

# 遗传神经网络在累积性环境风险评价中的应用

陈凯, 黄蕾, 方强

(污染控制与资源化研究国家重点实验室, 南京大学环境学院, 江苏 南京 210046)

**摘要:**以太湖流域常州段为研究对象, 构建了累积性水环境风险评价指标体系, 利用主成分分析法选取输入变量, 并应用MATLAB建立遗传神经网络综合评价模型。运用遗传算法对BP神经网络的权值和阈值进行优化, 将遗传算法全局搜索能力和BP算法局部搜索能力相结合, 提高了收敛速度和精度。应用模型对2004—2009年常州市累积性水环境风险进行了综合评价, 结果表明, 2004—2009年风险综合指数总体上处在中级与高级之间, 累积性水环境风险较大; 2008—2009年风险综合指数不断增大, 趋于低级; 农业和畜禽养殖业等面源风险源、污水处理和风险管理投资等控制机制以及人口和环境敏感目标等风险受体是造成太湖流域常州段累积性水环境风险较大的主要原因。

**关键词:** 遗传神经网络; 累积性环境风险评价; 太湖流域

中图分类号: X824

文献标识码: A

文章编号: 1674-6732(2012)-02-0001-06

## Application of Accumulative Environmental Risks Assessment Based on Genetic Neural Network Model

CHEN Kai, HUANG Lei, FANG Qiang

(State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing, Jiangsu 210046, China)

**ABSTRACT:** An assessment index system for accumulative water environmental risks was established according to the situation of Changzhou, part of Taihu lake basin. Principal component analysis was used to select input variables. Then, MATLAB was used to establish a comprehensive neural network model for evaluation. In this model, genetic algorithm was used to optimize the weight and threshold of network in order to improve the constringency rate and accuracy by taking advantage of both the overall searching ability of GA and the local search of BP. This model was applied to evaluate the index system for accumulative water environmental risks in Changzhou during 2004 to 2009. The results showed that during 2004 to 2009, Changzhou's composite index of accumulative water environmental risk was between intermediate level and severe state; the risk composite index was on the rise during 2008 to 2009, and tending toward lower level; non-point risk sources caused by agriculture and livestock farming, risk control mechanism such as sewage treatment and, investment of risk management and risk receptors including people and environment-sensitive targets were the main reasons which led to the high accumulative environmental risks in Changzhou.

**KEY WORDS:** genetic neural network; accumulative environmental risks assessment; Taihu lake basin

近几十年来,随着太湖流域经济的快速发展和人口规模的迅速扩大,社会经济和环境保护之间的矛盾日益尖锐,太湖水污染日趋严重<sup>[1]</sup>。频繁发生的水污染事件对太湖流域水环境风险评价提出了更高要求。累积性环境风险是指人类开发活动中潜在的会对人类健康、生态环境产生危害的行为,这种风险一般并不立刻显现,而且比较隐蔽,但对人类健康、生态环境却具有长远的影响<sup>[2]</sup>。

目前,国内针对区域累积性水环境风险的综合评价研究不多<sup>[3,4]</sup>。黄晓容等利用层次分析法构建了重庆三峡库区渐变水环境污染事故预警指标体系<sup>[3]</sup>;李万莲从水生态系统的社会—经济—自然属

性角度构建了蚌埠城市水生态环境评价指标体系,并用层次分析法进行了预警分析<sup>[5]</sup>;郭松影等从水安全风险的角度出发,建立了水安全风险评价指标体系,并应用系统动力学方法建立了水安全风险评价模型<sup>[6]</sup>;王立刚等根据中国农区水体环境质量的实际情况,提出了农区水体环境质量预警指标体系<sup>[7]</sup>。

收稿日期: 2011-06-20; 修订日期: 2011-08-08

基金项目: 国家水体污染控制与治理科技重大专项项目(2009ZX07527-008, 2009ZX07528-005); 江苏省环境监测科研基金项目(916); 中美国际合作项目(2010DFA91910)。

