

基于径向基函数神经网络的污水生物处理模拟

Simulation of Biological Wastewater Treatment Based on Radial Basis Function Neural Network

刘俊萍 严敏 (浙江工业大学建筑工程学院 杭州 310014)
 胡坚 (镇江水工业公司排水管理处 镇江 212003)
 王亚宜 (浙江工业大学建筑工程学院 杭州 310014)

摘要:针对复杂的非线性污水生物处理过程,开发了径向基函数的人工神经网络模型,并用实测污水厂进、出水数据进行模拟。采用最近邻聚类学习算法确定径向基函数的宽度、聚类中心和权值。其中神经网络的输入为进水水质和控制参数等 5 个影响因子,网络输出为 COD 或 TN。结果表明,用径向基函数神经网络模拟污水生物处理拟合性好,出水 COD 和 TN 达到预期的模拟精度。

关键词:污水生物处理 径向基函数 神经网络 最近邻聚类算法

循环活性污泥法(Cyclic Activated Sludge System, CASS)是在活性污泥法的基础上将 SBR 和生物选择器结合起来的一种新工艺,它具有抗冲击负荷、防止污泥膨胀、可脱氮,运行控制灵活等特点,也易于实现自动控制。但是在这种生物处理的运行过程中会受到很多因素的影响,这些因素之间又呈现出非常复杂的非线性关系。要保证生物处理过程的正常进行,并实现自动控制,必须研究并掌握这些因素之间的规律。目前国际上虽然进行了大量活性污泥数学模型的研究工作,但是由于模型中用的大部分参数随水质和所选处理工艺的不同而变化,要确定这些参数非常复杂,而且有些参数至今还未能实现实时检测,因此模型的建立和应用都有一定的困难^[1]。

人工神经网络(Artificial Neural Networks, ANN)是一种数学模型,尤其适用于一些非线性因素组合的情况。它可以将一些难以实时检测的参数与污水处理的一些易检测过程参数(如 DO 等)联系起来,建立其可靠的相关关系,从而间接实现对污水水质参数在线实时检测的“软测量”技术。

本文概述径向基函数模型的开发,模拟了非线性污水生物处理过程,考察在多个影响因素输入情况下所能取得的效果,并测试了 CASS 工艺中采用神经网络的可行性。

1 径向基函数神经网络模型

评价人工神经网络学习算法优劣的指标有学习所需的时间、泛化能力、神经网络的结构复杂性、鲁棒性。在多种人工神经网络中,目前采用最普遍的是以多层传感器为基础的误差反向传播前馈模型,称为 BP 网络^[2],即将输入信息沿网络正向传播,将误差信号沿网络反向传播,并修正权值。BP 网络的初始权值是随机给定的,所以每次训练的结果也不同,这是由于误差曲面的局部极小点非常多造成的。BP 算法本质上是梯度算法,易陷入局部极小点,而且收敛速度慢,所设计的神经网络的泛化能力不能保证。本研究中采用的是径向基函数神经网络 RBF(Radial Basis Function Neural Network)模型,RBF 神经网络模型,无论在逼近能力、分类能力和训练速度等方面均优于 BP 网络,能够避免局部收敛,具有快速全局收敛的性质^[3-4]。

1.1 径向基函数神经网络结构

径向基函数(RBF)神经网络是 J Moody 和 C Darken 于 80 年代末提出的一种以函数逼近理论为基础的三层前馈网络,具有很强的生物学背景,它模拟人的大脑皮层区域中局部调节及交叠的感受域(Receptive Field)的特点,常用的径向基函数有以下几种形式^[5]:

$$f(x) = \exp^{-\alpha \|x\|^2} \quad (1)$$

$$f(x) = 1/(\sigma^2 + x^2)^\alpha, \alpha > 0 \quad (2)$$

$$f(x) = (\alpha^2 + x^2)^\beta, \alpha < \beta < 1 \quad (3)$$

以上这些函数都是径向对称的,但最常用的是高斯函数:

