

# 基于BP神经网络模型的森林空气质量评价

王 琨<sup>1</sup>, 王文帅<sup>1</sup>, 张 夏<sup>2</sup>, 孙丽欣<sup>1</sup>

(1. 哈尔滨工业大学 市政环境工程学院, 哈尔滨 150090, wang02kun@163.com; 2. 中央财经大学, 北京 100081)

**摘要:** 为解决森林地区的空气评价问题, 基于BP神经网络技术进行建立森林空气质量评价模型的研究, 根据森林空气低污染、负离子高含量的特点, 以SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>、O<sub>3</sub>和负离子为评价指标建立模型, 并运用MATLAB工具软件实现. 以伊春五营森林和城镇空气质量数据为实例, 对模型进行对比检测, 评价结果分别为9.13和8.56分, 表明模型实用且可行.

**关键词:** 森林空气; 负离子; BP神经网络; 空气质量评价

中图分类号: X823

文献标志码: A

文章编号: 0367-6234(2010)08-1278-04

## Evaluation of forest air quality based on BP neural network

WANG Kun<sup>1</sup>, WANG Wen-shuai<sup>1</sup>, ZHANG Xia<sup>2</sup>, SUN Li-xin<sup>1</sup>

(1. School of Municipal and Environmental Engineering, Harbin Institute of Technology, Harbin 150090, China, wang02kun@163.com; 2. Central University of Finance and Economics, Beijing 100081, China)

**Abstract:** To solve the problem about the evaluation of forest air, an evaluation model based on BP neural network was studied. According to the characteristics of forest air with low pollution and high-volume negative ions, the model was constructed by MATLAB with the indices of SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>10</sub>, O<sub>3</sub> and negative ion. The evaluation results are 9.13 and 8.56 respectively by comparing the statistic data of air quality in forest with those in the town in Yichun, which proves the feasibility of the model.

**Key words:** forest air; negative ion; BP neural network; air quality evaluation

模糊数学、灰色模型、神经网络等理论的发展,使空气质量评价研究有了突破性进展<sup>[1-3]</sup>.但对于森林等生态区,空气中污染物含量小、特征性高,传统的大气评价方法不能充分说明森林地区空气的优异程度,其空气质量评价研究处于初级水平.为此,本文采用BP神经网络建立森林空气质量评价模型,对于保护森林,衡量其生态价值,有着实际意义.

人工神经网络(artificial neural network, ANN)是生物神经网络的抽象和模拟,是由大量人工神经元广泛连接而成的超大规模非线性动力学系统,具有大规模并行处理能力、信息分布存储、容错性和稳健性、自学习功能等特点.人工神经网络主要有以下几种典型模型:BP网络、

Hopfield网络模型、自组织神经网络、联想记忆神经网络、CMAC(感知器)模型.在ANN模型中应用较多的是具有非线性传递函数神经元构成的前馈网络中采用误差反传算法作为其学习算法的前馈网络(back propagation algorithm),简称BP网络. BP网络中信息单向传递,同一层中的神经元没有相关联系,而层与层之间则多采用全连接方式,连接程度由连接权重 $w_{ij}$ 表示,每层神经元状态都将影响下一层的神经元状态.神经元都有一个作用函数 $f$ 和阈值 $a$ ,其中,输入层神经元阈值为0,且 $f(x) = x$ ;而隐层和输出层的作用函数为非线性的Sigmoid型函数,一般采用 $f(x) = 1/(1 + e^{-x})$ <sup>[4]</sup>.基于BP神经网络的评价方法并不需要具体函数,由潜在函数可以得出任何一组质量指标对应的等级系数,可实现定量比较分析,且软件辅助快捷方便.本文选用这种方法进行森林空气质量评价.

收稿日期: 2008-11-26.

基金项目: 中国博士后科学基金(2005037655); 黑龙江省科技攻关项目(GC05C603-01).

作者简介: 王 琨(1964—),女,博士,副教授.