作者简介: 陈凯(1987—), 男, 本科, 从事环境风险分析与管理方向的研究。

当前国内主流的评价指标筛选和综合评判方法包括层次分析法、专家咨询法、主成分分析法、灰关联分析法、熵值法等<sup>[8]</sup>。这些方法在实际应用中具有极大的主观性,处理非线性问题能力较差,且缺乏自学习、自适应能力。BP 神经网络(Back-propagation Neural Network)虽然是处理非线性问题的有效工具,但也存在一些缺陷:学习收敛速度很慢,容易陷入局部最小点,网络结构不易确定等<sup>[9]</sup>。遗传算法(Genetic Algorithm, GA)具有全局搜索能力,同时具有较强的鲁棒性和自适应性<sup>[10]</sup>,将 GA 与 BP 神经网络相结合,用于优化网络的初始连接权值和阈值,可以提高算法的全局搜索能力和局部搜索能力,使结果性质得以极大地改善<sup>[11]</sup>。笔者利用遗传神经网络模型,对 2004—2009 年太湖流域常州段的累积性水环境风险等级进行评价,找出影响常州段累积性水环境风险的主要因素。

### 1 遗传神经网络模型的构建方法

BP 神经网络由 RUMELHART、MCCLELLAND 等研究并设计,基于误差反向传递算法的一种多层前馈型神经网络<sup>[11]</sup>,对解决非线性问题有着独特的优势,同时它还具有很强的输入输出非线性映射能力和易于学习、训练的优点<sup>[12]</sup>。遗传算法是模仿自然界生物进化机制发展起来的随机全局搜索和优化方法,它借鉴了达尔文的进化论和孟德尔的遗传学说,本质上是一种高效、并行、全局搜索的方法<sup>[13]</sup>。应用遗传算法优化 BP 神经网络的步骤见图 1。

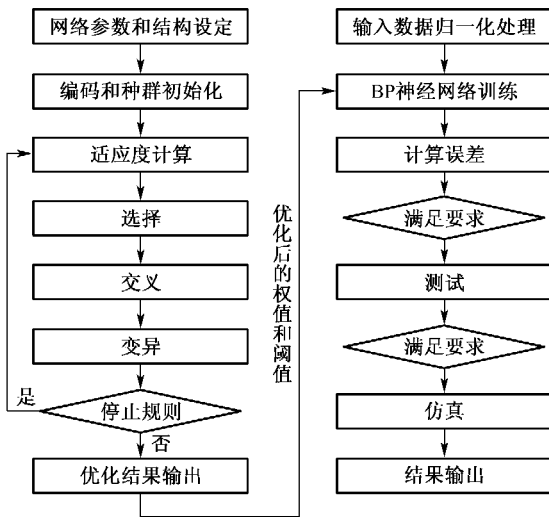


图 1 遗传算法优化 BP 神经网络的步骤

笔者采用浮点数编码方式,将权值和阈值连在一起组成染色体,染色体的长度  $S = R \times S_1 + S_1 \times S_2 + S_1 + S_2$ 。

式中: $R$ ——输入数量; $S_1$ ——隐含层结点数; $S_2$ ——输出数量。

染色体  $i$  的适应度值  $f_i$  用其对应的网络误差  $E_i$  的倒数表示。遗传算子操作方面,选择基于归一化的优先选择法(normGeomSelect),以排序策略为基础,进行选择操作。由于笔者采用实数编码方法,因此采用浮点型的算术交叉(arithXover)进行交叉操作<sup>[14]</sup>。变异操作采用非均匀变异(nonUnif-Mutation)<sup>[14]</sup>,对原有的基因值作一随机扰动,以扰动后的结果作为变异后的新基因值。