$$R_i(x) = \exp(-\|x - c_i\|^2 / 2\sigma_i^2), i=1,2,\dots,m \quad (4)$$

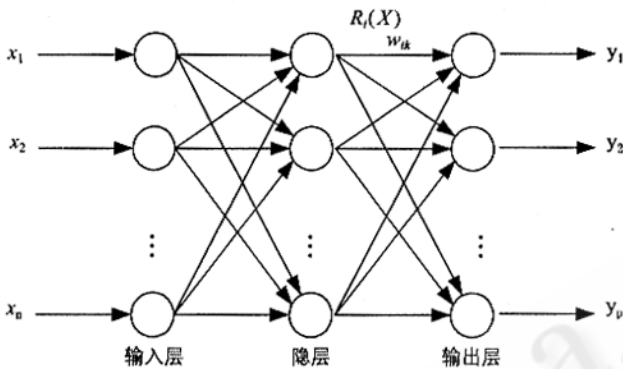


图 1 RBF 神经网络结构

式中: x 是 n 维输入向量; c_i 是第 i 个基函数的中心,与 x 具有相同维数的向量, σ_i 是第 i 个感知的变量,它决定了该基函数围绕中心点的宽度; m 是感知单元的个数(隐含层节点数)。 $\|x - c_i\|$ 是向量 $x - c_i$ 的范数,它通常表示 x 与 c_i 之间的距离; $R_i(x)$ 在 c_i 处有一个唯一的最大值,随着 $\|x - c_i\|$ 的增大, $R_i(x)$ 迅速衰减到零。对于给定的输入,只有一小部分靠近 x 的中心被激活。

RBF 神经网络的隐层作用基函数选用高斯函数,设输入层的输入为 $X = (x_1, x_2, \dots, x_1, \dots, x_n)$, 实际输出为 $Y = (y_1, y_2, \dots, y_k, \dots, y_p)$ 。输入层实现从 $X \rightarrow R_i(X)$ 的非线性映射,输出层实现从 $R_i(X) \rightarrow y_k$ 的线性映射,输出层第 k 个神经网络输出为:

$$\hat{y}_k = \sum_{i=1}^m \omega_{ik} R_i(X), k=1,2,\dots,p \quad (5)$$

式中: n 为输入层节点数; m 为隐层节点数; p 为输出层节点数; ω_{ik} 为隐层第 i 个神经元与输出层第 k 个神经元的连接权值; $R_i(X)$ 为隐层第 i 个神经元的作用函数。

因此,当确定了 RBF 网络的聚类中心 c_i 、权值 ω_{ik} 以后,就可求出给定某一输入时,网络对应的输出值。基于上述理论,本文基函数选用高斯函数,并构建一个 RBF 神经网络结构图。

图 1 所示为 RBF 神经网络结构,它由一个输入层、一个隐含层和一个输出层组成。输入层包括 5 个输入

变量,即化学耗氧量 COD、总氮 TN、温度 T、溶解氧 DO 和进水负荷 Q_{in} ,输出层则含 1 个输出变量,即 COD 或 TN。

1.2 RBF 网络的学习方法

RBF 网络常用的学习算法有聚类法、梯度训练法、正交最小二乘学习方法等^[6],通过这些算法来确定隐层的节点数。本文采用最近邻聚类学习算法来确定隐层单元的节点数和聚类中心,该算法是一种在线自适应聚类学习算法,隐层单元的节点数不需要事先人为确定,可实现动态在线学习。

2 RBF 污水生物处理模型应用实例

本研究将开发的径向基函数神经网络模型模拟了镇江市征润洲污水处理厂的出水水质。征润洲污水处理厂采用的是 CASS 工艺,它集生物选择区和反应区为一体,每个运行周期是 3 个小时,依次完成进水、曝气、沉淀、滌水等过程。经过 40 个周期的运行,实测得到的进、出水数据被用来进行对网络的训练和模拟。在生物处理过程中,影响出水水质的因素有很多,但是在 CASS 运行中,主要通过对溶解氧(DO)的控制,使废水经过一个厌氧-缺氧-好氧-缺氧-厌氧的交替过程,实现较强的脱氮作用。因此前述 5 个输入变量中,溶解氧 DO 是很重要的影响因素,被认为是模型中重要的影响因子,模型采用的其余影响因子还有平均进水负荷、温度等因素。

