

文章编号:1672-058X(2012)02-0047-05

# 改进遗传神经网络在水质评价中的应用\*

曾文华<sup>1</sup>, 崔 侠<sup>2</sup>, 刘 峰<sup>3</sup>

(1. 广州奥格智能科技有限公司, 广州 510663; 2. 广州市环境保护科学研究院, 广州 510620;  
3. 广州地理研究所 广州 510070)

**摘 要:**根据地表水环境质量评价标准,选取 5 项监测指标作为评价因子,在各指标对应级别取值区间内按均匀分布方式随机内插生成训练样本,采用改进自适应交叉和变异算子的遗传算法对 BP 网络连接权和阈值进行优化,以长江口某河段水质评价为例,阐明了该方法在水质评价中的作用,并与综合指标法进行比较;结果表明:该方法具有较好的客观性和实用性,能有效地对地表水环境质量进行评价,丰富了水质评价的方法体系,为地表水环境管理与决策的提供依据。

**关键词:**遗传算法;神经网络;水质评价

**中图分类号:**TP372

**文献标志码:**A

水质评价是水资源合理开发、利用和保护的重要依据,是在确立评价目标的情况下,选择相应的水质参数、水质标准和评价方法,对水质状况做出评定。传统的水质评价方法有很多,如综合指数法、模糊综合评价法、灰色聚类法等,但由于这些方法需要精确的数学模型来进行描述,须设计各评价指标对各级标准的隶属函数及各指标的权重,导致模型过于复杂造成计算困难,存在较大的主观随意性,很难真实模拟水环境系统的非线性变化过程<sup>[1-3]</sup>。理论和实践都证明,水质评价是一个典型的模式识别问题,运用不确定性、非线性理论可以克服传统水质评价方法在处理非线性问题中的实际困难。根据地表水环境质量评价标准,选取描述地表水环境质量的 5 项监测指标(TN、NH<sub>3</sub>-N、TP、CODMn 和总铅)作为评价因子,在各指标对应级别的取值区间内进行均匀分布方式随机内插得到训练样本集,利用遗传算法对 BP 神经网络连接权和阈值进行优化,以长江口某河段水质监测数据对网络进行测试,将测试结果与综合指标评价法所得结果进行比较分析。

## 1 遗传神经网络

### 1.1 人工神经网络

人工神经网络是一个非线性的动态系统,具有自组织、自学习能力和较强的非线性容错性等特点。BP (Back Propagation) 网络是 1986 年由 Rumelhart 和 McClelland 提出,是一种按误差逆传播算法训练的多层前馈网络,其结构如图 1 所示。

收稿日期:2010-12-20;修回日期:2011-06-23.

\* 基金项目:科技型中小企业技术创新基金资助(10C26214414756).

作者简介:曾文华(1978-),男,广东梅州人,从事遥感与地理信息系统、人工智能研究.

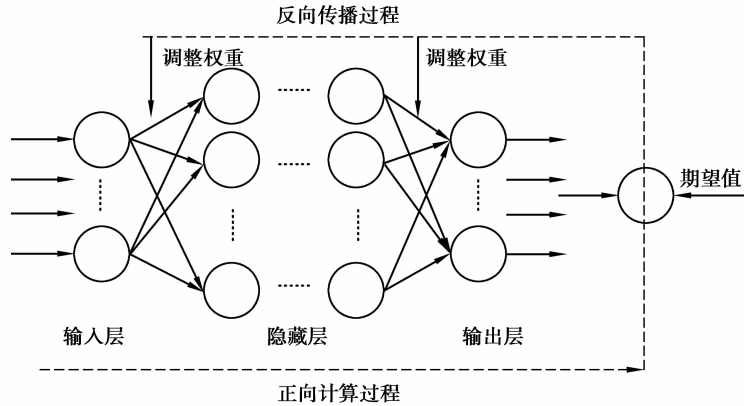


图 1 BP 网络结构

理论上 BP 神经网络可以任意精度逼近任何非线性连续函数,但在实际应用中,由于 BP 神经网络的连接权决定了网络的性能,其优化算法采用梯度下降法,该方法无法自动跳出局部搜索区间实现在全局范围内寻找最优值,极易陷入局部极值点而不是全局最优点,网络的精度因此而受到极大的制约。