# 1 森林空气评价体系建立

$$\Delta S = k \frac{\Delta c}{c} \quad (2)$$

## 1.1 森林生态区空气质量特点

1) 污染物含量低. 森林一般远离市区, 周围少有工厂, 居民也不多, 污染物排放量很少. 且森林植物对污染物有降解、吸收作用, 森林空气中污染物含量很低, 甚至有部分污染物无法检出<sup>[5]</sup>. 传统评价模式只能说明森林空气质量高于一级标准, 却无法说明其同类区中的相对优良程度.

2) 空气负离子含量高. 空气负离子作为森林大气成分中一个重要组成部分, 其功能越来越被最近的研究所发现. 负氧离子对氮氧化物、硫化物、碳氧化物、卤化物、臭氧都有着不同的降解作用. 研究<sup>[6]</sup>表明, 人体吸入有“空气维生素”美誉的空气负离子后, 会使人产生兴奋, 进而调节中枢神经系统, 增加人体对氧气的吸入量和二氧化碳的排除量, 促进血液循环和新陈代谢. 而且在森林中, 空气负离子的含量很高<sup>[7]</sup>. 所以将负离子作为森林空气质量评价的指标之一.

基于以上两点, 急需建立森林等一类生态区的专用评价模式.

## 1.2 空气负离子指标评价体系

通常大气质量评价采用的指标是  $SO_2$ 、 $NO_2$ 、 $PM_{10}$ 、TSP 等 4 个指标, 但是由于本研究中 TSP 指标用的是粒度计数法, 同时无法取得当地空气粒子粒度计数和质量计数的关系. 再考虑到本研究中是洁净空气,  $PM_{10}$  可以代表空气中颗粒物含量水平. 因此, 放弃采用 TSP 指标, 选用  $O_3$  指标, 加入负离子指标, 共同形成森林空气评价指标体系.

根据我国部分省市气象部门各自公布的数据, 目前有 2 种空气负离子含量的等级标准: 5 级分类和 7 级分类. 这 2 种等级标准与城市大气三级分类均有差距, 为形成引入负离子的森林空气评价体系, 寻求一致, 本文采用等标标度法建立负离子三级等级标准.

### 1.2.1 等标标度法的基本原理

韦伯用心理物理试验证明了人对差别的感受不取决于它们之间的绝对差数, 而取决于它们之间的相对差数或比率. 而费希那则进一步推导出, 对于中等强度的刺激, 人的主观感觉量  $S$  和客观刺激量  $r$  之间的函数关系为

$$S = k \log r \quad (1)$$

其中:  $k$  为韦伯常数.

根据这一定理, 人们对大气污染物由于质量浓度变化对人体产生的主观差别感觉变化应为:

其中:  $c$  代表某种污染物的客观质量浓度, 它与式(1)中的  $r$  相当.

式(2)表明, 当污染物质量浓度成等比变化, 人的主观差别感觉成等差变化. 因此, 在指定某种大气污染物的三级标准质量浓度限值时, 尽管人体对该种污染物质量浓度值的客观刺激量产生的主观差别感觉变化是等距分级的, 而对该污染物质量浓度变化应予以等比赋值的指数标度<sup>[8]</sup>.

### 1.2.2 空气负离子标准限值的确定

若将空气负离子含量对人体有益的质量浓度限值规定为  $c_\infty$ , 对人体产生危害的空气负离子质量浓度限值规定为  $c_0$ , 在变化范围内等差地化为 10 个等级  $k_i (i = 0, 1, \dots, 9)$ , 其相应的  $k = 0, k = 9$  的主观感觉量分别为  $S_0, S_9$  相应的大气污染物的质量浓度限值. 其中  $S_3, S_5, S_7$  分别为对应一、二、三级标准质量浓度限值. 若设相邻两级的客观重要性比率为  $\alpha (\alpha > 1)$ , 则任意两级的客观重要性比率为

$$\frac{c_k}{c_l} = \alpha^{k-l} (k = 0, 1, \dots, 9; l = 0, 1, \dots, 9). \quad (3)$$