2.1 出水 COD 模拟

出水 COD 模拟共采用了 40 个样本资料,前 35 个样本用于训练网络,后 5 个样本用于出水水质模拟。RBF 网络的输入变量为 5 个影响因子,输出为 COD。因此程序中训练样本的个数为 $N=35$,输入层节点数为 $n=5$,输出层节点数为 $p=1$ 。在训练样本之前,先将样本的输入和输出进行了归一化处理。

图 2 所示为 35 个样本出水 COD 的拟合结果,图中 COD 网络输出为人工神经网络的模拟结果,而 COD 实测为水厂实测数据,从图 2 可以看出,模型输出和实测数据极相符合,模型的拟合误差非常小,拟合精度很高。

将训练好的网络用于模拟其余 5 个样本的水质情况,5 个被模拟样本的进水水质和控制参数见表 1。

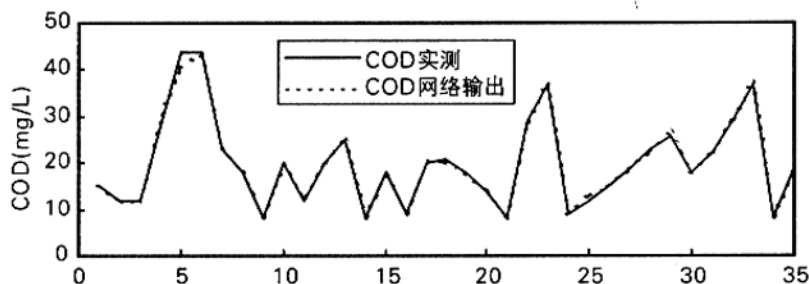


图 2 RBF 网络 COD 拟合结果

表 1 RBF 模拟输入节点

样本	进水水质			控制参数	
	COD /mg/L	TN /mg/L	DO /mg/L	Qin/ m3/min	T /°C
1	124	12.712	1.08	56.7	13.2
2	364	25.826	1.51	43.2	13.2
3	154	20.771	2.87	48.3	13.5
4	147	21.321	1.7	54.6	13.6
5	198	20.508	1.8	48.3	13.4

模拟得到的 5 个样本的出水 COD 值见图 3 和表 2。

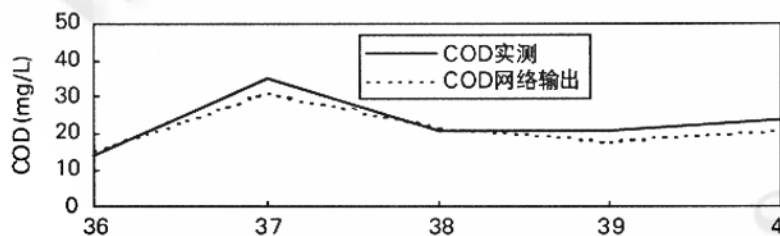


图 3 实测和 RBF 网络模拟的出水 COD

表 2 RBF 神经网络 COD 模拟误差

样本	实测 COD /mg/L	模拟 COD /mg/L	绝对误差 /mg/L	相对误差 /%
1	14	14.9	0.92	6.58
2	35	31.1	-3.88	-11.07
3	21	21.7	0.72	3.43
4	21	17.9	-3.14	-14.95
5	24	21.0	-3.00	-12.49

5 个样本的绝对误差分别为 0.92mg/L、-3.88mg/L、-0.72mg/L、-3.14mg/L、-3.00mg/L，相对误差分别为 6.58%、-11.07%、3.43%、-14.95% 和 -12.49%，平均的相对误差绝对值为 9.70%，具体见表 2。

