

# 基于 GA—BP 网络的人工湿地污水净化效果研究

黄娟, 王世和, 雒维国, 钱卫一, 鄢璐

(东南大学市政工程系, 南京 210096)

**[摘要]** 人工湿地作为污水处理领域的一种新型生态技术, 逐渐得到应用并倍受关注。人工湿地污水净化过程呈多参数影响的复杂非线性状态, 采用传统的数学方法难以建立准确的数学模型并进行多因素分析, 也无法精确确定出湿地系统最佳运行工况。基于大量可靠的试验数据, 首次采用遗传神经网络方法模拟湿地除污系统, 详细论述了网络拓扑结构优化和训练数据预处理等关键问题, 建立了可靠的 GA—BP 模型, 并采用该模型仿真湿地系统正交试验, 依据正交试验结果对影响因素进行分级, 确定了最佳运行工况, 并在此基础上有针对性地提出可行的强化除污措施。

**[关键词]** 人工湿地; 污水净化; GA—BP 网络; 正交试验

**[中图分类号]** X703.1 **[文献标识码]** A **[文章编号]** 1009—1742 (2007) 02—0079—05

## 1 前言

人工湿地是 20 世纪 70 年代蓬勃兴起的一种污水处理生态工程, 由于具有建造、运行和日常管理费用低廉, 处理效果稳定且适用面广等优点, 除处理城镇生活污水外, 也能广泛应用于农业、畜牧业、食品、矿山等工业废水的处理, 越来越受到世界各国的重视<sup>[1, 2]</sup>。人工湿地污水处理系统中, 有机物质的去除受到多因素的影响, 呈复杂的非线性生态过程, 难以建立准确的数学模型。目前大部分研究手段仍局限于单因素分析方法, 难以精确模拟污染物降解的复杂过程, 准确界定各因素的影响程度。近年来, 人工神经网络技术发展迅速, 广泛应用于各个领域, 也为上述问题提供了解决方法。鉴于此, 笔者首次提出采用遗传神经网络对人工湿地有机物降解系统进行非线性模拟, 合理确定网络拓扑结构、样本规模和样本数据归一化方法, 建立了稳定可靠的遗传神经网络模型, 并将其应用于仿真人工湿地有机物降解系统正交试验, 从而确定各

因素对有机物降解的影响程度和最佳运行工况, 并在此基础上提出可行的强化除污方法。

## 2 湿地除污系统遗传神经网络建模

### 2.1 遗传神经网络算法的基本原理

目前多层感知器网络 (BP)、径向基网络 (RBF)、Hopfield 网络和自适应共振网络 (ART) 属于应用较广泛的神经网络方法, 但这些方法都容易陷入神经网络结构的局部极小点, 而且其搜索空间只是整个结构空间中一个极小的子空间<sup>[3]</sup>。为此, 将遗传算法 (GA, genetic algorithm) 引入 BP 神经网络模型的学习过程, 形成 GA—BP 训练算法, 达到整体优化网络的目的。遗传算法利用简单的编码技术和繁殖机制来表现复杂的现象, 解决复杂的问题<sup>[4]</sup>。遗传 BP 学习算法的基本原理<sup>[5, 6]</sup>是利用遗传算法取代传统 BP 网络的权值学习算法 (“爬山”或“下山”法), 通过遗传和变异操作使初始网络权值不断更新换代, 使网络的系统总误差 SSE (sum square error) 趋于最小。算法流程如图 1 所示。

[收稿日期] 2005—12—30; 修回日期 2006—01—20;

[基金项目] 国家自然科学基金资助项目 (50278016)