## 1.2 改进遗传算法优化 BP 神经网络

遗传算法(GA, Genetic Algorithm)是一种模拟生命进化机制的搜索和优化方法,通过对对象系统第一代群体及其后代群体中的个体不断地优胜劣汰与随机性变异来获得对象系统的非线性映射。与传统搜索方法不同的是,GA 不是对具体参数搜索空间的一个解进行评估,而是对整个搜索空间的大量可行解同时并行搜索,这样就能克服反向传播算法可能陷入局部最优的不足<sup>[4,5]</sup>。但传统遗传算法存在一些具有超高适应度的个体,在选择算子时,会以很大的概率参加繁殖,引起群体平均适应度饱和、个体间竞争力减弱,导致运算收敛速度下降,甚至是未成熟收敛,出现“早熟”现象<sup>[6]</sup>。交叉概率  $p_c$  和变异概率  $p_m$  是决定遗传算法性能的关键,其取值直接影响算法的收敛性和收敛速度,实际应用中很难找到一个适应每个问题的最佳值。徐永春等通过改进自适应交叉和变异算子,设计自适应的交叉和变异公式<sup>[7]</sup>,如式(1)和式(2)所示。

$$p_c = \begin{cases} \frac{p_{c1}}{1 + \exp(10\lambda(t-1)/2(t_{\max}-2)) + k} & 0.5 < p_c < 1 \\ 0.5 & p_c < 0.5 \end{cases} \quad (1)$$

$$p_m = \frac{p_{m1}}{1 + \exp(-2t/t_{\max}) - 0.45} \quad (2)$$

式(1)、(2)中, $t$  为遗传迭代次数, $t_{\max}$  为最大遗传代数, $\lambda$ 、 $k$  为常数,根据种群分布大小取值,一般为 $[5, 10]$ , $p_{c1}$ 、 $p_{m1}$  分别为固定交叉概率和变异概率。通过式(1)、(2)可控制  $p_c$  取值在 $[0.5, 1]$ 之间,而  $p_m$  取值在 $[0, 0.5]$ 之间。通过自适应交叉和变异处理的遗传算法,出现适应值高于群体平均值的个体有更大的概率进入下一代,而适应值低于群体平均值的个体则将被淘汰,从而可大大加快遗传算法收敛。

## 2 改进遗传神经网络水环境评价模型

### 2.1 样本数据来源及预处理

样本数据采用《地表水环境质量标准(GB3838—2002)》所规定的水环境各类指标评价标准,选取 TN、NH<sub>3</sub>-N、TP、CODMn、总铅作为评价指标,分为 5 个级别,确定每一级别各类指标的标准浓度值范围,

如表 1 所示。

指标	I	II	III	IV	V
TN	[0,0.2]	(0.2,0.5]	(0.5,1.0]	(1.0,1.5]	(1.5,2.0]
NH <sub>3</sub> -N	[0,0.15]	(0.15,0.5]	(0.5,1.0]	(1.0,1.5]	(1.5,2.0]
TP	[0,0.02]	(0.02,0.1]	(0.1,0.2]	(0.2,0.3]	(0.3,0.4]
COD <sub>Mn</sub>	[0,2]	(2,4]	(4,6]	(6,10]	(10,15]
总铅	[0,0.005]	(0.005,0.01]	(0.01,0.025]	(0.025,0.05]	(0.05,0.1]

为使建立的 BP 神经网络能够学习到各种可能出现的情况,实现对检验数据的正确评价,需要对样本数据进行插值,在训练样本范围内按均匀分布方式随机内插生成训练输入样本,并对输入样本进行排序。同时,为了保证各指标数据具有可计算性,避免某些数值小的特征被淹没,需要对训练输入和输出数据进行归一化处理。在此采用最大最小值法进行归一化处理,该方法对数据的处理是一种线性变换,能够较好地保留其原始的意义,不会造成信息丢失。

## 2.2 BP 网络设计

采用 3 层 BP 神经网络,其中,输入层神经元个数  $n=5$ ,分别对应 TN、NH<sub>3</sub>-N、TP、CODMn 和总铅的观测值,输出层神经元个数为  $p=3$ ,对应评价结果。经过多次测试最终确定隐层神经元个数  $m=15$ ,隐层传输函数采用 sigmoid 型双正切函数,其导函数作为误差反向传输函数,输出层的神经元采用线性变换函数 purelin。遗传算法参数中,染色体采用实数形式编码,种群规模设置为 100,遗传代数设为 100,交叉概率和变异概率按式(1)和式(2)随迭代次数增加而收敛,误差平方和的倒数作为适应度。

