位点及石油需求的预测<sup>[10-11]</sup>,但目前尚无人将此法应用于天气预报中.本文将这种以交叉熵为误差函数的

1 福建省建阳气象雷达站,建阳 354200

2 福建省南平市气象局,南平 353000

神经网络方法(Artificial Neural Network using Cross Entropy, ANN-CE) 引入到二分类事件的预报中, 对照 ANN-MSE 介绍了该方法的特点及主要思路, 然后利用 2003—2008 年的 ECMWF 预报场资料, 把该法用于福建省南平市 4—6 月部分大雨或以上降水的 96 h 预报中, 分别用原始因子和 PCA 降维后的主因子建立了 ANN-CE 预报模型和 ANN-MSE 预报模型, 用这些模型编制了预报系统, 系统对 2009—2010 年进行了试报, 通过比较这些模型的预报结果展示了 ANN-CE 方法的有效性及其优点。

## 1 ANN-CE 的误差函数及训练算法

### 1.1 ANN-CE 结构及性质

ANN-CE 网络结构取三层, 其中输出层节点只取一个(图 1), 单元激活函数用 Sigmoid 函数. 文献 [9] 从贝叶斯观点证明了该网络的输出是输入的概率函数. 文献 [8] 中还证得交叉熵误差函数是网络输出的相对误差的函数, 交叉熵最小化的结果也就是使网络输出的相对误差之和达最小, 而误差平方和的目的是使绝对误差之和达最小. ANN-CE 将使网络输出值较小的样本(属于小概率的那些样本)仍有较大的相对误差, 即在小概率事件的预报上, 它比 ANN-MSE 更合理。

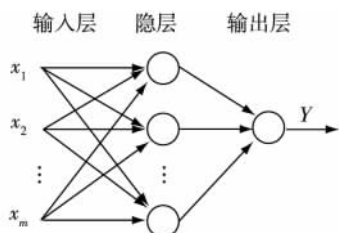


图 1 ANN-CE 网络结构示意图

Fig. 1 Network structure of ANN-CE

### 1.2 ANN-CE 的误差函数

神经网络参数(各连接权)是通过向训练样本(输入输出对)学习确定的, 训练过程中通过不断调整权试图使实际输出和网络输出误差(称误差函数)达最小化, 训练结束后网络权也就确定了, 此时就在输入和输出之间隐含建立了非线性映射关系. 对于只有一个输出节点的 ANN-MSE, 其误差函数<sup>[8]</sup>

为  $E = \frac{1}{2} \sum_n (y_n - t_n)^2$ . 其中  $y_n$  为第  $n$  个样本的网络输出,  $t_n$  为相应的目标值(实际值). 该误差函数是在实际输出满足正态分布假定下导出的, 因此 ANN-MSE 建模方法和回归方法有同样的局限性. 事实上,

在分类预报问题中, 更合理的误差函数可用交叉熵, 对两分类问题<sup>[8]</sup>为

$$E = - \sum_n \{ t_n \ln y_n + (1 - t_n) \ln(1 - y_n) \}. \quad (1)$$

式(1)就是 ANN-CE 的误差函数, 其在推导中不含有正态分布假定。

### 1.3 交叉熵神经网络训练算法

与传统 BP 网络一样, ANN-CE 的训练算法可用逐样本修订权值的反向传播算法, 该算法对每个样本按以下 3 个阶段进行处理。

1) 施加输入信号( $x_n$ ), 向前传播到输出端, 产生网络输出( $y_n$ ).