[作者简介] 黄娟 (1980—), 女, 江苏南通市人, 东南大学博士研究生, 主要研究方向为水污染控制及水处理技术

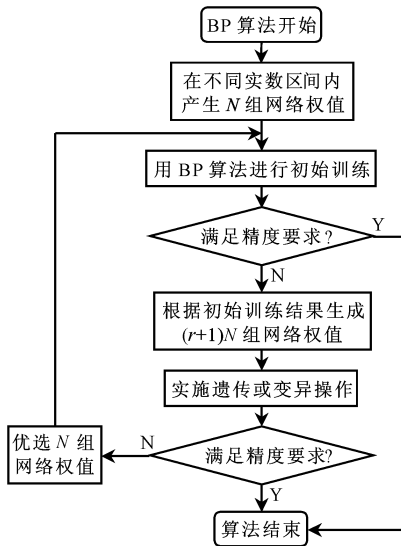


图 1 遗传 BP 算法流程框图

Fig.1 The flow chart of GA-BP arithmetic

遗传 BP 算法采用的适应度评价函数

$$F = [1 + \alpha[1 - N_y / N_{y_{\max}}]] / SSE$$

式中  $\alpha$  为隐含层神经元数量的影响系数,  $N_{y_{\max}}$  为隐含层神经元最大数量,  $N_y$  为对输出层节点有作用的隐含层节点的数量。遗传算子采用多位杂交法, 变异算子采用等概率突变操作<sup>[4]</sup>。

## 2.2 网络拓扑结构的确定

判定一个网络优劣的最主要标准: 网络是否具有好的泛化能力。网络的泛化能力是指经样本学习后的网络对学习样本外的数据做出正确反应的能力。影响网络泛化能力的 3 个最主要因素为网络拓扑结构、单元结点间的连接权值和训练样本的预处理。只要网络对训练样本的学习算法具有良好的全局收敛性好和稳定性, 则最终网络收敛的单元节点间连接权值也具有很好的精度。大量的研究和分析表明, 遗传神经网络算法正是一种稳定的全局收敛算法。因此在网络学习算法确定后, 最主要的工作就是要合理地确定网络拓扑结构和对训练样本进行预处理。

典型 BP 网络包括输入层、隐含层和输出层。Cybenko 从理论上证明了三层前向神经网络可以逼近任一连续函数<sup>[3]</sup>。大量试验数据表明湿地有机物去除率与各影响因素之间的关系是接近连续函数的。在人工湿地污水处理系统中, COD 是评价有机物去除效果的主要指标, 因此确定网络输出层为 1 个单元即 COD 去除率。根据大量的试验分析选择

7 个影响湿地 COD 去除率的因素作为输入参数, 如图 2 所示。隐含层的确定包括两个方面: 隐含层的层数和单元个数。为逼近一个连续函数, 只有一个隐含层的网络便已足够, 增加隐含层的层数反而会使网络的结构变得复杂, 稳定性变差<sup>[3]</sup>。隐含层的单元数目可采用 VC 维 (VC dimension) 来确定, VC 维可以测试系统复杂度对学习能力的影 响<sup>[7]</sup>。对于一个 VC 维为  $d$  的学习系统, 如果成功地用  $m$  个样本对其进行了训练, 则在最坏情况下, 其泛化程度也以很高的置信度小于给定容许误差  $e$ ,  $e \leq O((d/m) \ln(m/d))$ 。单隐层前馈神经网络的 VC 维  $d = MN + NP$ , 其中  $M$ ,  $N$ ,  $P$  分别为输入层、隐层、输出层的神经元数。 $M$  和  $P$  已知, 在样本数目允许范围内改变  $m$ , 并使  $N$  取不同数值, 可获得不同的  $e$ 。选择使  $e$  尽可能小的  $M$  和  $N$  的组合, 便可在确定最优网络拓扑结构的同时, 得到最适合该网络的训练样本规模  $m$ 。根据计算, 当  $m$  为 64,  $N$  为 8 时,  $e$  最小。由此网络拓扑结构得到确定, 如图 2 所示, 其中  $x_1$  为进水化学耗氧量 (COD),  $x_2$  为进水氨氮浓度,  $x_3$  为进水 pH 值,  $x_4$  为进水总磷浓度 TP,  $x_5$  为停留时间,  $x_6$  为水位,  $x_7$  为温度, 导师值  $y$  为出水 COD 的去除率, 并可编制与网络拓扑结构相应的遗传 BP 算法程序。

