遗传算法优化神经网络用于大气污染预报

王 芳，程水源，李明君，范 清
(北京工业大学 环境与能源工程学院，北京 100124)

摘 要：大气污染预报可以对大气污染提出警示，保护人体健康和生活环境．为了对北京市PM_{10}的质量浓度进行预报，建立了用于大气污染预报的遗传神经网络模型，该模型运用遗传算法优化神经网络的权值和阈值，有效提高了网络的收敛性和预报准确率．用改进后的神经网络对北京市PM_{10}的质量浓度进行了模拟，并将模型模拟结果与美国第3代空气质量模型 Models-3(CMAQ)的数值模拟结果进行了比较．实验结果表明：遗传神经网络模型和数值模型的模拟结果的平均相对误差分别为0.21和0.26，用于空气污染物质量浓度短期预报时，神经网络模型的预测精度与数值模型的预测精度相当．对于没有条件开展空气污染数值预报的城市或地区，神经网络是一种有效的替代方法．

关键词：PM_{10}；神经网络；遗传算法；CMAQ
中图分类号：X 823 **文献标志码**：A **文章编号**：0254－0037(2009)09－1230－05

目前，用于空气污染预报的方法主要有潜势预报、统计预报和数值预报．数值预报方法可以对各种大气污染物在不同尺度下的不同类型污染过程进行模拟，是未来空气污染预报的发展趋势．但空气污染数值预报需要高分辨率的气象数据和污染源数据作为数据支撑，在一些监测水平较低的中小城市较难实现．

空气污染物质量浓度的时空分布受到气象场、排放源、复杂下垫面、理化生过程的耦合等多种因素的影响，具有较强的非线性特性，神经网络模型可以很好地把握该复杂过程的非线性特征，因此可以得到更好的预测效果．近年来，神经网络模型被用来对城市可吸入颗粒物(inhalable particulate matter，简称PM_{10})的质量浓度进行不同时间尺度的预报，均取得了较好的效果[1-2]．此外，经遗传算法优化的神经网络也被应用于径流预报、气象预报等方面[3-4]．实验研究表明，神经网络模型的预测效果优于许多传统的统计模型，如多元线性回归、分类回归树、自回归等[5]．本文利用遗传算法改进的神经网络模型对北京市的$\rho(PM_{10})$进行模拟预报，将其模拟结果与数值模型模拟结果进行比较，并讨论了二者的适应性．

1 遗传神经网络模型的构建

神经网络的初始权值、阈值的选择缺乏依据，具有很大的随机性，很难选取具有全局性的初始点，因而求得全局最优的可能性较小．遗传算法是使用逐次迭代法搜索寻优，是以全局并行搜索方法来进行搜索的，这种群体搜索使遗传算法得以突破邻域搜索的限制，可以实现整个解空间上的分布式信息采集和探索，能找到满足要求的最优个体，具有全局搜索能力以及简单、快速、稳定性强等特点．因此可将神经网络的训练分成2个部分：首先用遗传算法来优化网络的初始权值和阈值，然后再用算法来训练输入数据，得到优化后的网络模型．遗传神经网络实现流程如图1所示．

利用遗传算法进行网络权值、阈值优化的优化程序如下：

1) 编码和初始群体的生成

随机产生 N 个初始串结构数据，对每个结构编码．将神经网络的各个权值和阈值按次序编成一个实

收稿日期：2007-12-28．
基金项目：国家重点基础发展计划973资助项目(2005CB724201)．
作者简介：王 芳(1983－)，女，河北邢台人，博士研究生；程水源(1958－)，男，河北邯郸人，教授，博士生导师．

数数组，作为遗传算法的一个染色体．染色体的长度 $S=R*S_1+S_1*S_2+S_1+S_2$，其中，S_1 为隐层结点数量，S_2 为输出数量，R 为输入数量．遗传操作在这样的染色体群中进行．

2) 目标函数和适应度函数

选用神经网络的误差平方和，即网络误差 σ 作为遗传算法的目标函数，则遗传操作的评价函数即适应度为 $F=1/\sigma$，网络误差越小，适应度越大．

3) 选择、交叉、变异

根据适应度、选择率进行选择操作，依据选择好的交叉率，采用两点交叉对父代个体的基因部分交换重组，产生新个体．变异的目的是为了保持染色体的多样性，防止早熟现象发生，实现全局搜索．

