

# 基于遗传神经网络的采暖季 PM<sub>10</sub> 浓度预测

刘艳华, 张承中, 王小平, 龚葳 (西安建筑科技大学环境与市政工程学院, 陕西西安 710055)

**摘要** 用 GA 算法和 BP 算法相结合的算法预测西安市 PM<sub>10</sub> 污染浓度, 首先采用 GA 算法优化 BP 神经网络模型的初始权重, 再用 BP 算法进行精确训练, 在此基础上进行浓度预报。实例表明 GA-BP 神经网络解决了 BP 神经网络收敛速度慢和易于陷入极小值的问题, 提高了预测精度。

**关键词** 遗传算法; BP 神经网络; 预测; PM<sub>10</sub> 浓度

**中图分类号** S127 **文献标识码** A **文章编号** 0517 - 6611 (2009) 16 - 07650 - 02

## Prediction of PM<sub>10</sub> Pollution Concentration in Heating Season Based on Genetic Algorithms and Neural Network

LIU Yan-hua et al (Environment and Municipal Engineering Institute, Xian University of Architecture and Technology, Xian, Shaanxi 710055)

**Abstract** The GA algorithm combining with BP algorithm was used to forecast PM<sub>10</sub> pollution concentration of Xian. Firstly, GA algorithm optimized BP neural network models of initial weight, using BP algorithm train the network accurately, forecasting the concentration. The example showed that GA-BP neural network solved the problems of BP neural network slow convergence speed and apted plunging infinitesimal data and increased the forecasting accuracy.

**Key words** Genetic algorithm; BP neural network; Forecasting; PM<sub>10</sub> concentration

目前国内外大气污染预报广泛采用污染潜势预报、污染浓度数值预报、污染浓度统计预报等模式。统计预报在国内外应用的较为广泛, 主要采用回归分析、时间序列、分类法、趋势分析法等预测方法<sup>[1]</sup>。近年来, 人工神经网络发展迅速, 其中 BP 神经网络具有很强的处理非线性问题的能力, 然而从全局范围内准确地预测存在一定的缺陷。遗传算法具有自适应性、全局优化性和隐含并行性, 体现出很强的全局搜索能力<sup>[2]</sup>。笔者利用遗传算法 (GA) 优化 BP 神经网络的 GA-BP 模型预测西安市 PM<sub>10</sub> 污染浓度, 为 PM<sub>10</sub> 污染浓度的精确预测提供科学依据。

### 1 遗传神经网络原理

**1.1 BP 神经网络** BP 神经网络由输入层、隐含层和输出层组成 (图 1), 层间各个神经元实现全连接, 同层神经元之间不连接。首先按随机给定的初始权值, 从输入层经隐含层逐层计算各单元的输出值, 然后计算模型实际输出值与期望输出值之间的误差, 如其不符合要求, 则沿误差最大的方向反传回去, 对各层权值阈值进行调整, 直至误差小于预定值或达到预定训练次数。

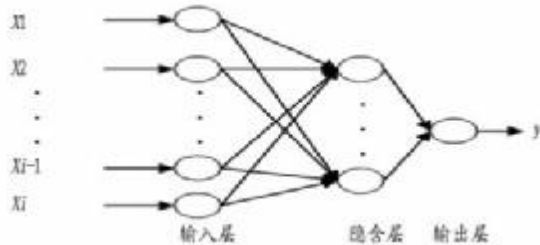


图 1 BP 神经网络结构

Fig. 1 BP neural network structure

**1.2 遗传算法** 遗传算法 (GA) 是一种基于自然选择和自然遗传的全局优化算法, 根据适应度的大小对神经网络的权系数和阈值的编码进行遗传操作, 得到最优解。遗传算法的

基本步骤为:

(1) 产生初始种群。随机生成包含  $n$  个个体的群体  $X = (X_1, X_2, \dots, X_n)^T$ , 每个个体代表一个神经网络的初始权值阈值分布, 每个基因值代表一个权值阈值, 用实数编码方法对权值阈值进行编码。

(2) 适应度函数评价。用训练样本对种群中的每个个体进行训练, 计算每个个体的学习误差, 从而确定适应度值, 学习误差如下:

$$E = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^p (y_j^k - o_j^k)^2 \quad (1)$$

式中,  $m$  表示训练样本个数;  $p$  表示输出层节点数;  $y_j^k, o_j^k$  表示测试样本的实际输出和期望输出。因此适应度函数为  $\text{fitness} = 1/E$ 。

