

农孟松, 黄海洪, 孙崇智, 等. 基于主分量神经网络的降水集成预报方法研究[J]. 气象, 2011, 37(3): 352-355.

基于主分量神经网络的降水集成预报方法研究^{* 1}

农孟松^{1,2} 黄海洪² 孙崇智² 郑凤琴² 陈伟斌²

1 南京信息工程大学大气科学学院, 南京 210044

2 广西气象台, 南宁 530022

提 要: 运用人工神经网络与主分量分析(PCA)相结合的方法, 对同一降水预报量的各种数值预报产品进行集成预报研究。结果表明: 主分量人工神经网络方法所构造的集成预报模型, 不仅对历史样本的拟合精度好于各个子预报产品, 独立样本的实验预报结果也显示出更好的预报准确率及稳定性。业务应用前景良好。

关键词: 主分量, 神经网络, 集成预报

A Neural Network Model Based on Principal Component Analysis for Ensemble Precipitation Prediction

NONG Mengsong^{1,2} HUANG Haihong² SUN Chongzhi² ZHENG Fengqin² CHEN Weibin²

1 School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044

2 Meteorological Observatory of Guangxi, Nanning 530022

Abstract: Using the method of artificial neural network and principal component analysis (PCA) to study a variety of numerical forecast products for the same precipitation forecast. The results show that the fitting accuracy of the principal component analysis artificial neural network ensemble model is better than each sub-product, and the experimental results of the independent sample also show its better prediction accuracy and stability. The model is a good prospect for operational applications.

Key words: principal component analysis, neural network, ensemble prediction

引 言

随着数值预报业务水平的不断提高, 各类数值预报产品对预报员最终的综合预报影响越来越大。目前在省级以下台站的集成预报还是以经验为主, 兼有部分定量统计集成方法^[1-2]。较为常用的定量统计集成方法有回归、平均和加权、典型相关等预报方法。这些常用的集成预报方法本质上是属于线性的顺序处理技术。为了提高模式输出的降水预报的精度, 使模式结果能满足不断提高的应用要求, 用神经网络方法替代传统的回归方法对结果进行后处理, 可以得到满意的结果^[3], 美国国家天气局已经有

几个人工智能系统投入使用^[4], 如 Gillispie 研制了一系列人工神经网络系统, 其中定量降水预报网络(Quantitative Precipitation Forecast, QPF)已经投入业务使用。国内许多学者也成功地运用了神经网络进行降水、沙尘暴、大风等的预报^[5-10]。曹晓钟等^[5]采用自组织神经网络进行天气分型, 集成多种数值预报产品, 分型建模进行分级降水试报, 其结果优于单一模式。赵翠光^[6]使用人工神经网络方法建立了我国沙尘暴短期预报模型, 输出我国有无沙尘暴的预报结果。

反向传播人工神经网络, 即 BP(back-propagation)模型^[11], 具有物理概念清晰、通用性好、拓扑结构灵活可变、非线性映射能力强的特点。理论已经

* 中国气象局新技术推广项目预报员专项(CMATG2008Y07)和广西科技厅攻关项目(桂科攻 0993002-1 和 0816006-9)共同资助
2010 年 1 月 7 日收稿; 2010 年 4 月 7 日收修定稿
第一作者: 农孟松, 主要从事天气预报和中尺度灾害性天气研究. Email: nmsong1997@163.com

证明,一个3层的神经网络能够实现任意精度的连续函数映射^[12]。本文在主分量分析基础上构造BP神经网络低维学习矩阵的预报建模方法,浓缩预报信息,降维去噪,并针对神经网络方法在预报建模中存在的“过拟合”现象及提高泛化性能问题展开讨论,以达到提高预报准确率的目的。

1 基本原理和方法

1.1 主分量分析方法(PCA)

主分量分析方法(PCA)是利用降维的思想,把原来多指标转化为少数几个相互独立的综合指标的一种多元统计分析方法,该方法被广泛应用于大气科学的预报和研究工作中^[13-14]。其基本思想是:设某一预报对象有 n 个预报因子 $X = (x_1, x_2, \dots, x_n)^T$,可以通过主分量分析方法构造 n 个新综合因子变量 $Z = (z_1, z_2, \dots, z_n)^T$ 。每一个新综合因子变量是原因子变量的线性组合,新综合因子变量之间是相互正交的,即新因子之间的相关系数为零,新综合因子变量的特征值明确表示了它对原因子组的贡献大小。