2) 用网络输出( $y_n$ )与目标输出( $t_n$ )计算出误差信号, 然后把它由输出端节点开始逐层向后(左)传播到每个节点。

3) 调整网络权. 当所有样本输入完后, 用式(1)计算出总的误差函数值, 当其足够小时停止, 否则进行下一轮的训练。

设输出层节点误差信号为  $\delta_n$ , 对于传统基于 MSSE 的 BP 神经网络<sup>[8]</sup>有  $\delta_n = y_n(1 - y_n)(t_n - y_n)$ , 而对于 ANN-CE, 文献 [8] 推导出  $\delta_n$  计算式为

$$\delta_n = y_n - t_n. \quad (2)$$

可见, ANN-CE 和 ANN-MSE 在数学形式上只有两处不同: 误差函数形式及输出层节点误差计算公式, 因此, ANN-CE 在实现时只要在传统 BP 神经网络程序基础上稍作改动即可。

## 2 资料与方法

因子资料用福建省气象台接收保存的 ECMWF (Micaps 格式) 2003—2008 年 4—6 月逐日 12 时 (GMT) 发布的 72~96 h 预报场(格距  $2.5^\circ \times 2.5^\circ$ ), 范围  $110 \sim 122.5^\circ\text{E}$   $22.5 \sim 32.5^\circ\text{N}$ (图 2) 相应雨量资料为 4 月 5 日—7 月 4 日期间的南平市(共 10 个站)逐日 20—20 时(BT)雨量, 当某日前 4 个最大雨量的平均值  $\geq 25.0$  mm 时, 称为有“部分大雨或以上降水”, 以此作为预报对象(记为  $y = 1$ , 否则  $y = 0$ ). 可见预报对象包含局部性大降水天气, 还可看出所建立的预报模型属于 MOS 方法的一种. 因子资料经预处理排除缺测值或异常值样本(共 56 个)后实际有效样本为 490 个, 其中预报对象出现 92 d, 几率为 0.19。

预报方法确定情况下, 模型好坏关键在于因子, 因子挑选大体有两个途径: 一个是从天气形势入手, 该法可能导致对某些类型预报好, 而对另一些却失

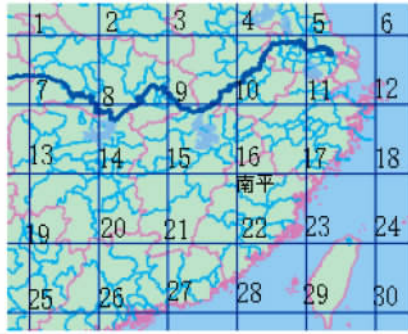


图2 预报模式计算范围及格点编号

Fig. 2 Computation range and grid number for forecast model

败,导致模型不稳定,在模型预报季节长时更为明显;另一途径是从降水形成基本条件(热力、动力、水汽等)及大降水形成必要条件(如水汽供应)入手.本文从后一种入手构造初选因子,然后用逐步回归方法挑选出复相关系数足够高的作为预报因子.这样建立的非线性模型中已含有良好的线性关系,可能会有较长的生命力.当把 ANN-MSE 或线性回归方法用于预报小概率事件时,建模前消空处理是常用且有明确效果的措施,但 ANN-CE 方法并不含正态分布假定,故未采用消空处理.

### 3 强降水预测模型及其效果检验

#### 3.1 因子初选

用逐步回归进行因子初选,按以下步骤进行.

1) 计算南平市前 4 个最大雨量的平均值,并将其进行正态化(4 次开方运算)得到序列  $y_k$ .

2) EC 资料有地面气压  $p$ (850、700、500 和 200 hPa),各层  $u$   $p$  850 hPa 温度  $t$  500 hPa 高度  $h$  850、700 hPa 相对湿度  $r$  共 13 个量,计算每个量 72~96 h 预报的平均值及变化值;然后进行因子组合得到 53 个候选因子,这些因子和  $y_k$  的相关系数皆通过显著性水平  $\alpha = 0.05$  时的  $t$  检验,再用逐步回归( $F_\alpha = 2.0$ )筛选出 19 个预报因子(表 1),表 1 中因子按其入选方程时的方差大小来排列,复相关为 0.69.