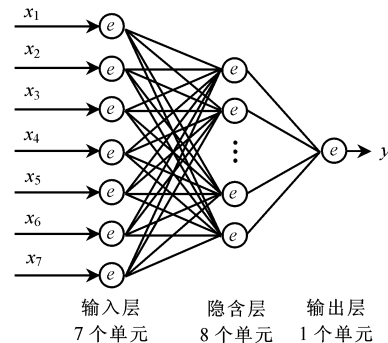


图 2 神经网络拓扑结构

Fig.2 Topological structure of ANN model

## 2.3 训练数据的获得与预处理

以南京锁金村污水厂初沉池出水为处理对象, 建立人工湿地中试试验系统, 经过长期观测, 获得大量试验数据, 其中剔除一些明显错误的数 据, 并依据尽量扩大输入参数取值范围的原则, 从中选择 64 组试验数据作为神经网络的训练样本, 5 组作为经训练的网络预测性能的检验样本, 限于篇幅, 表 1 给出 10 组样本数据, 其中前 5 组为检验样本。

表 1 GA—BP 网络样本原始数据列表  
Table 1 Origin data list of ANN sample

序号	$x_1/\text{mg}\cdot\text{L}^{-1}$	$x_2/\text{mg}\cdot\text{L}^{-1}$	$x_3/\text{pH}$	$x_4/\text{mg}\cdot\text{L}^{-1}$	$x_5/\text{d}$	$x_6/\text{cm}$	$x_7/^\circ\text{C}$	$y/\%$
1	235.3	46.5	7.2	4.2	5.5	20	25	68.2
2	195.1	56.2	7.1	4.2	5.5	40	29	78.8
3	213.5	41.2	7.1	4.2	5.5	20	24	75.6
4	205.3	20.1	7.8	4.5	5.0	20	25	70.6
5	261.3	36.2	7.1	4.1	5.5	60	30	88.9
6	193.4	56.2	8.0	5.0	5.5	60	35	94.5
7	186.2	56.6	7.2	6.8	5.5	40	37	83.9
8	181.3	50.2	7.8	5.2	3.0	60	29	54.7
9	210.2	56.5	7.8	3.8	5.5	60	34	92.4
10	185.6	37.2	7.6	5.3	6.5	60	25	88.5

为使网络具有良好的泛化性能，需要对训练数据进行归一化预处理。由于归一化方法的选择对网络性能影响较大，因此采用线性归一、对数归一和概率归一方法对训练数据进行预处理，原理如下：

线性归一  $((x_i - x_{\min}) / (x_{\max} - x_{\min})) (x_H - x_L) + x_L, i = 1, 2, \dots, M$

对数归一  $((\ln x_i - \ln x_{\min}) / (\ln x_{\max} - \ln x_{\min})) \cdot (x_H - x_L) + x_L,$

概率归一  $(x_i - E) / \sigma。$

式中  $i = 1, 2, \dots, M, x_{\max}$  和  $x_{\min}$  为样本最大值和最小值， $x_H$  和  $x_L$  为归一后区间上界和下界， $E$  和  $\sigma$  分别为样本均值和均方差。根据不同归一化方法的预测检验结果，选择精度最高的一种作为最终确定的归一化方法。

对于人工湿地高度非线性系统，以原始试验数据为基础，采用 Matlab 软件建立的多元线性回归模型为

$$y = 245.9019 - 0.0127x_1 - 0.0903x_2 - 7.6697x_3 + 0.6830x_4 - 16.9577x_5 - 0.7438x_6 - 0.8377x_7。$$