将主分量分析得到的新综合因子变量用于构造神经网络学习矩阵,可以很直观地去掉特征值为零的对应的主分量,因为这些主分量几乎没有包含原变量的信息,只保留特征量相对较大的且与预报量相关程度高的主分量。另外在实际计算中可以发现,一般前几个主分量具有更大的方差贡献且与预报量相关性较好,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音影响。由上述分析不难发现,利用主分量分析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,对原预报因子矩阵有很好的降维作用。

1.2 神经网络方法

人工神经网络是一个可自动提取一组预报变量和另一组自变量之间非线性关系的数据处理系统,该网络的建立过程称为人工神经网络的训练过程,是用自适应算法递归迭代求解因变量与自变量之间的非线性关系。训练后的神经网络模型可用来估算或预报预测变量。到目前为止,已经出现许多神经网络模型及相应的学习方法。其中反向传播神经网络模型(简称BP模型)是应用较广泛的一种,这种模型在输入层和输出层之间插入若干个隐含层,相邻层次之间的神经网络元之间用连接权系数作相互连接,而各层内的神经元之间没有连接,本文使用的神经网络模型只含1个隐含层, $f(x)$ 为网络激活函数,采用S型函数:

$$f(x) = 1/(1 + e^{-x})$$

具体的算法可参阅文献^[15]。

2 集成预报建模

2.1 基本资料的处理

本文选取的资料时段为2002年4—9月及2003年4—8月,资料范围是广西11个主要城市,个例标准为有超过3站日降雨量大于25mm的降水过程。经统计共有150个个例,其中120个样本用于预报建模,30个样本用于方程检验。本文选取的子预报产品为国家气象中心T213、T639、广州区域中心中尺度预报模式和日本数值预报的降水预报产品。上述四种产品都是格点值,为了与预报值形式相一致,应用高斯权重插值方法,把格点值插值到相应的11个城市点上。另外本文采用的评分方法为2005年中国气象局确定的《中短期预报质量检验办法(试行)》。

2.2 主分量学习矩阵模型的建立

根据前述的主分量分析方法,对上述各因子作主分量分析。以主分量特征值相对较大,与预报量相关系数较高作为取舍标准,确定神经网络学习矩阵的输入。表1是南宁市降水量各主分量特征值及主分量与降水预报量的相关系数,可以看出,相对来说,前三个主分量与预报量的相关系数较高,并且相对应的特征值也较大,第四个主分量的相关系数相对于第三个主分量小将近2个量级,因此用前三个主分量来构造神经网络的学习矩阵。其他站点也有类似的结果。本文采用三层前馈网络计算方法,隐节点数为3,学习因子和动力因子分别取0.9和0.7。利用1.2节给出的公式,对标准化的学习矩阵作反复的学习训练,当学习矩阵训练达5000次时,误差函数趋于稳定。由网络模型所确定的连接权系数及阈值可以得到相应的预报模型,并以此确定的参数和网络结构作为未来独立样本和实际预报的模型,在本文中也用于对历史样本的预报拟合对比分析。

表 1 南宁市降水量各主分量特征值及
主分量与降水预报量的相关系数

Table 1 Eigenvalues of principal component and
correlation coefficients between principal
component and predictor for Nanning

序号	1	2	3	4
特征值	1.98	0.91	0.61	0.50
相关系数	0.54	-0.15	-0.25	-0.008

表 2 降水预报历史拟合评分

Table 2 Scores of precipitation forecast historical record fitting

	南宁	桂林	柳州	梧州	北海	防港	百色	河池	钦州	玉林	贵港	平均
T213	0.29	0.30	0.20	0.28	0.14	0.30	0.34	0.24	0.32	0.34	0.34	0.28
T639	0.22	0.18	0.18	0.22	0.17	0.08	0.19	0.18	0.12	0.13	0.17	0.17
JP	0.46	0.37	0.38	0.37	0.20	0.39	0.37	0.29	0.35	0.39	0.40	0.36
GZ	0.17	0.23	0.36	0.28	0.21	0.18	0.20	0.20	0.16	0.21	0.20	0.22
集成	0.53	0.44	0.43	0.44	0.42	0.43	0.51	0.33	0.34	0.41	0.43	0.43