表 1 中  $x_1$  用 4 个格点(22、17、10、15)(图 2)的平均风速计算,由于格距是固定的,其计算结果和第 16 格点的涡度成正比,类似的因子还有  $x_4$  和  $x_8$ ,这 3 个因子和  $y_k$  的相关系数分别为 0.4、0.31、-0.3,可见低层为气旋式涡度高层为反气旋式涡度是闽北较大范围降水过程的重要特征; $x_{30}$  和  $x_{51}$  和西南气流有关,代表水汽输送条件; $x_{32}$  代表相对湿度条件; $x_{10}$  代表 850 hPa 南北风向东西向锋区运动的强弱,其

值越大反映南风迎锋面(或等熵面)上的爬升运动越显著也就越有利于降水,该因子是由两个量相乘得到的,其构造思想上运用了数学上乘法原理,其要点在于只要有一个量为零(如南北风为零)则该因子值就为零,就不起作用,由此避免了因子有时会起反作用的情形,类似的因子还有  $x_{11}$ 、 $x_{13}$ 、 $x_{17}$ 、 $x_{22}$ ,注意  $x_{13}$  和  $x_{22}$  的计算公式中的温度用 850 hPa 温度代替; $x_{38}$  用格点 16 和 17 地面气压的 24 h 变化值计算出,它和地面冷高压自西向东影响闽北过程有关; $x_{33}$  反映 850 hPa 降温过程; $x_{40}$  反映地面气压越低越有利于降水; $x_{65}$  的物理意义可以这样来理解:当地面有冷空气自东北路影响闽北过程中,北面的 10 格点会加压,而南面的 22 格点则减压,由此会产生由东指向西的变压风(作用力),此时西面 15 格点处的温度往往比东面 17 格点处的温度更高(用 850 hPa( $\bar{t}_{15} - \bar{t}_{17}$ )代表),则二者相乘就反映出冷空气从东北路影响闽北过程中的锋面作用效应,类似因子还有  $x_{64}$ ,但  $x_{64}$  则反映出冷空气从西北路影响闽北过程中的锋面作用效应; $x_{25}$ 、 $x_{26}$ 、 $x_{23}$  的构造思路和  $x_{10}$  类似,但它们反映的是增湿过程的作用.

表 1 原因子列表

Table 1 Original factors list

因子	层次	计算公式	方差	相关系数
$x_1$	850 hPa	$(\bar{u}_{22} + \bar{v}_{17}) - (\bar{u}_{10} + \bar{v}_{15})$	0.058 8	0.40
$x_{30}$	700 hPa	$(\bar{u}_{16} + \bar{u}_{17} + \bar{u}_{21} + \bar{u}_{22}) / 4$	0.034 7	0.47
$x_{32}$	850 hPa	$(\bar{r}_{15} + \bar{r}_{16}) / 2$	0.034 1	0.61
$x_{51}$	500 hPa	$(\bar{v}_{16} + \bar{v}_{17} + \bar{v}_{22}) / 3$	0.032 9	0.33
$x_{10}$	850 hPa	$\bar{v}_{16} \times (\bar{t}_{16} - \bar{t}_{10})$	0.026 1	0.20
$x_{38}$	地面	$\Delta p_{16} - \Delta p_{17}$	0.023 6	0.20
$x_{33}$	850 hPa	$(\Delta t_{10} + \Delta t_{15} + \Delta t_{16}) / 3$	0.021 9	-0.31
$x_4$	700 hPa	$(\bar{u}_{22} + \bar{v}_{17}) - (\bar{u}_{10} + \bar{v}_{15})$	0.019 1	0.31
$x_8$	200 hPa	$(\bar{u}_{22} + \bar{v}_{17}) - (\bar{u}_{10} + \bar{v}_{15})$	0.018 6	-0.30
$x_{17}$	850 hPa	$\Delta v_{16} \times (\Delta t_{16} - \Delta t_{10})$	0.018 4	0.14
$x_{11}$	850 hPa	$\bar{v}_{16} \times (\bar{t}_{16} - \bar{t}_{22})$	0.015 7	-0.21
$x_{13}$	700 hPa	$\bar{v}_{16} \times (\bar{t}_{16} - \bar{t}_{10})$	0.015 1	0.25
$x_{40}$	地面	$(\bar{p}_{15} + \bar{p}_{16}) / 2$	0.013 9	-0.21
$x_{22}$	700 hPa	$\Delta v_{16} \times (\bar{t}_{16} - \bar{t}_{22})$	0.012 1	0.10
$x_{65}$	地面	$(\Delta p_{22} - \Delta p_{10}) \times (\bar{t}_{15} - \bar{t}_{17})$	0.011 1	-0.11
$x_{64}$	地面	$(\Delta p_{15} - \Delta p_{17}) \times (\bar{t}_{10} - \bar{t}_{22})$	0.009 1	-0.18
$x_{25}$	700 hPa	$\bar{v}_{16} \times (\bar{r}_{16} - \bar{r}_{22})$	0.008 3	0.19
$x_{26}$	700 hPa	$(\bar{u}_{16} + \bar{u}_{17} + \bar{u}_{21} + \bar{u}_{22}) / 4$	0.008 3	0.13
$x_{23}$	850 hPa	$\bar{v}_{16} \times (\bar{r}_{16} - \bar{r}_{22})$	0.005 1	0.11