取  $l = 0$ , 则式(3)变为

$$c_k = c_0 \alpha^k \quad (4)$$

由  $c_k |_{k=9} = c_\infty$ , 确定相邻两级重要性比率

$$\alpha = \left( \frac{c_\infty}{c_0} \right)^{\frac{1}{9}} \quad (5)$$

将式(5)代入式(4)得

$$c_k = c_0 \left( \frac{c_\infty}{c_0} \right)^{\frac{k}{9}} \quad (6)$$

取  $k = 3, 5, 7$ , 分别计算空气负离子的三级标准质量浓度限值. 问题关键在于  $c_\infty$  和  $c_0$  的选择. 根据对人体的试验, 负氧离子为低于 40 个/ $cm^3$ , 将会诱发生理障碍, 引起头痛, 失眠神经衰弱, 过敏性疾病. 因此将其确定为  $c_0$ . 当高于 5 000 个/ $cm^3$  会增强人体免疫力, 于人类产生明显有益的影响, 因此确定为  $c_\infty$ . 于是计算出  $k = 3, 5, 7$  时的  $c$  值分别为 200, 584. 9, 1 709. 97, 近似取 200, 600, 1 700, 作为三级分类的质量浓度限制, 见表 1.

表 1 空气负离子三级分类标准

分级	一级	二级	三级
含量限值/(个· $cm^{-3}$ )	200	600	1 700

陈德明等<sup>[9]</sup>于 1994 年证明, 利用等标标度法来确定的三级空气质量标准的限值与环境空气质量标准基本一致, 2 种标准除  $CO$  和  $O_3$  的一级标准质量浓度限值差别较大以外, 其余污染物的标准级别质量浓度限值仅有微小差异. 而且这种差

异不会引起人的主观差别感觉有明显的变化,因此使用等标标度法等够很好符合我国大气质量标准体系. 本文所建立的空气负离子质量指标体系,与国家大气质量标准的分类,在方法上保证了统一. 两种标准具有含量与级别的隐含映射关系.

### 1.3 森林空气评价指标归一化

森林空气评价体系要求各项空气指标计量级别统一衡量. 大气质量标准中的污染物都用质量浓度单位,为  $\text{mg}/\text{m}^3$ ,而空气负离子采用计数浓度,即个/ $\text{cm}^3$ . 在单位上不统一,计量级别也不统一. 在进行数学处理时,为避免物理意义上会产生混淆,一般是将计数浓度转化为质量浓度.

为了解决这一问题,将采用数据的归一化来预处理数据. 归一化的具体作用是归纳统一样本的统计分布性. 归一化在 0~1 之间是统计的概率分布,归一化在 -1~1 之间是统计的坐标分布. 其基本的方法是线性函数转换,表达式如下:

$$y = (x - V_{\min}) / (V_{\max} - V_{\min})$$

其中:  $x, y$  分别为转换前、后的值,  $V_{\max}, V_{\min}$  分别为样本的最大值和最小值. 通过归一化处理的数据没有单位,分布在 -1~1 之间. 而且归一化处理后能够加快训练网络的收敛性.

## 2 森林大气质量 BP 模型的 MATLAB 实现

大气质量包含多个指标,最后要得出综合评价的一个等级或者数值,可根据大气质量指标来训练一个神经网络(函数关系),然后通过输入带评价的样本指标进行评价.

首先,使用大气质量标准数据来训练神经网络,对应的输出结果是质量的综合评分. 对应的理想状态,即各种污染物为零含量时打分为 10 分,背景质量浓度限值对应的打分为 9 分,一级标准为 8 分,二级标准为 6 分,三级标准为 4 分,建立一个大气质量和评分的复杂关系. 然后输入待评价的质量指标,按照上面训练的网络进行模拟,得

出相应的评分值. BP 模型涉及复杂的计算,可通过 MATLAB 软件进行, MATLAB 自带有神经网络工具箱,里面有各种神经网络的函数,方便用户的使用. 主要的函数是 newff, train, sim; 在这个函数里有几个参数十分重要,可以人为设定,分别是: 最大训练步数: net.trainparam. epochs, 简称为 epochs; 目标函数误差: net.trainparam. goal, 简称为 goal; 学习速率: net.trainparam. lr, 简称为 lr. Newff 在确定网络结构后,会自动调用 init 函数,用缺省函数来初始化网络中各个权重,产生一个可训练的前反馈网络,即函数的返回值 net.