## 2.3 网络训练

将改进遗传算法优化后的连接权及阈值代入设计好的 BP 神经网络,从插值后的分类标准数据中随机选取 80 个样本作为网络训练样本,其余 20 个样本作为测试数据,网络误差期望精度设置为  $\varepsilon=0.0001$ ,学习效率  $\eta=0.8$ ,对网络进行训练。

## 2.4 计算结果及分析

种群的进化轨迹曲线如图 3 所示,可以看出经过上世纪 80 代左右的进化,即可搜索到最优点,其进化速

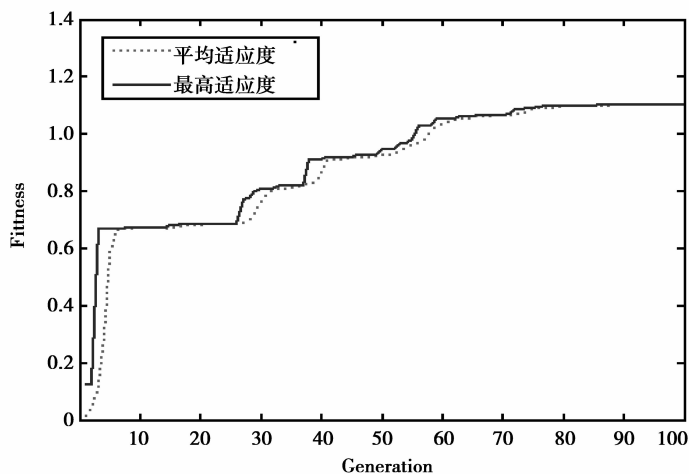


图 2 种群进化轨迹

度是比较快的。经改进遗传算法优化后的权值和阈值对 BP 网络进行初始化,随机抽取 100 条样本对网络进行训练,网络的收敛曲线如图 4 所示,经过约 4 次迭代网络即达到了预期的精度。

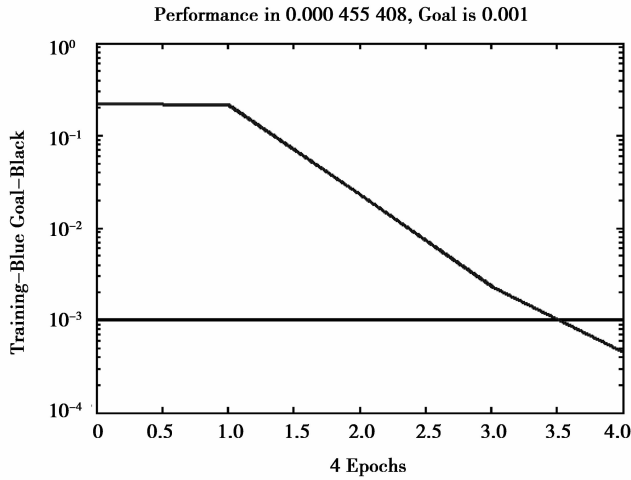


图 3 网络训练曲线

为了检验网络的分类能力,采用长江口某河段低平潮的观测数据<sup>[8]</sup>对网络进行测试。将训练样本归一化时采用的参数对测试数据进行预处理,处理后代入网络进行仿真,并将网络输出进行反归一化处理,水质监测数据、综合指数评价结果及模型输出结果如表 2 所示。结果表明:模型可有效地对水质监测数据进行分类,具有较高的分类能力和较强的通用性。