由表 2 和图 3 可看出，线性回归模型预测结果不仅误差较大（最大达 70% 以上），并且有时其预测值会存在严重的失真现象（如第 2、第 5 组数据所示），而遗传 BP 网络的预测结果则与实际情况符合较好。通过对网络训练样本原始数据不同归一方法的比较，选择预测精度最高的对数归一方法作为原始训练样本的预处理方法。

表 2 网络模型与线性回归预测结果比较

Table 2 Comparison of prediction values with ANN and linear regression

序号	GA—BP 网络预测 $y/\%$			多元线性回归预测 $y/\%$	实测 $y/\%$
	对数归一	线性归一	概率归一		
1	62.9	63.2	74.5	57.3	68.2
2	79.1	73.5	76.5	39.5	78.8
3	77.5	79.8	77.9	59.6	75.6
4	69.8	78.2	61.5	64.1	70.6
5	86.2	90.2	85.6	24.6	88.9

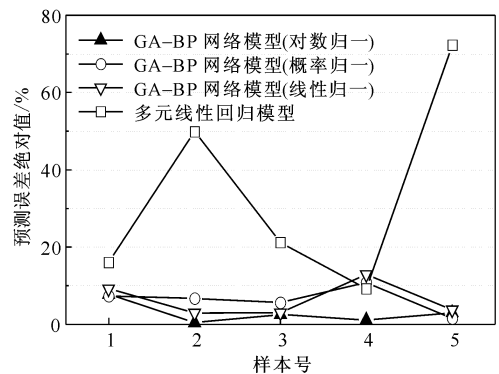


图 3 误差检验图

Fig.3 Comparative scheme of prediction values

### 3 基于系统遗传 BP 网络的正交试验

采用上节建立的遗传 BP 网络模型仿真正交试验。正交试验仿真结果如表 3 所示。由此可确定各因素对人工湿地 COD 去除率的影响程度由大到小顺序：停留时间，水位，进水 COD 含量，温度，进水氨氮含量，进水磷含量，进水 pH；并划分为

3 个等级：影响显著（水流停留时间、水位）、影响较大（进水 COD 含量、温度）、影响较小（进水氨氮含量、进水磷含量、进水 pH）。首先，控制水力停留时间为 6 d 时 COD 去除率