建立网络后就用 train 函数对网络进行训练,其语法是: Net = train(net, p, t), 其中 p 和 t 分别为输入和输出矩阵, net 为由 newff 产生的要训练的网络. Train 会根据在 newff 函数中确定的训练函数,来训练网络.

对训练后的网络就何以用来仿真评价,其语法是: Y = sim(net, a), 其中 net 为训练后的网络; a 为输出数据, y 为所求的评价结果.

建立适合森林大气质量评价的 BP 模型的参数的选择主要包括:网络的层数,隐含层的神经元个数,学习速率和期望误差的选取. 以隐含层的神经元数的选择为例,神经元太少,网络学习不够,需要多次训练,训练精度也不高,网络隐含层神经元的个数  $s_1$  越多,功能越强,但当神经元数太多,训练时间随之增加. 另外,在进行函数逼近时,  $s_1$  过大,可能导致不协调的拟合.

本文以五营森林空气质量数据进行模拟评价. 首先对相同的 lr、epoch、goal 值,取不同的  $s_1$  进行评价,结果见表 2. 不同的  $s_1$  取值,最后的评价结果有明显偏差,有部分结果甚至不合理. 当  $s_1$  取 2 时,结果超过 10,明显不正确. 因此,选择一个合适正确的  $s_1$  值,需要经验判断. 对本研究中待评价的空气质量的基本判断是,其空气质量的每一项指标要低于环境的背景限值,也就是说评分应在 10~9 之间.

表 2 相同 lr, epoch 和 goal 不同  $s_1$  的评价结果

$s_1$	real epoch 层数	Performance 运算值	评价结果	$s_1$	real epoch 层数	Performance 运算值	评价结果
2	5	$4.03 \times 10^{-8}$	10.393 8	8	3	$1.13 \times 10^{-9}$	11.421 5
3	3	$3.29 \times 10^{-7}$	7.317	9	3	$2.10 \times 10^{-11}$	8.419 7
4	4	$1.65 \times 10^{-9}$	8.703	10	3	$7.31 \times 10^{-10}$	9.146 4
5	3	$5.67 \times 10^{-9}$	9.430 9	12	3	$4.37 \times 10^{-7}$	10.088 9
6	3	$7.63 \times 10^{-12}$	8.510 8	15	3	$2.81 \times 10^{-7}$	8.429 1
7	3	$2.68 \times 10^{-9}$	8.137 4	20	2	$4.03 \times 10^{-9}$	8.936 9

注: lr=0.03, epoch=1 000, goal=0.000 001.

从上表中看出,  $s_1$ 取 5 和 10 比较合理. 用同样的方法对哈尔滨市大气质量<sup>[10]</sup>和森林空气质量进行比较.  $s_1$ 取不同的值,结果如表 3.

表 3 不同  $s_1$  取值下城市和森林空气评价结果

$s_1$	五营森林	哈尔滨市	$s_1$	五营森林	哈尔滨市
2	10.393 8	4.519 3	8	11.421 5	4.374 0
3	7.317	2.901 6	9	8.419 7	3.704 5
4	8.703	5.084 1	10	9.146 4	6.654 4
5	9.430 9	4.422 8	12	10.088 9	5.455 4
6	8.510 8	4.561 1	15	8.429 1	9.208 2
7	8.137 4	6.505 7	20	8.936 9	2.859 0

该文献中的地区大气质量评价为 4 级,按上面的标准,也就是说得分在 5 分以下,比较可以看出,  $s_1$ 取 2,3,5,8,9,20 比较合理. 综合两个地区以及经验  $s_1$ 的选取,取值为 5 最为合理. 仿照上面的思路对其他参数进行选择,最终的参数选择如表 4.