由于进水 COD 的波动幅度很大，从表 1 中可以看出进水 COD 的差值为 240 mg/L，训练用的网络用 35 个样本的 COD 的差值更大，即使在这种情况下，取得的结果也是令人满意的。

2.2 出水 TN 模拟

同样采用以上 40 个样本资料，用前 35 个样本训练网络，后 5 个样本用于出水水质模拟。RBF 污水水质模拟模型的输入相同，即 5 个影响因子，输出为 TN。

训练样本的模拟结果见图 4 所示。从图中可知，污水水质模拟模型的训练段拟合误差非常小，拟合精度很高。

将表 1 中的进水水质和控制参数输入 RBF 水质模拟模型，得到 5 个样本的模拟出水 TN 指标，模拟结果见表 3 和图 4。

5 个样本的绝对误差分别为 1.545mg/L、-0.620mg/L、0.927mg/L、0.516mg/L、0.281mg/L，相对误差分别为 9.77%、-3.13%、5.19%、2.83% 和 1.53%，平均的相对误差绝对值为 4.49%。图 4 中模拟段为这 5 个样本出水水质 TN 指标的 RBF 神经网络模拟效果，由此可见，RBF 神经网络的模拟精度较高，用于 CASS 污水生物处理的模拟效果很好。

3 结语

本文所建立的径向基函数神经网络污水生物处理模型，具有较强非线性处理能力和逼近能力。文中采用最近邻聚类学习算法选取聚类中心，使该算法具有学习时间短，网络运算速度快，性能稳定等优点。经过采用 40 个污水厂实际进出水水质

表 3 RBF 神经网络模拟误差

样本	实测 TN /mg/L	模拟 TN /mg/L	绝对误差 /mg/L	相对误差 /%
1	15.816	17.361	1.545	9.77
2	19.770	19.150	-0.620	-3.13
3	17.868	18.795	0.927	5.19
4	18.268	18.784	0.516	2.83
5	18.337	18.618	0.281	1.53

(下转第 58 页)

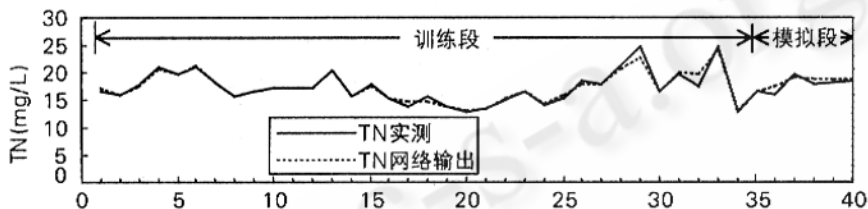


图 4 RBF 网络 TN 实测及模拟

样本, 选用了 5 个影响因子, 对出水水质中 COD 指标和 TN 指标分别进行训练和模拟, 结果表明, 该污水生物处理模型有相当高的模拟精度, 即使在进水 COD 波动幅度较大的情况下, 也有较高的预测效果, 因此该模型有其实用价值, 可帮助采用生物处理工艺的污水处理厂运行时预测出水水质, 有条件在事先为可能出现的水质波动采取预防措施。

参考文献

1 刘建勇、薛罡、顾国维, 神经网络在污水处理自动控制

制中所起的作用[J], 给水排水, 2003, 29(7): 85 - 88.

2 Maged M. Hamed, Moga G. Khalafallah, Ezzat A. Hassanien. Prediction of Wastewater Treatment Plant Performance Using Artificial Neural Networks[J]. Environmental Modelling & software, 2004, 19: 919 - 928.

3 陆爽、张子达、李萌, [J]. 中国工程科学, 2004, 6(2): 56 - 60.

4 孙健、申瑞民、韩鹏, 一种新颖的径向基函数 RBF 网络学习算法[J]. 计算机学报, 2003, 26(11): 1562 - 1567.

5 焦李成, 神经网络系统理论[M], 西安: 西安电子科技大学出版社, 1996: 1 - 15.

6 魏海坤, 神经网络结构设计的理论与方法[M], 北京: 国防工业出版社, 2005: 31 - 45.