#### 3.2 因子降维

原因子有 19 个,对其进行 PCA 变换后以其与  $y_k$  相关不为 0.0 为原则,挑出 13 个主成分对应的主

因子作为神经网络预报模型的输入因子,这 13 个主成分累积方差贡献率已达 0.77.

### 3.3 预报模型的建立

把 13 个主因子数据归一化到  $[0.1, 0.9]$ ,把预报对象进行 01 化 ( $y=1$  表示有,  $y=0$  表示无),把样本次序随机化,再把样本分成训练集(300 个)和验证集(190 个)2 部分,采用具有动量更新的 BP 算法,学习因子和动量因子取 0.5,初始权和阈值设为  $-0.5/\text{fan\_in}$  到  $0.5/\text{fan\_in}$  区间均匀分布的随机数 ( $\text{fan\_in}$  是输入数),网络结构为 13-1-1 型,用交叉验证办法训练网络,即用训练集的交叉熵误差来调整网络权,在训练完一轮后分别计算训练集和验证集的总的误差函数值  $E$ ,如此得到图 3,取网络训练 2 000 次时的网络权值作为预报模型参数,还以 TS 评分 (TS 评分 = 报对次数 / (报对次数 + 空报次数 + 漏报次数)) 最高为原则从历史样本中确定预报临界值为 0.19,这个值恰好等于预报对象在历史样本中出现几率.作为对照,用同样方法建立了相同结构的主因子 ANN-MSE 预报模型,其误差变化曲线如图 4,取训练 1 000 次时的网络权作为预报模型参数.

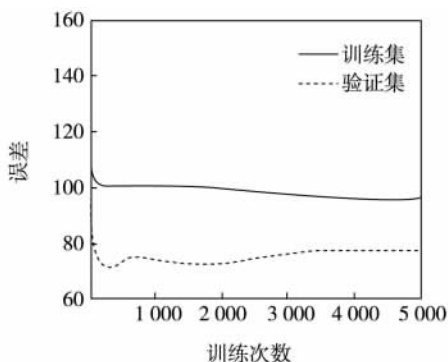


图 3 主成分 ANN-CE 误差随训练次数变化曲线

Fig.3 Error of principal component ANN-CE model varying with training times

同样出于比较目的,还建立了原始因子(19 个) ANN-CE 预报模型,结构为 19-3-1 型,其误差变化曲线如图 5,取训练 9 次时的网络权作为预报模型参数,此时验证集的交叉熵误差函数值达最小.还建立了相同结构的原始因子 ANN-MSE 预报模型,取训练 190 次时的网络权作为预报模型参数,此时验证集的误差函数值亦达最小.从图 3—5 可看出:

1) 图 3 和 4 中误差函数值的变化随着训练次数增加趋于稳定,与此明显不同的是,图 5 中随着训练次数增加,在训练集误差减小的同时验证集误差

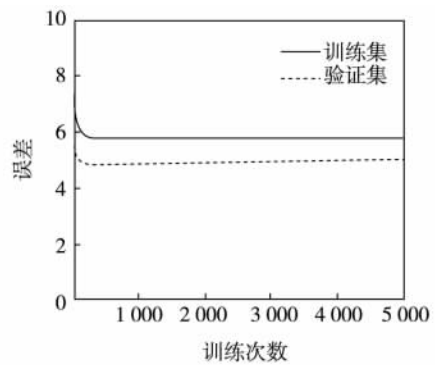


图 4 主成分 ANN-MSE 误差随训练次数变化曲线

Fig.4 Error of principal component ANN-MSE model varying with training times

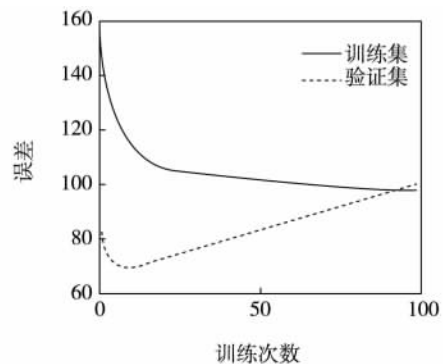


图 5 原始因子 ANN-CE 误差随训练次数变化曲线

Fig.5 Error of original factor ANN-CE model varying with training times

却开始增大,这是训练过度的表现;

2) 图 3 误差函数值最大为 153,而图 4 误差函数值最大为 7.4,二者相差很大,这是因为交叉熵误差函数是网络输出的相对误差的函数,而误差平方和与绝对误差有关,相对误差的量级自然比绝对误差大得多;

3) 图 3 中的误差函数值下降幅度明显比图 4 大,这似乎反映 ANN-CE 网络对预报对象随输入变量的分布密度函数的逼近更合理.

### 3.4 预报系统的建立

把前述 4 个模型用 C++ 语言编制了预报系统,系统运行流程如下:从 Micaps 文件中读出格点资料后计算出 19 个原因子值,期间要排除缺掉或异常值的情况,再对其进行标准化,然后用 19 个特征向量和特征值把标准化后的原因子变换为 13 个主因子,再把主因子值变换到  $[0.1, 0.9]$ ,最后用模型参数及模型输入值(因子值)计算出网络输出值(预报值).

预报值在每天 08 时以后就已自动产生,预报员可在一个窗口程序中查看预报结论,主导结论由主因子 ANN-CE 模型给出,其他模型结论供参考.系统在 2009 年 12 月 1 日开始在南平市气象台运行,此后运行状况良好.

### 3.5 预报模型检验

前述 4 个模型的预报检验结果如表 2—4,表 2 中有“+”号的式子是指训练样本和检验样本之和.模型以历史样本预报对象出现几率(0.19)为预报临界值(当网络输出 $\geq 0.19$ 时报有,否则报无).其中 2009 年有效样本有 88 d,预报对象出现 15 次(几率 17%),属正常年份,而 2010 年有效样本有 90 d,预报对象出现 31 次(几率 33%),属多雨年份,该年闽北出现了百年少见持续性暴雨过程及特大洪水,其强度超过 1998 年.综合表 2—4 的结果可以看出:

表 2 模型对训练集和验证集的评分  
Table 2 TS grades of forecast models on training set and validation set

预报模型	报对次数	漏报次数	空报次数	TS 评分
主成分 ANN-CE	51 + 33 = 84	4 + 4 = 8	47 + 66 = 113	0.41
主成分 ANN-MSE	50 + 35 = 85	5 + 2 = 7	68 + 48 = 116	0.41
原因子 ANN-CE	50 + 28 = 78	9 + 5 = 14	84 + 58 = 142	0.33
原因子 ANN-MSE	54 + 26 = 80	5 + 7 = 12	75 + 54 = 129	0.36