(3) 遗传操作。遗传操作包括选择、交叉、变异 3 个部分, 选择是从群体中按照它们的适应度选出优良个体, 作为产生下一代的父体, 再对群体进行交叉、变异, 从而进一步实现群体的进化。

**1.3 GA-BP 混合算法神经网络** 将 GA 算法和 BP 算法结合起来, 形成 GA-BP 混合训练算法, 以 GA 优化 BP 网络的初始权值和阈值, 避免 BP 网络易陷入局部极小问题, 达到优化网络的目的, 提高模型的预测精度<sup>[3]</sup>。

GA-BP 网络具体实现步骤<sup>[4]</sup>如下: ①参数初始化, 包括遗传算法和 BP 神经网络的各个参数, 输入样本数据; ②在区间  $[-1, 1]$  上随机产生神经网络的一组权值阈值, 采用实数编码对其进行编码; ③适应度函数评价; ④遗传操作到设置的最大代数; ⑤将遗传操作后的个体解码为 BP 神经网络的权值、阈值初始值; ⑥用 BP 网络训练初始权值阈值, 直到到达一定的精度或一定的次数, 应用训练好的模型进行 PM<sub>10</sub> 浓度预报。

## 2 遗传神经网络预测 PM<sub>10</sub> 浓度

**2.1 气象因子相关性分析** 由于 PM<sub>10</sub> 的浓度与气象条件密切相关, 笔者采用西安市环境监测站提供的 2001 ~ 2008 年采暖季 (11 月 15 日 ~ 3 月 15 日) PM<sub>10</sub> 浓度实测资料和对的气象资料。气象要素包括 17 个, 分别是: 日均气压、日均

基金项目 陕西省自然科学基金项目 (2002E2-12)。  
 作者简介 刘艳华 (1982 - ), 女, 河北唐山人, 硕士研究生, 研究方向: 大气污染控制与技术。  
 收稿日期 2009-03-17

气温、最高气温、最低气温、日照时数、总辐射量、净辐射量、日均湿度、总云量、08 气压、08 气温、08 湿度、08 露温、平均风速、08 能见度和 14 能见度以及前日 PM<sub>10</sub> 浓度。采用二元变量相关分析法进行相关性分析。图 2 为前日 PM<sub>10</sub> 浓度和日均湿度与 PM<sub>10</sub> 浓度的相关性分析, 因子的相关系数越大对

PM<sub>10</sub> 浓度的影响越显著; 表 1 为 PM<sub>10</sub> 浓度与各个气象因子的相关系数。由表 1 可知, 选取前日浓度、14 能见度、08 能见度、最低气温、08 气温、日均气温、08 露温、最高气温、平均风速和日均气压 10 个气象影响因子为 GA-BP 网络模型的输入因子, 当日的 PM<sub>10</sub> 浓度为输出因子。

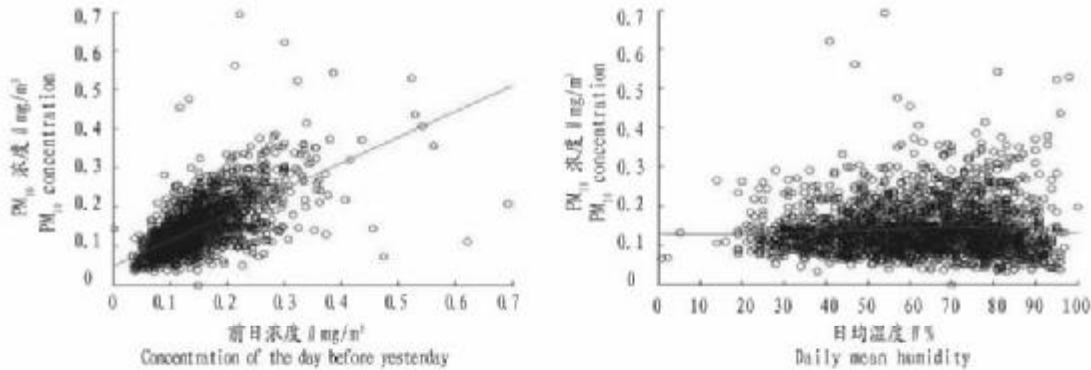


图 2 气象因子的相关性分析

Fig.2 Correlation analysis of meteorological factor

表 1 PM<sub>10</sub> 浓度与各个气象因子的相关系数

Table 1 PM<sub>10</sub> concentration and correlation coefficient of each meteorological factor