表 2 长江口某河段水质监测数据及评价结果

样点							mg/L	
	TN	NH <sub>3</sub> -N	TP	COD <sub>Mn</sub>	总铅	指数法	GA-ANN	
1	0.897	0.35	0.11	2.4	0.036	Ⅲ	100(Ⅲ)	
2	0.952	0.29	0.08	2.1	0.037	Ⅲ	100(Ⅲ)	
3	1.736	0.32	0.09	2.3	0.038	Ⅳ	110(Ⅳ)	
4	0.893	0.31	0.07	2.6	0.034	Ⅲ	110(Ⅳ)	
5	0.744	0.39	0.08	2.1	0.031	Ⅲ	100(Ⅲ)	
6	0.787	0.42	0.08	2.4	0.031	Ⅲ	100(Ⅲ)	
7	0.713	0.6	0.09	2.4	0.044	Ⅲ	110(Ⅳ)	
8	0.925	0.38	0.09	2.2	0.039	Ⅲ	100(Ⅲ)	
9	0.819	0.31	0.08	2.3	0.039	Ⅲ	100(Ⅲ)	
10	1.076	0.5	0.08	2.6	0.034	Ⅲ	100(Ⅲ)	
11	0.827	0.49	0.09	2.4	0.041	Ⅲ	110(Ⅳ)	
12	0.925	0.43	0.12	2.2	0.032	Ⅲ	100(Ⅲ)	
13	1.106	0.44	0.11	2	0.03	Ⅲ	100(Ⅲ)	
14	1.934	0.3	0.09	14.7	0.409	Ⅴ	111(Ⅴ)	

### 3 结束语

利用地表水环境评价标准,在各指标对应等级取值区间内进行均匀分布随机内插作为网络训练样本集,能为神经网络提供丰富的训练数据。计算结果与实测数据分析结果相比较后表明,遗传 BP 算法的学习

速度和计算精度均令人满意。通过遗传算子的改进,保证了优秀个体进入下一代,提高传统遗传算法的全局搜索和收敛效率,利用优化结果对 BP 神经网络进行初始化,可显著改善 BP 网络收敛速度慢、容易陷入局部最优解等不足。测试表明,经训练的网络具有较强的泛化能力,且具有较好的通用性,可作为地表水环境质量评价的一种实用方法。

#### 参考文献:

- [1] 董曼玲,黄胜伟. 径向基函数神经网络在水质评价中的应用[J]. 环境科学与技术,2003,26(1):23-25
- [2] QIAO J J,ZHEN X W,ZHANG Y R. The application of fuzzy comprehensive evaluation on the water quality of Changjiang River [C]. Proceedings of International Conference on Machine Learning and Cybernetics,2008:1467-1473
- [3] ZHU CH J,LIU Q H. Evaluation of water quality using grey clustering[C]. Proceedings of Second International Workshop on Knowledge Discovery and Data Mining,2009:803-805
- [4] 贺素良,王湘中,喻寿益. 遗传算法及模糊、神经网络融合技术的研究[J]. 计算机工程,2003,29(3):17-19
- [5] 黄建国,罗航等. 运用 GA-BP 神经网络研究时间序列的预测[J]. 电子科技大学学报,2009,38(5):687-692
- [6] 谢凯,排挤小生境遗传算法的研究与应用[D]. 合肥:安徽理工大学计算机科学与工程学院,2005
- [7] 徐永春,张森文. 改进遗传神经网络在甘蔗产量预测中的应用[J]. 华南农业大学学报,2010,31(3):102-104
- [8] 李伯昌,施慧燕. 长江口河段水环境现状分析[J]. 水资源保护,2003,21(1):39-44

## Application of the Neural Networks Using Improved Genetic Algorithm to the Surface Water Quality Evaluation

ZENG Wen-hua<sup>1</sup>, CUI Xia<sup>2</sup>, LIU Feng<sup>3</sup>

- (1. Augur Intelligence Technology (Guangzhou) Co., Ltd., Guangzhou 510663, China;
2. Guangzhou Research Institute of Environmental Protection, Guangzhou 510620, China;
3. Guangzhou Institute of Geography, Guangzhou 510070, China)

**Abstract:** Five indicators were selected as the evaluation factors according to Environmental Quality Standards for Surface Water, and the learning and training samples were created by uniform distributions random interpolating within the value of indicators. An improved genetic algorithm was used to optimize the initial weights and threshold of the BP neural network. The network was adopted to assess the water quality of one entrance of the Yangtze River and compared with comprehensive index method, the precision of the model was verified. The results showed that the genetic neural network algorithm was objective and practical and could effectively evaluate the quality of surface water environment and enrich water quality evaluation method and system, and could provide the basis for decision making for the surface water environmental management.

**Key words:** genetic algorithm; neural network; surface water quality evaluation

责任编辑:代小红