表 3 模型对 2009 年样本的评分  
Table 3 TS grades of forecast models on 2009 samples

预报模型	报对次数	漏报次数	空报次数	TS 评分
主成分 ANN-CE	11	4	13	0.39
主成分 ANN-MSE	9	6	13	0.32
原因子 ANN-CE	7	8	7	0.32
原因子 ANN-MSE	6	9	7	0.27

表 4 模型对 2010 年样本的评分  
Table 4 TS grades of forecast models on 2010 samples

预报模型	报对次数	漏报次数	空报次数	TS 评分
主成分 ANN-CE	27	4	16	0.57
主成分 ANN-MSE	29	2	21	0.55
原因子 ANN-CE	28	3	28	0.47
原因子 ANN-MSE	27	4	27	0.47

1) 主因子模式比原因子模式的 TS 评分明显要高,且漏报次数也要少很多,表明 PCA 降维技术对改进模式性能是显著的;

2) 就 ANN-CE 模型和 ANN-MSE 模型对比来说,对于历史样本和 2010 年二者差别不大,但在

2009 年,无论是原始因子还是用主因子前者的 TS 评分都更高一些并且漏报次数要少,这可能与该年预报对象出现次数较少且较分散有关,这似乎说明,在分类预报中,如果预报对象出现概率很小,选用 ANN-CE 模型比用 ANN-MSE 模型更好;

3) ANN-CE 模型对 2010 年 6 月的持续性暴雨过程表现出很好的预报性能,但其中 15、16、27、28 日出现空报,事实上这几天南平市分别有 8、12、12、10 个乡镇自动站出现大雨以上降水,各日最大雨量分别为 56.8、74.5、45.1 和 65.4 mm,可见实际仍是正确的,这种情形可能不在少数;

4) 主因子 ANN-CE 预报模型性能最为稳定,其对独立样本的 TS 评分和漏报率分别为 0.51 和 0.17,也比当地气象台的同期值(分别为 0.34 和 0.57)要好;

5) 预报模型的网络输出最大值为 0.49,未达到或接近 1.0.这是因为预报对象先验概率不大(0.19),根据贝叶斯法则,作为后验概率的输出值随之降低,另外,实际使用中要注意,输出值越大,代表当前输入状态下,预报对象出现机率也越大.

## 4 小结

1) 当用于分类预报时,ANN-CE 方法因其不含正态分布的假定,比 ANN-MSE 方法更合理,且可不必进行消空处理;交叉熵最小化的结果也就是使网络输出的相对误差之和达最小.这些特点使其适合于小概率事件的预报.

2) 用 PCA 降维后的主因子预报模型 TS 评分比原始因子预报模型高且漏报次数少;用交叉验证办法确定训练结束时机有效防止了过度拟合训练数据的问题,使模型更为稳定;对于预报对象出现较少的 2009 年,ANN-CE 模型比 ANN-MSE 模型预报性能更好.总体来看,主因子 ANN-CE 预报模型最为稳定,是一种适合于小概率事件预报的方法.

## 参考文献

### References

- [1] 陆汉城. 中尺度天气原理和预报[M]. 北京: 气象出版社 2002: 280-283  
LU Hancheng. Principle and forecast of medium range weather[M]. Beijing: China Meteorological Press 2002: 280-283
- [2] 施能. 气象科研与预报中的多元分析方法[M]. 2 版. 北京: 气象出版社 2002: 1-5, 72-82  
SHI Neng. Multivariate analysis for scientific research