重复以上操作，直到进化代数达到要求或网络误差满足条件时结束遗传算法，选择网络误差最小的一组权值和阈值，再利用 BP 算法进行训练．

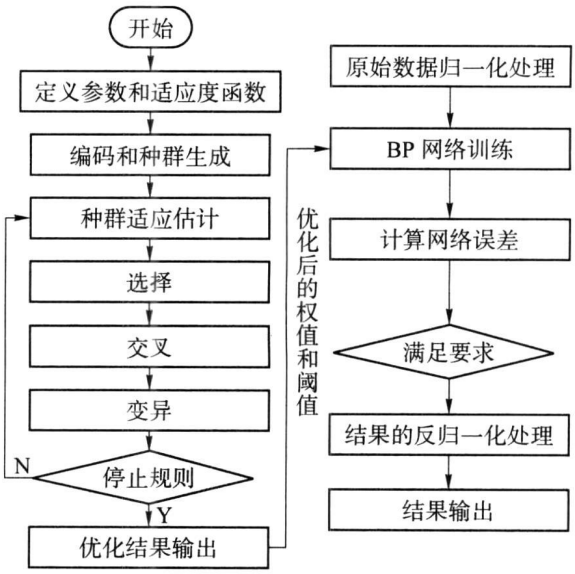

图 1　遗传算法优化的神经网络实现流程

Fig. 1　Schematic representation of the BP networks procedure optimized by GAs

2　实验过程与结果分析

2.1　数据获取

由于气象数据和污染数据在各季度内的变化具有相对一致的特点，因此 BP 网络预报模型需分季节建立，以捕捉不同季节气候污染物相应的变化特征[6]．以夏季为例，利用北京市(车公庄、古城、前门、农展馆、天坛、东四和奥体 7 个国控监测站点)2002 年夏季代表月份 7、8 月的气象数据及相应时段的 $\rho(PM_{10})$ 监测平均值共 492 组数据作为实验数据，其中 2002-07-01－2002-08-19共396 组作为训练数据，2002-08-20－2002-08-31共96 组作为验证数据．由于从气象台站获取的气象数据分辨率为 3 h，因此拟对目标区域颗粒物的质量浓度进行提前 3 h 预报．

2.2　输入参数选择

大气中颗粒物的质量浓度的时空分布是许多参数相互作用的结果，分为一次颗粒物和二次颗粒物．一次颗粒物主要由化石燃料燃烧、机械加工、熔炼等人为源产生，二次颗粒物主要来自于大气中的化学反应、转化和凝聚．这个转化过程受大气中其他污染物的质量浓度、气象条件等因素的影响．雨量和风影响大气中颗粒物沉积和迁移，云量、温度和湿度影响二次颗粒物的形成．前一时刻的颗粒物质量浓度在模型中作为表征初始条件的参数输入．在模型中利用多元逐步回归和经验判断[1-2]从气象站十多项常规气象观测数据中最终筛选出如下预报因子：

1) 初始条件　$\rho(PM_{10})_t$ 为当时的 PM_{10}的质量浓度平均值；

2) 预测气象条件　$S_{W,t+3}$为预测时段的平均风速；$D_{W,t+3}$为预测时段的平均风向；$H_{R,t+3}$为预测时段的相对湿度；C_{t+3}为预测时段的总低云量；T_{t+3}为预测时段的环境温度；R_{t+3}为预测时段的 6 h 降雨量．在实际预报中，气象数据将用气象模式预报值代替．

2.3　实验分析

将上述初始条件和气象条件共 7 项预报因子作为输入层；为了使实验结果更切合实际，隐含层神经元个数通过试错法(即通过试验，选择模型预测效果最好时的隐含层数目)得到，经反复试算，隐含层节点数选为 10；输出层为由输入层各因素所决定的 3 h 后的 $\rho(PM_{10})$ 平均值．神经网络预报模型最终确定为

7—10—1 结构，指定精度为 0.01. 训练样本集的部分原始数据见表 1. 原始数据经归一化处理后开始网络训练，预报结果经反归一化处理后输出.

表 1 神经网络训练样本集的部分数据

Table 1 Part of the training datasets of neural network

时间	$\rho(PM_{10})_t$/($\mu g \cdot m^{-3}$)	$S_{W,t+3}$/($m \cdot s^{-1}$)	$D_{W,t+3}$/(°)	$H_{R,t+3}$/%	C_{t+3}/成	T_{t+3}/℃	R_{t+3}/mm	$\rho(PM_{10})_{t+3}$/($\mu g \cdot m^{-3}$)
2002—07—20 T 00:00:00	88	0	0	89.5	7	22	0	33
2002—07—20 T 03:00:00	33	0	0	86.8	7	22.2	0.8	29
2002—07—20 T 06:00:00	29	1	320	81.2	7	22.8	0	49
2002—07—20 T 09:00:00	49	2	40	75.9	8	23.9	0	63
2002—07—20 T 12:00:00	63	2	360	65.5	7	26.6	0	73
2002—07—20 T 15:00:00	73	2	90	59.4	5	28.8	0.01	38
2002—07—20 T 18:00:00	38	1	110	78.6	8	25.7	0	82
2002—07—20 T 21:00:00	82	1	20	73.8	7	25	0.1	92
2002—07—21 T 00:00:00	92	0	0	80.3	5	24.2	0	84

采用 MATLAB 遗传算法程序包和神经网络工具箱有关函数编程求解. 利用遗传算法优化 BP 神经网络的操作参数为：权重初始化空间为[－1,1]，种群规模为 50，最大进化代数为 100，选择率为 0.09；交叉率为 0.9；变异率为 0.08.