## 2 基于主成分—遗传神经网络的累积性水环境风险评价实证研究

### 2.1 指标体系的构建

环境污染事件是在环境风险场的复合作用机制下形成的,环境风险场包括风险源、环境风险控制机制、风险受体 3 个因素,它们共同决定了水环境风险的水平<sup>[2]</sup>。笔者从风险源的危险性、控制机制的有效性以及风险受体的易损性 3 个方面入手,构建太湖流域累积性水环境风险评价指标体系,如表 1 所示。指标分为正影响指标和负影响指标:正影响指标值越大,风险综合指数越大;负影响指标则反之。

表 1 太湖流域累积性水环境风险评价指标体系

一级指标	二级指标	三级指标	四级指标
风险源	面源污染	农业	Y1 化肥施用强度(折纯)/(kg·hm <sup>-2</sup> )(-)
		污染	Y2 农药使用强度/(kg·hm <sup>-2</sup> )(-)
	畜禽养殖业	畜禽	Y3 畜禽养殖业 TN 排放量/t(-)
		养殖业	Y4 畜禽养殖业 TP 排放量/t(-)
	生活污水	生活	Y5 人均日生活用水量/[L·(人·d) <sup>-1</sup> ](-)
		污水	Y6 单位 GDP 工业废水 COD 排放量/(kg·万元 <sup>-1</sup> )(-)
点源风险源	工业废水	Y7 单位 GDP 工业废水排放量/(t·万元 <sup>-1</sup> )(-)	
控制机制	初级控制机制	排污达标	Y8 工业废水排放达标率/%(+)
	次级控制机制	污水处理	Y9 城镇生活污水集中处理率/%(+)
	风险管理投资	环保投资	Y10 环保投资占 GDP 比重/%(+)

续表 1

一级指标	二级指标	三级指标	四级指标
风险受体	人口特征	流域内人口数量	Y11 人口密度/(人·hm <sup>-2</sup> )(-) Y12 人口自然增长率/%(-)
		人群生活水平	Y13 人均GDP/(万元·人 <sup>-1</sup> )(+)
	环境敏感目标	人口结构	Y14 老年(60岁以上)人口比重/%(-)
		水环境质量	Y15 集中式饮用水水源地水质达标率/%(+) Y16 水环境功能区水质达标率/%(+)
	基础设施	医疗卫生条件	Y17 每千人拥有卫生机构床位数/(个·千人 <sup>-1</sup> )(+)
		交通便利程度	Y18 人均实有铺装道路面积/(m <sup>2</sup> ·人 <sup>-1</sup> )(+)

注:(-)代表负影响指标,(+)代表正影响指标。

### 2.2 数据的获取和处理

笔者选取2004—2009年常州市累积性水环境风险评价指标数据进行评价,指标数据如表2所示。数据主要来自《常州市统计年鉴2006—2010》、《江苏统计年鉴2006—2009》和《常州市区环境状况公报2005—2009》。指标的最大值、达标值和最小值的选取主要根据国际、国家、行业和地方法规和标准,并结合常州市的社会经济发展情况和太湖治理要求综合考虑设定。

表2 常州累积性水环境风险评价指标数据

指标	2004年	2005年	2006年	2007年	2008年	2009年	最小值	达标值	最大值
Y1/(kg·hm <sup>-2</sup> )	516.72	509.66	538.77	475.05	452.71	470.11	120	225	364
Y2/(kg·hm <sup>-2</sup> )	25.40	37.34	34.33	31.08	27.57	26.75	3	10	20
Y3/t	6 089	5 911	5 218	5 204	5 643	5 779	3 000	5 000	6 000
Y4/t	2 228	2 213	1 974	1 977	2 135	2 167	1 500	1 800	2 000
Y5/[L·(人·d) <sup>-1</sup> ]	176.45	181.6	204.98	217.6	234.9	232.1	200	240	300
Y6/(kg·万元 <sup>-1</sup> )	3.90	3.94	2.28	1.62	1.25	1.05	1	4	10
Y7/(t·万元 <sup>-1</sup> )	20.94	24.9	17.2	16.7	14.4	14.3	10	30	50
Y8/%	98.8	98.6	98.1	98.79	98.85	100	60	95	100
Y9/%	19.09	29.46	60.57	63.43	76	78.86	60	85	95
Y10/%	2.83	2.4	3.55	2.46	2.32	2.4	1.5	3.5	5
Y11/(人·hm <sup>-2</sup> )	798	804	811	993	1 005	1 015	132	452.3	752
Y12/%	1.1	2.1	1.9	1	0.2	-0.4	1	4	9
Y13/(万元·人 <sup>-1</sup> )	3.2	3.2	3.7	4.4	5	5.7	0.25	3.3	6.94
Y14/%	15.26	15.89	16.51	17.2	17.92	18.77	10	12	17
Y15/%	99.7	100	99.8	100	100	100	80	95	100
Y16水/%	39.4	38.5	58.5	29.3	41.5	66.2	60	85	100
Y17/(个·千人 <sup>-1</sup> )	3.54	3.55	3.75	4.13	4.21	4.49	3	3.72	5
Y18/(m <sup>2</sup> ·人 <sup>-1</sup> )	4.72	5.34	6.08	6.29	7.27	7.81	3	6	10

