

基于神经网络的活性污泥系统建模及其控制

楼文高^{1,2} 刘遂庆¹

(1. 同济大学环境科学与工程学院, 上海 200092; 2. 上海理工大学, 上海 200093)

摘要 针对松江污水厂污水处理活性污泥系统, 采用神经网络技术进行建模试验研究, 在对实际运行数据剔除异常数据后, 将样本数据随机分成训练样本、检验样本和测试样本。用试凑法确定合理的神经网络隐层节点数, 用检验样本实时监控训练过程从而避免“过训练”现象, 用多次改变网络初始连接权值求得全局极小点, 从而建立了泛化能力较好的基于神经网络的活性污泥系统数学模型。利用建立的神经网络模型, 对活性污泥系统运行情况的仿真与控制进行了分析研究。示例研究表明: 神经网络技术能较好地应用于活性污泥系统的建模与控制, 有很好的理论与实践意义。

关键词 活性污泥系统 神经网络 建模 仿真 控制 泛化能力

中图分类号 TP273 **文献标识码** A **文章编号** 1008-9241(2006)08-0047-05

Application of neural network to modeling and simulation on activated sludge system

Lou Wengao^{1,2} Liu Suiqing¹

(1. School of Environmental Science and Engineering, Tongji University, Shanghai 200092;
2. University of Shanghai for Science and Technology, Shanghai 200093)

Abstract The actual operation data of activated sludge system in Shanghai Songjiang Sewage Treatment Plant (SSSTP) were used to establish a neural network-based (NN-based) model of activated sludge system. The abnormal data were deleted according to the physical and scope principle and an effective data set was filed. The total data were divided into three parts, namely, training data set, verification (validation) data set and test data set. The proper number of neurons on hidden layer was determined by trail-and-error. The verification data set was used to monitor the training process real-timely and dynamically and to find out over-training phenomena. The global minimum of error-function was got by randomly changing the initial value of connection weights more than thirty times. The reasonable, reliable NN-based model for activated sludge system was thus set up in this paper. The simulation and operation control of activated sludge system was carried out by using the established NN-based model. The case study shows that neural network can be successfully applied to describe and control the non-linear and complicated activated sludge system. The NN-based model possesses good generalization.

Key words activated sludge system; neural network; modeling; simulation; control; generalization

现有的基于基质降解和微生物生长规律和反应器理论的数学模型(如莫诺方程式、劳伦斯-麦卡蒂方程式等), 曾对污水处理技术发展起到了积极的推动作用^[1], 但污水处理系统是一个复杂的非线性系统, 上述模型也存在一些难以解决的问题。

另一方面, 神经网络技术具有好的自适应性、自学习性和容错性的非线性建模技术, 已在很多复杂系统的建模中取得了成功, 得到了广泛的应用^[2-4]。采用神经网络技术建模首先要解决2个问题:(1) 确定合理的神经网络结构;(2) 训练过程中避免出现“过训练”现象并求得全局极小点邻域内的可行解。自20世纪90年代以来(国外自1990年, 国内自1998年)以来^[5,6], 在污(废)水处理领域已先后

发表了几十篇有关应用神经网络技术的论文(详见文献^[7]), 取得了一些探索性的研究成果。但由于多数作者对神经网络技术缺乏较深入的理解, 在建模过程中采用了不太合理(过大)的网络结构, 无法判定训练过程中是否发生了“过训练”现象, 也无法判定模型的泛化能力, 实用价值较差。

本文提出了神经网络建模的基本原则和步骤,

基金项目: 上海市教委高等学校科学技术发展基金资助项目 (01H03)

收稿日期: 2005-05-04; **修订日期:** 2005-06-28

作者简介: 楼文高(1964~), 男, 教授, 博士, 主要从事人工神经网络理论、多指标综合评价等现代数据处理技术在环境科学与工程中的应用研究与教学工作。

E-mail: wglou@shfu.edu.cn; wglou@usst.edu.cn

并用试凑法确定合理的神经网络结构,用检验样本实时监控训练过程以避免出现“过训练”现象的影响,多次随机改变网络初始连接权值以求得全局极小点邻域内的可行解,从而确保神经网络模型的泛化能力和实用价值。