在误差逼近的过程中，BP 网络初始权值和阈值对应的网络误差为 0.317 941，训练 80 次时误差为 0.093 316，163 次时达到指定精度；遗传算法优化的 BP 网络初始权值和阈值对应的网络误差为 0.082 655，训练 50 次时误差为 0.032 654，84 次时即达到指定精度. BP 和遗传 BP 的预测平均相对误差分别为 31%和 26%. 可见，遗传算法优化网络初始权值和阈值的方法有效地提高了网络的收敛速度和预报准确率. 用遗传算法优化后的神经网络模型的预测结果如图 2 所示. 可以看出，模型对于峰值的把握略有滞后，但总的来说能较好地跟踪污染物的质量浓度的变化趋势，具有良好的泛化能力.

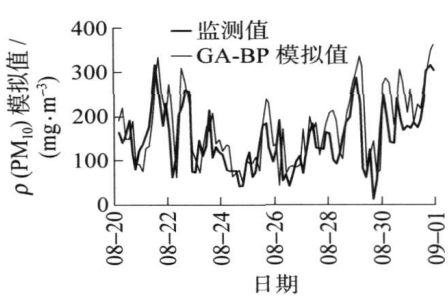

图 2 2002-08-20—2002-08-31模拟值与监测值对比

Fig. 2 Comparison between simulative and observed PM_{10} concentration during 2002-08-20—2002-08-31

3 与数值预报方法的对比

为将遗传算法改进的神经网络模型用于常规空气污染预报，把模型的输入和输出作了相应的调整，7、8 月份前 50 d 作为训练样本，后 12 d 作为验证样本，重新调试网络驯化和遗传算法操作参数，用来预报 $\rho(PM_{10})$次日日平均值.

将遗传算法优化后的 $\rho(PM_{10})$日平均值模拟结果与美国第 3 代空气质量模型 Models-3(CMAQ)的数值模拟结果作对比分析. 数值模拟结果引自北京工业大学于 2003 年承担的北京绿色奥运大气污染控制项目的研究报告. 项目通过区域环境现状调查、污染源清单建立、激光雷达观测、温风廓线仪遥测等工作，建立起 CMAQ 空气质量模式系统，数值模式采用 MM5、ARPS 气象模式与 CMAQ 空气质量模式的耦合模式，北京及周边区域数值模型调试为 3 层网格嵌套[7-8]. 应用该模型系统对北京市大气环境质量进行了

数值模拟[9-10]. 2 种方法的模拟结果对比如图 3 所示.

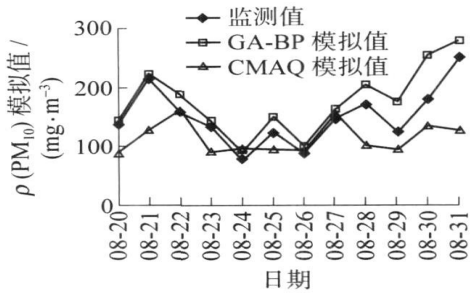

图 3　2002-08-20—2002-08-31两种型模拟效果比较

Fig. 3　Comparison of results between two models during 2002-08-20—2002-08-31

遗传算法改进的神经网络模型和 CMAQ 数值模型的模拟结果的平均相对误差分别为 0.21 和 0.26. 可见，神经网络模型在用于空气污染物质量浓度短期预报时，其效果并不逊于数值模型，是一种非常实用的空气污染预报方法.

由于神经网络模型通过输入与输出之间的响应关系得到的只是污染物在某时空内的质量浓度的平均值，而数值模式涉及多尺度范围(城市尺度及天气尺度)、多种污染物、多种过程考虑(污染物输送、扩散、化学转化以及清除过程)，具有高时空分辨率的特性(数千米分辨的污染物质量浓度小时变化)[11]，在进行常规质量浓度预报的同时还可以得到污染物的时空分布特征，对污染的形成机理和传输过程都有更深刻的反映.

4　结论

1) 相比传统的 BP 网络，采用遗传算法改进的 BP 神经网络进行空气污染预报，具有收敛速度快，准确率高的优点. 利用该法进行空气污染物短期质量浓度预测，能为城市(区域)空气质量管理提供依据.

2) 训练样本数的增加对网络泛化能力的提高有很大影响，利用神经网络进行空气污染预报需要海量的数据支持. 输入参数的选取是神经网络预测的关键因素，尝试获取更多与污染物的形成和转化相关的气象参数，可使预报精度进一步提高.