由于BP神经网络的隐含层一般采用Sigmoid转换函数,要求输入数据的值在0~1之间<sup>[15]</sup>。因此,不同量纲、没有可比性的各项指标数据需要进行标准化处理。笔者采用极差标准化方法。

对于正影响的指标采用公式:

$$Ri = (Xi - Xi_{min}) / (Xi_{max} - Xi_{min}) \quad (1)$$

对于负影响的指标采用公式:

$$Ri = (Xi_{max} - Xi) / (Xi_{max} - Xi_{min}) \quad (2)$$

式中:Ri——指标量化值;Xi——各指标实际

值;  $X_{i_{max}}$ ——指标最大值;  $X_{i_{min}}$ ——指标最小值。对于正影响指标,若其实际值大于最大值,指标量化值为1,若其实际值小于最小值,指标量化值为0;负影响指标反之。

### 2.3 主成分分析

主成分分析法(Principal Component Analysis, PCA)是利用降维的思想,把多指标转化为少数几个综合指标,即主成分,而这几个主成分可以反映原来多个变量的大部分信息的一种统计方法<sup>[16]</sup>。该法不仅可以反映原有指标的信息量,而且可以解决指标之间信息重叠和权重选取问题,从而达到降维并减少计算量的目的。由于神经网络模型的精度和泛化能力易受输入样本变量的影响,利用主成分分析方法可以消除变量间的相关性,对输入空间进行重构,消除重叠信息,提取主导因素,优化网络结构,提高模型精度<sup>[17]</sup>。

利用 SPSS 软件对指标数据进行主成分分析,得到旋转后的主成分贡献率如图 2 所示。4 个主成分的累积方差贡献率为 98.631%,因此可用这 4 个主成分因子来代替 18 个指标作为输入变量。主成分因子 2 中 Y3、Y4 有较大载荷,即反映畜禽养殖业 TN 和 TP 排放量的信息;主成分因子 3 中 Y10、

Y16 有较大载荷,即反映环保投资占 GDP 比重和水环境功能区水质达标率的信息;主成分因子 4 中 Y2 有较大载荷,即反映农药使用强度的信息;主成分因子 1 中剩余 13 个变量有较大载荷,即反映了除畜禽养殖业外的风险源信息、初级控制机制信息和除主成分因子 4 中 Y2(反映农药使用强度的信息)外的风险受体信息。求得各项主成分因子的得分函数,进而得到各主成分因子的得分矩阵和各主成分因子的最小值、达标值和最大值。之后进行标准化处理,得到位于(0,1)之间的主成分输入数据,如表 3 所示。