1 污水处理活性污泥系统简述

松江污水厂采用普通的活性污泥处理方法。根据设计要求,对污水厂尾(出)水的化学需氧量(COD)、五日生化需氧量(BOD_5)、悬浮固体(SS)和氨氮(NH_3-N)4个指标有明确的排放要求,它们的设计标准(排放)分别为100、30、30和5 mg/L。

经分析报表提供的数据,确定影响上述4项尾水指标的因素主要有原污水水质指标COD、 BOD_5 、SS、 NH_3-N 、pH、TKN和运行控制参数的水力平均停留时间(h,沉砂池或曝气池容积/进水流量)、曝气池内的污泥浓度MLSS和MLVSS、污泥沉降比SV(%)等11项指标(注:除pH值和标明的单位外,文中其他变量的单位均为mg/L)。

根据要求,需要建立的是某时出水水质与该部分水进入污水厂时的水质及其运行控制参数之间的数学模型,而不能建立同一时刻出水水质与进水水质、运行控制参数之间的关系,这一点务必不能混淆。根据该厂设计资料知污水从进厂到出厂的水力平均停留时间为21 h,实际运行的水力平均停留时间为25 h。因此,建立某天的出水水质与前一天同一时刻的进水水质与当天运行控制参数之间的数学模型是比较合理的。

2 人工神经网络模型

人工神经网络在20世纪80年代后得到迅速发展,是目前最活跃的前沿学科之一,尤其适合于处理非线性系统。在现有的人工神经网络应用中,80%~90%采用BP网络模型(包括用拟牛顿法、L-M法等)。它由一个输入层、一个输出层和若干个隐(含)层组成,工作原理详见文献[2~4]。

2.1 BP网络模型存在的不足

BP网络建模首先要解决的核心问题是:(1)如何取得合理的网络结构;(2)训练过程中如何避免出现“过训练”现象和求得全局极小点邻域内的可行解。网络结构太大,是出现“过训练”的内因,训练时很容易出现“过训练”现象,而结构太小难以达到满意的精度。训练时出现“过训练”现象,建立的

模型就没有泛化能力,模型就没有实用价值。合理的隐层节点数与输入层、输出层节点数、问题的复杂程度等因素有关,但迄今还没有理论计算公式。值得注意的是,BP网络模型良好的非线性逼近能力和泛化能力是在具有合理隐层节点数和不发生“过训练”现象时才具有的,而不是模型本身所固有的。

2.2 建立BP网络模型的基本原则和步骤

为了确保BP网络模型的泛化能力,必须解决好上述核心问题,即建模必须遵循一定的基本原则和步骤。

2.2.1 样本数据的筛选与分组

BP神经网络建模属于数据驱动型建模技术,具有garbage in and garbage out的特性。因此,为确保模型的有效性,首先应对样本数据进行筛选,剔除异常数据和错误数据(见第3部分第1段)。

将经筛选后得到的有效数据随机分成训练样本、检验样本(10%以上)和测试样本(10%以上)3部分。训练样本用来根据一定的算法调整网络连接权值而使误差函数趋于极小。检验样本用来实时监控训练过程,并判定训练过程是否出现了“过训练”现象^[2-4,8-10]。随着训练次数的增加,训练样本误差始终减小,而检验样本误差从开始时的减小继而出现增大的趋势,则表明出现了“过训练”现象。测试样本用来判定建立的模型的泛化能力。如果检验样本、测试样本的误差与训练样本误差基本相等或稍大,表明模型具有较好的泛化能力。目前国内(污)水处理系统神经网络建模的所有论文均没有检验样本(有的论文把测试样本误称为检验样本)^[4,5],无法判断这些论文的建模(训练)过程是否发生了“过训练”现象!