表 4 森林大气质量评价模型 MATLAB 实现参数选择

网络层数	隐含层神经元数	学习速率	期望误差
3	5	0.03	0.000 1

### 3 伊春五营森林空气评价及模型检验

依据《环境空气质量标准》(GB 3095—1996)、清洁空气主要成分和森林空气特点,确定空气监测主要物质和物理参数,具体为  $O_2$ 、 $CO_2$ 、 $CO$ 、 $NO_x$ 、 $SO_2$ 、 $NH_3$ 、臭氧、苯系物、可吸入颗粒物  $PM_{10}$ 、空气负离子.

在伊春五营原始森林保护区(丰林区)选取了 2 个测点,平原选取了 2 个测点,其中丰林区测点 1 海拔最低,丰林区测点 3 海拔最高,且属于旅游风景点. 为了和丰林保护区的数据进行比较,又在五营城镇区居民定居点和室内选取了 2 个对比测试点. 采用数据均为日平均数据,见表 5.

表 5 空气指标日平均数据

测点	质量浓度/ $(\mu g \cdot m^{-3})$				负离子含量/ ( $\text{个} \cdot \text{cm}^{-3}$ )
	$SO_2$	$NO_2$	$O_3$	$PM_{10}$	
背景限值	20	20	50	20	—
林区	4.77	1.50	2.23	7.74	2 027
城镇	19.00	10.93	2.44	36.53	1 530

MATLAB 最终运行结果是五营林区和城镇分别为 9.13 分和 8.56 分. 由表 5 可以看出,五营森林空气质量低于环境背景限值,使用普通的分级评价标准,任何方法测评均属于一级标准,并不能说明其空气作为森林环境空气的优秀程度,本文正是针对这一点,提出能够量化森林空气质量的方法,更好地说明森林等空气洁净生态区的空气

质量,同时也验证了本文所建立模型的可靠性.

### 4 结 论

1) 考虑到空气负离子在森林大气质量中的重要作用,用等标度法建立了三级空气负离子指标体系. 应用归一化方法,将空气负离子和污染物指标综合起来,形成了适合森林大气质量的指标体系. 引入空气负离子指标进行空气质量评价,为专类空气评价研究提供了新的指示和参考.

2) 空气质量评价体系应用 MATLAB 软件实现,经过训练学习得到最终参数:神经层数为 3,隐含层神经元数为 5,学习效率为 0.03,期望误差为 0.000 1.

3) 以伊春五营林区与城镇数据为实例对比评价,结果表明:林区为 9.13 分,城镇为 8.56 分,量化性说明了五营林区空气的优良程度,验证了所建立模型的可靠性及必要性.

### 参考文献:

- [1] CANEVALE C, FINZI G, PISONI E, *et al.* Neuro-fuzzy and neural network systems for air quality control [J]. Atmospheric Environment, 2008,7(64):1-11.
- [2] TZU-YI P, KEISUKE H, HSIN-HSIEN H, *et al.* Using grey system theory to evaluate transportation effects on air quality trends in Japan [J]. Transportation Research Part D, 2007, 1(7):158-166.
- [3] 熊忠华,陈琦,郑秀梅,等. 基于遗传算法的人工神经网络大气环境评价[J]. 环境科学与技术,2005(4):83-85.
- [4] MARTIN J D, MORTON Y T, ZHOU Qihou. Neural network development for the forecasting of upper atmosphere parameter distributions [J]. Advances in Space Research,2005,(36):2480-2485.
- [5] 刘东生. 我国森林资源监督管理的理论与实践研究 [D]. 北京:北京林业大学,2008.
- [6] JOVANIC B R,JOVANIC S B. The effect of high concentration of negative ions in the air on the chlorophyll content in plant leaves [J]. Water Air & Soil Pollution, 2001, 129: 259-265.
- [7] 曾曙才,苏志尧,陈北光. 我国森林空气负离子研究进展[J]. 南京林业大学学报:自然科学版,2006,30(5):107-110.
- [8] 李祚泳. 大气环境质量评价的标度指数法[J]. 中国环境科学,1998,18(5):433-436.
- [9] 陈德明,李祚泳. 大气环境质量的物元分析评价法[J]. 环境科学研究,1994(2):24-28.
- [10] 袁春欢. 哈尔滨市大气可吸入颗粒物组分及来源研究[D]. 哈尔滨:哈尔滨工业大学,2006.