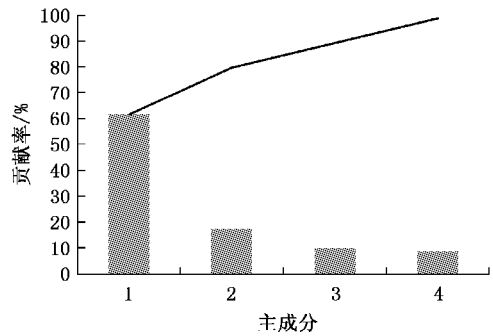


图2 旋转后主成分方差贡献率

表3 归一化后主成分因子数据

主成分	2004年	2005年	2006年	2007年	2008年	2009年	最小值	达标值	最大值
C1	0.12	0.14	0.44	0.62	0.81	0.93	0.00	0.44	1.00
C2	-0.91	-0.66	0.17	0.01	-0.27	-0.40	0.00	0.40	1.00
C3	-0.17	-0.20	0.45	-0.36	-0.17	0.30	0.00	0.38	1.00
C4	0.12	-0.44	-0.26	-0.07	0.13	0.20	0.00	0.72	1.00

### 2.4 遗传神经网络模型的构建

应用 MATLAB 神经网络工具箱和遗传算法工具箱(GAOT)编写程序,实现遗传神经网络模型的构建。模型构建步骤如下。

#### 2.4.1 BP 网络设计与遗传算法参数设定

根据 Kolmogorov 定理,对含一个隐含层的 3 层 BP 网络,只要隐含层神经元数目足够,便能以任意精度逼近有界区域上的任意连续函数<sup>[18]</sup>。因此,笔者建立的是 3 层 BP 网络。输入层神经元数目取决于输入矢量的维数<sup>[19]</sup>,即为 4。隐含层神经元个数的确定尚没有绝对准确的方法,一般通过经验公式和试错法获得<sup>[15]</sup>。笔者根据 Kolmogorov 定理,经多次测试后发现(2n + 1)个隐含层神经元数(n 为输入层神经元数)可以使模型学习和预测效

果最佳,因此确定隐含层神经元数为 9。输出层包含一个神经元,即风险综合指数 P。根据各主成分因子数据(归一化后)的最小值、达标值、最大值,设定风险综合指数范围和相应的风险等级,如表 4 所示。

表4 水环境风险评价分级标准

指标	高级	中级	低级	极低级
C1	<0	0~0.44	0.44~1	>1
C2	<0	0~0.4	0.4~1	>1
C3	<0	0~0.38	0.38~1	>1
C4	<0	0~0.72	0.72~1	>1
P	<0	0~0.6	0.6~1	>1

隐含层和输出层神经元的传递函数分别取

tansig 函数和 purelin 函数。选取以学习率可变的动量 BP 算法 traingdx 函数进行网络训练。设定学习率为 0.01, 误差为 0.000 1, 循环次数为 3 000 次。所构建的 BP 网络模型结构如图 3 所示:

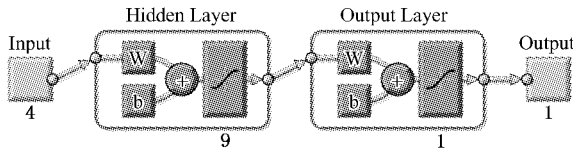


图 3 BP 神经网络结构

种群的规模大小直接影响到遗传算法的收敛性,过小容易收敛到局部最优解,过大则会造成计算速度降低,一般设定在 10 ~ 200 之间<sup>[13]</sup>。经过多次试验后选定种群规模设定为 190;进化代数设定为 500 代。按照 HOUCK 关于 GAOT 工具箱的设定,归一化优先选择的参数设定为 [0.08], 算术交叉的参数设定为 [2 0], 非均匀变异的参数设定为 [4 gen 3]<sup>[14]</sup>。

2.4.2 训练和检验样本的生成

由于环境风险评价的神经网络研究尚处于起步阶段,没有一个通用的评价标准,因此笔者参照有关文献的选取方法<sup>[10,13]</sup>,根据所用样本数据的最大值、达标值、最小值之间的区间,进行 3 次样条内插,设定风险等级,构建 41 组训练样本和 6 组检验样本(包括各项指标标准化值和综合指数值)。