表 3  $L_{18}(2 \times 3^7)$  的正交试验数据表\*

Table 3 The orthogonal test table of  $L_{18}(2 \times 3^7)$

实验号	空列	$x_1/\text{mg} \cdot \text{L}^{-1}$	$x_2/\text{mg} \cdot \text{L}^{-1}$	$x_3/\text{pH}$	$x_4/\text{mg} \cdot \text{L}^{-1}$	$x_5/\text{d}$	$x_6/\text{cm}$	$x_7/^\circ\text{C}$	$y/\%$
1	(1)	(1) 150	(1) 20	(1) 6.5	(1) 1	(1) 3.0	(1) 20	(1) 30	55.16
2	(1)	(1) 150	(2) 30	(2) 7.0	(2) 3	(2) 5.5	(2) 40	(2) 20	78.47
3	(1)	(1) 150	(3) 40	(3) 7.5	(3) 5	(3) 6.0	(3) 60	(3) 15	86.19
4	(1)	(2) 200	(1) 20	(1) 6.5	(2) 3	(2) 5.5	(3) 60	(3) 15	76.38
5	(1)	(2) 200	(2) 30	(2) 7.0	(3) 5	(3) 6.0	(1) 20	(1) 30	71.19
6	(1)	(2) 200	(3) 40	(3) 7.5	(1) 1	(1) 3.0	(2) 40	(2) 20	59.90
7	(1)	(3) 250	(1) 20	(2) 7.0	(1) 1	(3) 6.0	(2) 40	(3) 15	86.93
8	(1)	(3) 250	(2) 30	(3) 7.5	(2) 3	(1) 3.0	(3) 60	(1) 30	82.40
9	(1)	(3) 250	(3) 40	(1) 6.5	(3) 5	(2) 5.5	(1) 20	(2) 20	72.21
10	(2)	(1) 150	(1) 20	(3) 7.5	(3) 5	(2) 5.5	(2) 40	(1) 30	71.81
11	(2)	(1) 150	(2) 30	(1) 6.5	(1) 1	(3) 6.0	(3) 60	(2) 20	90.77
12	(2)	(1) 150	(3) 40	(2) 7.0	(2) 3	(1) 3.0	(1) 20	(3) 15	35.58
13	(2)	(2) 200	(1) 20	(2) 7.0	(3) 5	(1) 3.0	(3) 60	(2) 20	72.18
14	(2)	(2) 200	(2) 30	(3) 7.5	(1) 1	(2) 5.5	(1) 20	(3) 15	68.39
15	(2)	(2) 200	(3) 40	(1) 6.5	(2) 3	(3) 6.0	(2) 40	(1) 30	86.96
16	(2)	(3) 250	(1) 20	(3) 7.5	(2) 3	(3) 6.0	(1) 20	(2) 20	79.62
17	(2)	(3) 250	(2) 30	(1) 6.5	(3) 5	(1) 3.0	(2) 40	(3) 15	66.06
18	(2)	(3) 250	(3) 40	(2) 7.0	(1) 1	(2) 5.5	(3) 60	(1) 30	95.31
$I_j$		417.98	442.08	447.54	456.45	371.29	382.15	462.83	
$II_j$		434.98	457.28	439.66	439.41	462.57	450.13	453.16	$\alpha=1, \dots, 18$
$III_j$		482.54	436.15	448.31	439.64	501.65	503.22	419.52	$j=1, \dots, 8$
$R$		64.56	21.13	8.65	17.04	130.36	121.07	43.31	

\* I, II, III 为正交试验中间分析结果, R 为正交试验最终分析结果,  $\alpha$  为试验组号 (18 组),  $j$  为自变量 (8 个), (1), (2), (3) 表示各参数的水平标号, 均为正交表固定格式

明显高于 3 d, 这是因为停留时间过短, 水流湍动剧烈, 吸附于基质表面的颗粒性有机物来不及降解而被带出系统, 有机物降解率降低。

其次, 水位越高, 有机物降解效果越好。这是由于湿地上层覆盖土壤基质, 而植物根系主要位于湿地表层以下 20~40 cm 处, 因此当水位淹没土壤基质和植物根系时, 强化了基质、植物、微生物三方面的协同作用: 其一, 土壤基质和植物根系表面好氧微生物大量繁殖, 有利于好氧反应降解有机物; 其二, 充分发挥了淹没于湿地水位以下的植物根系对营养物质的吸收作用; 其三, 植物根系在土壤基质中交错纵生, 有利于对颗粒性有机物拦截吸附, 进而被微生物降解。

第三, 人工湿地对有机物降解效果较为稳定, 抗冲击负荷能力较强, 出水 COD 普遍较低, 因此进水 COD 越高, 相对去除率越低。而湿地系统中有机物主要依靠好氧微生物降解去除, 温度越高, 好氧微生物生长状态越佳, 有利于有机物的降解。

此外, 进水氨氮主要依靠微生物硝化反硝化作用去除, 而湿地系统中溶氧浓度偏低, 硝化菌为弱势菌种, 对有机物降解影响较小; 湿地磷的去除主要依靠物理化学作用, 对 COD 去除率几乎无影响; 而进水基本处于中性 pH 值条件, 各水平相差不大, 对 COD 去除率影响较小。

由表 3 还可看出, 进水 COD 为 250 mg/L, 氨氮为 40 mg/L, pH 为 7.0, TP 为 1 mg/L 时, 对应最大

COD 去除率 (95.31 %) 的最佳运行参数组合为: 停留时间 5.5 d, 水位 60 cm 及温度 30℃。各参数在此组合附近时, COD 去除率达到最大。