气象因子 Meteorological factor	相关系数 Correlation coefficient	气象因子 Meteorological factor	相关系数 Correlation coefficient
前日浓度	0.660	日均气压	0.200
14 能见度	0.404	08 气压	0.197
08 能见度	0.344	日照时数	0.170
最低气温	0.325	净辐射量	0.140
08 气温	0.319	总辐射量	0.063
日均气温	0.313	08 湿度	0.054
08 露温	0.295	日均湿度	0.051
最高气温	0.290	总云量	0.037
平均风速	0.259		

2.2 数据预处理 在对神经网络训练前, 为避免输入向量物理意义和单位的不同对 BP 神经网络模型的影响, 减轻网络训练时的难度, 对输入向量以及网络期望输出进行标准归一化处理<sup>[5]</sup>, 将它们转化为区间 [0, 1] 内的数。变量标准归一化方法如下:

$$S_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

式中,  $x_i$  表示原始样本数据;  $x_{\max}$ ,  $x_{\min}$  分别表示原始变量中的最大值及最小值;  $S_i$  表示变换后的  $x_i$ 。

2.3 预测结果 遗传算法中采用实数编码, 交叉概率取 0.30, 选择概率取 0.80, 变异概率 0.05, 初始种群个数  $N = 100$ 。以 2001 ~ 2006 年采暖季的数据为样本, 经过 300 次遗传迭代后, 求得其最优个体, 将最优个体解码为 BP 网络的初始权值和阈值, 以此为基础进行 BP 网络训练。神经网络结构为 10-21-1, 学习率为 0.01, 训练次数为 1 000, 误差目标为 0.005。借助 Matlab 平台, 利用训练好的 GA-BP 网络模型对 2007 ~ 2008 年采暖季的数据进行仿真预测, 预测结果见图 3。

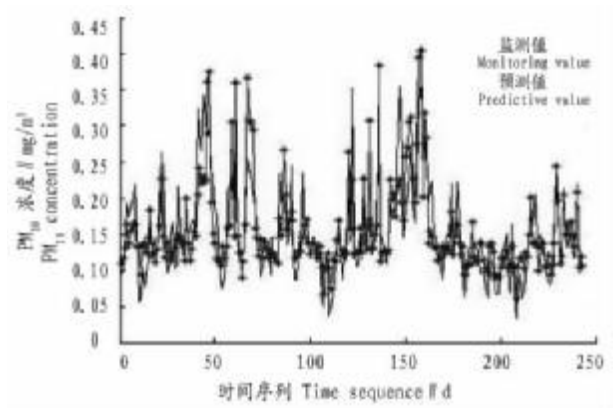


图 3 GA-BP 网络模型采暖季 PM<sub>10</sub> 浓度的预测曲线

Fig.3 Predictive curve of PM<sub>10</sub> concentrations in heating season based on GA-BP network model

### 3 结语

将 GA 算法和 BP 神经网络结合起来, 形成 GA-BP 混合训练算法, 以 GA 优化 BP 网络的初始权值和阈值, 再进行网络训练, 以此模型进行 PM<sub>10</sub> 污染浓度预报, 避免了 BP 神经网络收敛速度慢和易陷入局部极小的问题, 达到了优化网络的目的。笔者针对西安市采暖季 PM<sub>10</sub> 的浓度进行了预报, 实例应用证明 GA 算法和 BP 算法相结合的预报模型预测西安市 PM<sub>10</sub> 污染浓度具有较高的精度, 是提高预测精度的有效方法。

### 参考文献

- [1] 陈朝阳, 胡乐群, 万鹤群. 基于遗传算法的神经网络经济预测模型的建立[J]. 预测, 1997, 16 (1): 68 - 70.
- [2] 陈国良, 王煦法, 庄镇泉. 遗传算法及其应用[M]. 北京: 人民邮电出版社, 1996.
- [3] 苑希民, 李鸿雁. 神经网络和遗传算法在水科学领域的应用[M]. 北京: 中国水利水电出版社, 2002.
- [4] ZEALAND C M, BURN D H, SIMONOVIC S P. Short term stream flow forecasting using artificial neural net works [J]. Hydrol, 1999, 214: 32 - 48.
- [5] 吕锋, 石敏, 吴丹雯. 基于 GA 的 BP 神经网络模型的研究及应用[J]. 武汉理工大学学报, 2002, 24 (2): 63 - 66.