2.4.3 网络训练和检验

单独用 BP 神经网络进行训练时,当网络训练到第 441 次时,达到了系统误差  $se < 0.000 1$  的停止条件,而用遗传神经网络训练时仅需要 127 步即可达到(见图 4、图 5)。6 组检验样本测试结果表明,该网络对每一组测试样本的相对误差不超过 0.05,由此可以认为网络模型训练成功。

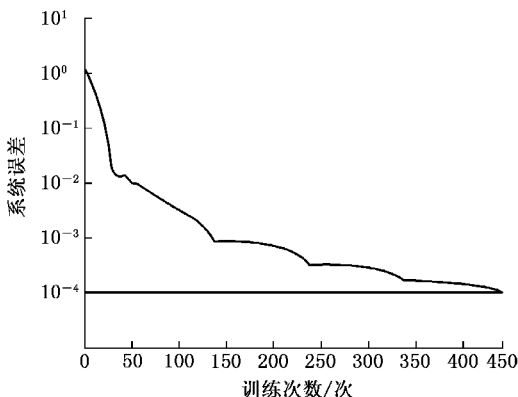


图 4 BP 神经网络训练误差

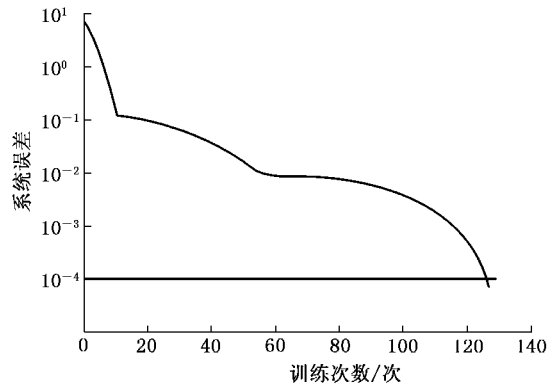


图 5 遗传神经网络训练误差

3 结果

应用构建好的遗传神经网络模型,将常州市 2004—2009 年的指标标准值输入,得到太湖流域常州段累积性水环境风险综合指数级别,如表 5 所示。

表 5 累积性水环境风险等级评价结果

	2004 年	2005 年	2006 年	2007 年	2008 年	2009 年
P	0.186 2	-0.101 2	0.457 1	-0.405 9	0.013 7	0.711 9
风险等级	中	高	中	高	中	极低

4 结论

(1) 总体上来看,2004—2009 年太湖流域常州段的累积性水环境风险评价等级徘徊在中级与高级之间,除 2006 和 2009 年外,风险综合指数均低于 0.2,累积性水环境风险较大。这与《太湖流域及东南诸河水资源公报》(2004—2009)中太湖常州段(竺山湖)水质均劣于 V 类的实际情况基本相符,说明运用遗传神经网络模型进行累积性水环境风险评价是可行的。

(2) 从趋势上来看,2004—2007 年风险综合指数呈波动变化,趋势并不明显。其中,2006 年风险综合指数大幅增加,这主要是由于农药使用强度、畜禽养殖业 TP 年排放量、单位 GDP 的废水排放量和单位 GDP 的 COD 排放量降低以及城镇生活污水集中处理率和环保投资占 GDP 比重增加所致。2007—2009 年,除人均日生活用水量、老年人口所占比重和畜禽养殖业 TP 年排放量明显增加外,其他指标大都有所好转,尤其是城镇生活污水集中处理率和水环境功能区水质达标率有了较大幅度提高,风险综合指数不断增加,到 2009 年达到

低风险级别。这表明2007年太湖蓝藻事件之后所采取的应对措施是有效的。

(3) 从评价等级上来看,2005和2007年的风险综合指数分别处于高风险等级。2005年与2004年相比,17项指标相差不大,风险综合指数都较低;而农药使用强度有了大幅增加,导致2005年的风险等级下降。这说明农业面源风险是导致2005年常州市水环境处于高风险的主要原因。2007年的风险综合指数远低于其他年份,主要是由于水环境功能区水质达标率、环保投资占GDP比重大幅下降和人口密度大幅增加有关。这说明次级控制机制有效性降低、风险受体易损性增加是导致2007年水环境处于高风险的主要原因。