#### 4 结语

以人工湿地污水处理系统为模拟目标, 针对 GA—BP 网络拓扑结构、训练样本规模、训练数据归一化方法等方面进行详细的讨论, 建立了优化 GA—BP 网络模型: 隐层单元为 8, 训练样本规模为 64, 训练数据采用对数归一方法。检验结果表明, 优化后的 GA—BP 网络模拟输出与实际试验结果相差在 7.8 % 左右, 满足精度要求。在此基础上, 利用优化后的 GA—BP 网络模型仿真正交试验分析, 根据分析结果将湿地有机物去除影响因素划分为三个等级: 影响显著 (水流停留时间、水位)、影响较大 (进水 COD 含量、温度)、影响相对较小 (进水氨氮含量、进水磷含量、进水 pH), 并确定了人工湿地有机物去除的最佳运行参数组合为停留时间 5.5 d, 水位 60 cm 及温度 30℃。

为了增强湿地对污染物的净化效果, 提出以下建议: **a.** 适当延长水力停留时间 (5.5~6.5 d 最佳), 或采用串联工艺有利于有机物的吸附降解; **b.** 湿地低温条件下污染物净化效果不理想, 可采取一定保温措施, 尽量减少低温域系统热损失, 缓解低温的不利影响; **c.** 尽量采取高水位 (60~70

cm 水深最佳) 运行潜流湿地, 使污水淹没植物根区, 并选择纵深方向根系生长旺盛的水生植物作为首选湿地植物。

#### 参考文献

- [1] Michael C K. Subsurface flow constructed wetlands treating municipal wastewater for nitrogen transformation and removal [J]. *Water Environment Research*, 1997, 67 (7): 1254~1264
- [2] Gopal B. Natural and constructed wetlands for wastewater treatment: potentials and problems [J]. *Water Science and Technology*, 1999, 40: 27~35
- [3] 张乃尧, 阎平凡. 神经网络与模糊控制[M]. 北京: 清华大学出版社, 1998
- [4] Michalewicz Z. Genetic Algorithm + Data Structure = Evolution Programs, 2nd ed [M]. New York: Springer-Verlag, 1994
- [5] Whitley D, Dominic S, Das R. Genetic reinforcement learning for neurocontrol problem [J]. *Machine Learning*, 1993, 13: 259~284
- [6] Koehn P. Combining Genetic Algorithms and Neural Networks: The Encoding Problem [D]. Knoxville: The University of Tennessee, 1994
- [7] Ehrenfeucht A, Detal H. A general lower bound on the number of example needed for learning[J]. *Information and Computation*, 1989, 82(3):247~251

## Study on the Purification of Wastewater in the Constructed Wetland Based on GA—BP Network

Huang Juan, Wang Shihe, Luo Weiguo, Qian Weiyi, Yan Lu

(Department of Municipal Engineering, Southeast University, Nanjing 210096, China)

**[Abstract]** As a new type of ecological technique in wastewater treatment field, Constructed Wetland was gradually put into application and received growing attention. Since the wastewater purification process in constructed wetland was a complex and nonlinear state affected by many interactional factors, it was hard to establish exact mathematics model to carry out multi-factor analysis and determine best operation condition by traditional means. Based on plenty of reliable experimental data, genetic neural network was first tentatively utilized to simulate the pollutant removal system of wetland, and some key problems were discussed, such as how to determine optimally the topological structure, sample scale and unitary method for training data, etc. Optimized GA—BP network was established to simulate orthogonal test of wetland system. According to the results of orthogonal test, best operation condition was decided and the factors (e.g. water level, hydraulic retaining time, etc) were classified. Therefore, proper feasible methods of pollutant removal were put forward.

**[Key words]** constructed wetlands; wastewater purification; GA—BP network; orthogonal test