- and forecast of meteorology [M]. 2nd Ed. Beijing: China Meteorological Press 2002: 1-5, 72-82
- [ 3 ] 林健玲, 金龙, 林开平. 神经网络方法在广西日降水预报中的应用 [J]. 南京气象学院学报, 2006, 29(2): 215-219  
LIN Jianling, JIN Long, LIN Kaiping. Application of artificial neural network to daily regional average rainfall forecast in Guangxi [J]. Journal of Nanjing Institute of Meteorology 2006 29(2): 215-219
- [ 4 ] 苗春生, 段婧, 徐春芳. 人工神经网络方法在短期天气预报中的应用 [J]. 江南大学学报: 自然科学版, 2007, 6(6): 648-653  
MIAO Chunsheng, DUAN Jing, XU Chunfang. Application of artificial neural network to forecast short-term precipitation in Nanjing [J]. Journal of Jiangnan University: Natural Science Edition 2007 6(6): 648-653
- [ 5 ] 曹晓钟, 闵晶晶, 刘还珠, 等. 分类与集成方法在降雨预报中的应用 [J]. 气象, 2008, 34(10): 3-10  
CAO Xiaozhong, MIN Jingjing, LIU Huanzhu, et al. Application of classification and integration to rainfall forecast [J]. Meteorological Monthly 2008 34(10): 3-10
- [ 6 ] Ham F M, Kostantac I. 叶世伟, 王海娟译. 神经计算原理 [M]. 北京: 机械工业出版社, 2007: 77-85  
Ham F M, Kostantac I. Principles of neurocomputing for science & engineering [M]. Columbus, OH: McGraw-Hill 2000
- [ 7 ] 金龙. 神经网络气象预报建模理论与应用 [M]. 北京: 气象出版社, 2004: 189-201  
JIN Long. Neural network modeling method & application for meteorological forecast [M]. Beijing: China Meteorological Press 2004: 189-201
- [ 8 ] 阎平凡, 张长水. 人工神经网络与模拟进化计算 [M]. 2版. 北京: 清华大学出版社, 2005: 437-447  
YAN Pingfan, ZHANG Changshui. Artificial neural networks and evolutionary computing [M]. 2nd Ed. Beijing: Tsinghua University Press 2005: 437-447
- [ 9 ] Mitchell T M. 曾华军, 张银奎译. 机器学习 [M]. 北京: 机械工业出版社, 2008: 79-92, 121-125  
Mitchell T M. Machine learning [M]. Columbus, OH: McGraw-Hill, 1997
- [ 10 ] 徐东, 王翼飞. ANN-CE: 一种预测 DNA 结合位点的改进神经网络方法 [J]. 应用科学学报, 2005, 23(2):  
XU Dong, WANG Yifei. ANN-CE: An improved neural network method for predicting DNA binding sites [J]. Journal Applied Sciences 2005 23(2):
- [ 11 ] Nasr G E, Badr E A, Goun C. Cross entropy function in neural networks: Forecasting gasoline demand [EB/OL]. [2010-08-28]. FLAIRS-02 Proceedings. <http://www.aaai.org/Papers/FLAIRS/2002/FLAIRS02-075.pdf>

## Application of BP neural network using cross-entropy to 96 hours forecast of heavy precipitation event in northern Fujian province

WU Mugui<sup>1</sup> JIANG Caiying<sup>2</sup> ZHANG Xinhua<sup>2</sup> LAI Rongqin<sup>1</sup>

1 Jianyang Meteorological Radar Station of Fujian Province, Jianyang 354200

2 Nanping Meteorological Office of Fujian Province, Nanping 353000

**Abstract** As a neural network based on MSSE, ANN-MSE is not an appropriate solution to the problem of predicting rare weather event. In this paper, an improved neural network method, ANN-CE is presented, which is a three layered back-propagation neural network with one output unit. The error function of ANN-CE is a cross entropy function. Then utilizing ECMWF forecast fields data, this method was applied to 96 hours forecast of heavy precipitation event in northern Fujian province. The ANN-CE model and the ANN-MSE model based on original factor and principle component after PCA reducing dimensions were respectively built. These models were applied to independent samples in 2009—2010, and the test results are as following: TS grade for model based on principal component is higher than that of model based on original factors; miss-rate for the ANN-CE model is lower than that of the ANN-MSE model. All in all, ANN-CE model based on principal component has best performance and stability, whose TS grade and miss-rate was respectively 0.51 and 0.17, so it was suited for forecasting rare event.

**Key words** BP neural network; cross entropy; categorical forecast; rare event