3) 利用神经网络方法建立的空气污染预报模型，与利用数值模拟方法建立的数值预报模型相比，具有数据易于获取、模型易于搭建、预测耗时短等优点，可以广泛用于空气污染物质量浓度的短期预报、空气污染指数预报，尤其适用于污染源排放、环境现状资料缺乏的城市和地区.

参考文献：

[1] GRIVAS G, CHALOULAKOU A. Artificial neural network models for prediction of PM_{10} hourly concentrations, in the greater area of Athens, Greece[J]. Atmospheric Environment, 2006, 40: 1216-1229.

[2] JEF H, CLEMENS M, GERWIN D, et al. A neural network forecast for dailyaverage PM_{10} concentrations in Belgium[J]. Atmospheric Environment, 2005, 39: 3279-3289.

[3] 黄牧涛，黄科娘. 混合遗传神经网络算法研究及其在径流预报中的应用[J]. 长江科学院院报，2007，24(4)：31-33. HUANG Mu-tao, HUANG Ke-lang. Daily runoff forecast system based on combined genetic algorithm and ANN[J]. Journal of Yangtze River Scientific Research Institute, 2007, 24(4): 31-33. (in Chinese)

[4] 金龙，吴建生，林开平，等. 基于遗传算法的神经网络短期气候预测模型[J]. 高原气象，2005，24(6)：981-987. JIN Long, WU Jian-sheng, LIN Kai-ping, et al. Short term climate prediction model of neural network based on genetic algorithms[J]. Plateau Meteorology, 2005, 24(6): 981-987. (in Chinese)

[5] ARCHONTOULA C, MICHAELA S, NIKOLAS S. Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens[J]. The Science of the Total Environment, 2003, 313: 1-13.

[6] 白晓平，张启明，方栋，等. 人工神经网络在苏州空气污染预报中的应用[J]. 科技导报，2007，25(3)：45-49. BAI Xiao-ping, ZHANG Qi-ming, FANG Dong, et al. Application of artificial neural network to air pollution prediction in Suzhou City[J]. Science & Technology Review, 2007, 25(3): 45-49. (in Chinese)

[7] CHEN D S, CHENG S Y, GUO X R, et al. An integrated MM5-CMAQ modeling approach for assessing trans-boundary PM_{10} contribution to the host city of 2008 Olympic summer games-Beijing, China[J]. Atmospheric Environment, 2007, 41: 1237-1250.

[8] CHEN D S, CHENG S Y, LI J B, et al. Application of LIDAR technique and MM5-CMAQ modeling approach for the assessment of winter PM_{10} air pollution: a case study in Beijing, China[J]. Water, Air and Soil Pollution, 2007, 181: 409-427.

[9] CHENG S Y, CHEN D S, LI J B, et al. An ARPS-CMAQ modeling approach for assessing the atmospheric assimilative capacity of the Beijing metropolitan region[J]. Water, Air and Soil Pollution, 2007, 181: 211-224.

[10] CHEN D S, CHENG S Y, GUO X R, et al. An integrated ARPS-MM5-CMAQ modeling approach for predicting PM_{10} concentration in the metropolitan region of Beijing in winter[J]. Environmental Informatics Archives, 2005, 3: 439-448.

[11] 房小怡, 蒋维楣, 吴涧, 等. 城市空气质量数值预报模式系统及其应用[J]. 环境科学学报, 2004, 24(1): 111-115.
FANG Xiao-yi, JIANG Wei-mei, WU Jian, et al. Study on the development of numerical model system to predict urban air quality[J]. Acta Scientiae Circumstantiae, 2004, 24(1): 111-115. (in Chinese)

Optimizing BP Networks by Means of Genetic Algorithms in Air Pollution Prediction

WANG Fang, CHENG Shui-yuan, LI Ming-jun, FAN Qing
(College of Environmental & Energy Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: Air pollution forecasting provides early warning before air pollution issue occurs, thus protects human health and living environment. A neural network model optimized by genetic algorithm was developed in order to predict PM_{10} concentrations in Beijing. The genetic algorithm was used to optimize the initial weights and threshold of the BP neural network in simulation. Astringency of network and accuracy of prediction were effectively improved. The improved network and Models-3 Community Multi-scale Air Quality (CMAQ) modeling system were both applied in the prediction of short-term PM_{10} concentration in autumn 2002 in Beijing. Results showed good prediction capability of both models, and the mean relative errors were separately 0.21 and 0.26. When applied in short-term air pollution forecasting, neural network is of similar prediction accuracy compared with CMAQ. It is an effective alternate method for air pollution forecasting in areas where mathematical model on air pollution can't be widely applied.

Key words: PM_{10}; neural network; genetic algorithm; CMAQ

(责任编辑 刘 潇)