对 120 个样本的降水预报历史拟合,集成预报方法是其中最优化的一种。在 11 个城市的拟合中,有 10 个站点评分为最高,总平均分也明显好于其他 4 种产品的平均值。这说明运用主分量人工神经网络方法所构造的集成预报模型,对各类子预报产品有较好的集成优化效果。

3.2 对独立样本的预报对比分析

在一般情况下,预报模型对历史样本的拟合精度很高并不一定说明其预报效果也一定很好。统计预报方法在实际预报建模的应用中,最大的障碍就是存在明显的“过拟合”现象,因此预报模型是否存在“过拟合”现象,是检验其应用价值的重要标准。

表 3 为主分量人工神经网络集成预报模型与其他 4 个子预报产品对 30 个独立样本的预报评分结果,由表 3 可以看出 3 个特点,一是集成预报的总平均评分是最高的,在 11 个城市的预报中有 6 个名列第一,在其余 5 个站点的预报评分也较高,说明该集成方法具有较好的集成优化效果;二是相对而言,桂

3 计算结果分析

3.1 对历史样本的拟合对比分析

集成预报的根本目的是要能够客观地综合所有子预报方程对预报量的预报信息,而考察集成预报方法的有效性,首先应该比较分析集成预报模型历史样本拟合率是否好于各子预报方程。由表 2 可知,

南的集成预报效果优于桂北,可能的原因是桂北地区降水主要受西风带系统影响,降水机理规律性较强,有线性变化特征,而桂南地区降水机理较复杂,非线性变化特征较明显,因此采用具有非线性准动力系统特征的人工神经网络方法进行预报集成效果更显著一些;三是集成预报方法没有明显的过拟合现象,无论是对各城市点或总的平均,对历史的拟合预报与对独立个例的预报差异不大。

上述分析表明,主分量神经网络集成预报模型不但没有明显的“过拟合”现象,而且预报精度也明显好于各子预报产品。其主要原因是,主分量神经网络预报模型所采用的预报因子是原因子群作 EOF 展开后的 3 个主分量,它们不仅浓缩包含了全部预报因子的主要预报信息,同时这 3 个主分量是正交的,没有复共线性关系。由于根据 3 个主分量构造的神经网络预报模型,其网络结构规模较小,应用研究成果表明,在能够包含预报量与预报因子间主要相关关系的条件下,网络规模越小预报模型的预报性能越好。

表 3 对独立样本的降水预报评分

Table 3 Precipitation forecast scores of independent samples

	南宁	桂林	柳州	梧州	北海	防港	百色	河池	钦州	玉林	贵港	平均
T213	0.36	0.35	0.21	0.20	0.38	0.41	0.48	0.40	0.31	0.41	0.38	0.35
T639	0.10	0.18	0.28	0.15	0.1	0.1	0.04	0.00	0.13	0.15	0.13	0.12
JP	0.35	0.38	0.41	0.35	0.34	0.32	0.32	0.30	0.39	0.48	0.62	0.39
GZ	0.05	0.22	0.13	0.14	0.19	0.08	0.08	0.19	0.11	0.12	0.17	0.13
集成	0.41	0.33	0.48	0.29	0.38	0.53	0.64	0.34	0.39	0.46	0.52	0.44

4 小 结

(1) 主分量人工神经网络方法所构造的集成预报模型,对几种降水数值预报产品有较好的集成优化效果。无论是对历史个例的拟合预报或对独立样本的预报,集成方法皆优于各子预报产品。