(4) 从各项指标达标情况来看,化肥施用强度(折纯)、农药使用强度、畜禽养殖业TP和TN年排放量、城镇生活污水集中处理率、环保投资占GDP比重、人口密度、老年人口所占比重、水环境功能区水质达标率等9项指标均低于达标值。因此,农业和畜禽养殖业等面源风险源、污水处理和风险管理投资等控制机制以及人口和环境敏感目标等风险受体是造成太湖流域常州段累积性水环境风险较大的主要原因。

#### [参考文献]

- [1] 沈建军,李柏山,许海萍.太湖水污染原因分析及治理措施[J].环境科学导刊,2009,28(2):27-29.
- [2] 毕军,杨洁,李其亮.区域环境风险分析和管理[M].北京:中国环境科学出版社,2006:17-94.
- [3] 黄晓容.重庆三峡库区水环境污染事故预警指标体系研究[D].重庆:西南大学,2009:1-46.
- [4] 毕军,曲常胜,黄蕾.中国环境风险评价现状及发展趋势[J].环境监控与预警,2009,1(1):1-5.
- [5] 李万莲.蚌埠城市水生态环境预警研究[J].环境科学导刊,2008,27(5):43-46.
- [6] 郭松影,周直,高成卫.水安全风险评价系统研究[J].中国水运,2007,7(10):48-50.
- [7] 张根明,向晓骥,王殊伟.基于BP神经网络的中部六省可持续发展能力研究[J].国土与自然资源研究,2006(4):24-25.
- [8] 杨帆.突发性水污染事故预警指标筛选及体系构建研究[D].北京:北京林业大学,2009:39-41.
- [9] 张德丰.MATLAB神经网络应用设计[M].北京:机械工业出版社,2009:262.
- [10] 蔡嗣经,陈海燕,郑明贵.基于遗传神经网络的北京市能源可持续发展能力评价[J].辽宁工程技术大学学报:自然科学版,2009,28(1):5-8.
- [11] 刘树锋,陈俊合.基于神经网络理论的水资源承载力研究[J].资源科学,2007,29(1):99-104.
- [12] 李明,李雪铭.基于遗传算法改进的BP神经网络在我国主要城市人居环境质量评价中的应用[J].经济地理,2007,27(1):99-103.
- [13] 雷英杰,张善文,李续武,等.MATLAB遗传算法工具箱及应用[M].西安:西安电子科技大学出版社,2005:2.
- [14] HOUCK C, JONES J, KAY M. A genetic algorithm for function optimization; A Matlab Implementation [Z].
- [15] 杨秋林,张淑贞.基于BP神经网络的水环境承载力评价[J].国土与自然资源研究,2009(4):70-72.
- [16] 陈希,周娜娜.遗传神经网络在铁矿石需求预测中的应用[J].天津科技大学学报,2010,25(6):67-70.
- [17] 张石,张瑞友,汪定伟.基于PCA-BPNN方法的中长期电力负荷预测[J].控制工程,2010,17(6):800-802.
- [18] 陈柳,马广大.大气中SO<sub>2</sub>浓度的小波分析及神经网络预测[J].环境科学学报,2006,26(9):1553-1558.
- [19] 欧阳钧,王爱枝.基于Matlab的BP神经网络在大气污染物浓度预测中的应用[J].环境科学与管理,2009,34(11):176-180.

## 声 明

本刊已加入中国学术期刊网络出版总库、中国学术期刊综合评价数据库、万方数据—数字化期刊群、中国核心期刊(遴选)数据库、中文科技期刊数据库和教育阅读网。本刊已许可其以数字化方式复制、汇编、发行、信息网络传播本刊全文。著作权使用费与本刊稿酬一并支付。作者向本刊提交文章发表的行为即视为同意本刊上述声明。

《环境监控与预警》编辑部