(2) 桂南的降水集成预报效果优于桂北。

(3) 主分量人工神经网络方法所构造的降水集成预报模型,无明显的“过拟合”现象,这说明主分量神经网络预报模型使预报泛化能力有显著的提高。

参考文献

- [1] 苗爱梅,郭玉玺,梁明珠,等. 夏季分县降水量集成预报试验[J]. 气象,2000,26(5):8-12.
- [2] 周家斌,张福海. 一种汛期降水分布客观集成预报方法[J]. 应用气象学报,2000,11(6):93-97.
- [3] Robert J K, Ana P B. Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks[J]. Weather and Forecasting, 1998, 13(4): 1194-1204.
- [4] 曾晓梅. 国外人工智能技术在天气预报中的应用综述[J]. 气象科技, 1994(4): 4-10.
- [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):3-11.
- [6] 赵翠光. 人工神经元网络方法在沙尘暴短期预报中的应用[J]. 气象,2004,30(4):39-41.
- [7] 孙军波,钱燕珍,陈佩燕,等. 登陆台风站点大风预报的人工神经网络方法[J]. 气象,2010,36(9):81-86.
- [8] 钱莉,兰晓波,杨永龙. 最优了集神经网络在武威气温客观预报中的应用[J]. 气象,2010,36(5):102-107.
- [9] 蔡敏,黄艳,朱宵峰,等. 基于BP神经网络的地质灾害细网格预报模型[J]. 气象,2009,35(7):95-100.
- [10] 张雪慧,官莉,王振会,等. 利用人工神经网络方法反演大气温度廓线[J]. 气象,2009,35(11):137-142.
- [11] 孙照渤,谭桂容,赵振国. 人工神经网络方法在夏季降水预报中的应用[J]. 南京气象学院学报, 1998, 21(1): 47-51.
- [12] Rumelhart D E. Parallel distributed processing: Explorations in the microstructure of cognition[M]. Cambridge MA: MIT Press, 1986:101-109.
- [13] 张邦林,丑纪范,孙照渤. 用前期大气环流预报对中国夏季降水的 EOF 迭代方案[J]. 科学通报,1991,36(23):1797-1798.
- [14] 丁裕国. 经验正交函数展开气象场收敛性的研究[J]. 热带气象学报,1984(4):316-325.
- [15] 李学桥,马莉. 神经网络工程应用[M]. 重庆:重庆大学出版社,1996:37-44.

征稿简则

1 《气象》主要刊登气象科学研究领域的综合评述及研究论文;天气、气候诊断分析与预报技术;气象业务技术及业务现代化建设经验;气象灾害的规律及防灾减灾决策;公共气象服务和专业气象服务技术方法;气象科技信息动态等。

2 来稿注意事项

- 2.1 来稿务必论点明确,数据可靠,文字精炼。文章的书写顺序为:中文题目(不超过20个汉字)、作者姓名、单位名称、邮政编码、中文提要(200~300字)、关键词(3~8个);英文题目、作者姓名、单位名称、邮政编码、英文提要及英文关键词;引言;正文;结论和讨论;致谢;参考文献。文章首页页脚处附作者信息,即姓名、从事专业工作和研究方向。Email等。
- 2.2 正文标题用阿拉伯数字连续编号,不同层次数字间用圆点间隔,如“1”,“1.2”,“2.1.3”等。标题左顶格,在数字编号后空一格再写标题,末尾不加标点符号。
- 2.3 基金资助的研究项目,请注明省部级以上基金名称和项目编号。
- 2.4 文稿的单位制采用《中华人民共和国法定计量单位》。图、表中量和单位间用“/”隔开,表示物理量的符号用斜体表示,并注意文种、大小写、正斜体、上下角码等。
- 2.5 科技术语和名词应使用全国自然科学名词审定委员会公布的名词。
- 2.6 文稿只附必要的图表。图、表要求准确,清晰,美观。在文中相应的位置插入图、表。图、表附相应的中英文图题、表题。附表请使用三线表。
- 2.7 参考文献按在文内出现顺序连续编码。在期刊上发表的文献书写格式为:[序号]作者姓名(列出前3位).题名[J].期刊名,年,卷(期):页码.在专著、教科书上发表的文献书写次序为:[序号]作者姓名.译者姓名.文题名.书名[M].出版地:出版社,出版年:页码.

3 请勿一稿两投。本刊一般不退稿,若6个月内未见刊用通知,可自行处理。

4 本刊收取发表费,并付稿酬,所付稿酬包含纸质版、数字版稿酬和刊物内容网络服务报酬等。凡不同意将其稿件纳入此两种版本进行交流者,请书面说明。

5 通讯地址:北京市中关村南大街46号《气象》编辑部 邮政编码:100081

网上投稿地址: <http://219.234.83.5>

电话:(010)68407936 58993789 68407336 Email:qixiang@cams.cma.gov.cn