2.2.2 确定合理的网络结构

针对一定数量的训练样本,总存在一个合理的隐层数和隐层节点数。合理的网络结构是指在满足精度要求的前提下取尽可能紧凑的结构,即网络连接权值数最少。必须注意2点:(1)一般取一个隐层;(2)对于3层网络,输入层和隐层节点数必须至少比训练样本数少一个,否则,网络总可以以任意精度逼近期望值(即训练样本误差趋近零)^[2-4,8-10],但可能没有泛化能力(即非训练样本误差比训练样本误差大得多)。一般地,训练样本数要多于网络连接权值数,如果没有检验样本实时监控训练过程,训练样本数是网络连接权值数的10倍以上时通常也能取得较好的结果^[4,6,11]。合理的隐层节点数不

仅与输入/输出层节点数有关,更与需解决问题的复杂程度等因素有关。目前各种文献提供的确定隐层节点数的计算公式都是针对训练样本任意多或个别特殊情况,不具有一般性,不宜直接采用。目前最有效的途径是,采用试凑法(逐步增加或减少节点)确定合理的隐层节点数。

2.2.3 随机改变网络初始权值

如果网络初始连接权值相同,则每次训练只能搜索到相同的极值点(局部或全局),很难求得全局极小点邻域。由于误差函数存在很多局部极小点,因此,程序必须具有能够随机改变网络初始连接权值的功能^[8,9]。

2.2.4 训练网络模型

训练 BP 网络就是根据算法不断调整网络连接权值,使训练样本的误差平方和达到最小或小于某一期望值。从存在性结论知,即使每个训练样本的误差都很小,非训练样本误差仍可能很大。必须用非训练样本(本文称为检验样本和测试样本)误差来表示和评价建立的模型是否已有效逼近训练样本所蕴含的规律,如果非训练样本误差与训练样本误差很接近^[2~4,9~11],说明建立的模型已有效逼近训练样本所蕴含的规律,否则只能说明模型仅在这些训练样本点上逼近而已。对于同一网络结构,通过选取多组(通常是几十组,由问题的复杂程度而定)不同的网络初始连接权值对网络进行训练,选取没有发生“过训练”、精度较高的网络连接权值(即可认为是全局极小点)。

2.2.5 确定合理的网络模型

合理的网络模型是指具有合理隐层及其节点数、训练时没有发生“过训练”现象而求得全局极小

点邻域可行解时建立的神经网络模型。

3 建立活性污泥系统神经网络模型

取松江污水厂 2003 年 1 月 1 日~2004 年 10 月 31 日实际运行的 669 组数据为研究的原始数据。对原始数据首先剔除不完整的数据,在校核无误的情况下,再根据污水本身固有的大小关系(污水处理机理准则)进行判断,校核数据。如无论是进水还是出水, BOD₅ 值都应小于 COD 值; MLVSS 值都应小于 MLSS 值,等等。经上述整理与预处理,得有效数据 630 组。

针对 630 组有效数据,取 29 组(2004 年 10 月份有效数据)为预测数据,另各随机抽取 60 组数据为检验样本和测试样本,其余 481 组数据为训练样本。

大量实践和示例试算表明:训练多个具有一个输出的网络模型比训练一个具有多个输出的网络模型要方便和简单得多,而且也不会影响后续的预测与控制。因此,根据本研究的原始数据,神经网络模型输入层有 11 个节点(变量),输出层有 1 个节点(本文以出水 COD 为例)。根据训练样本数要多于网络连接权值数的要求,隐层节点数不能多于 36 个(连接权值数为 471)。采用逐步增加节点的试凑法:每次增加 3~5 个隐层节点,针对每个结构,随机改变初始权值试算 50 次以上。

本研究采用 STATSOFT 公司出品的 STATISTICA Neural Network 软件^[9]。经试算,当隐层节点数为 35 个(连接权值数为 456)时,模型具有较好的泛化能力,训练样本、检验样本、测试样本的误差和数据本身特征值如表 1 所示。检验样本、测试样本的实测值与模型计算值如图 1 和图 2 所示。

表 1 训练样本、检验样本、测试样本和预测数据的各项模型性能指标值和数据特征值

Table 1 Various model performance values and eigenvalues of training, verification, testing and predicting set data

指 标	训练样本	检验样本	测试样本	预测数据
	模型值/实际数据	模型值/实际数据	模型值/实际数据	模型值/实际数据
RMSE	5.5526	5.7928	5.9322	4.7764
AAE	3.861	4.397	4.125	4.261
R	0.9300	0.9013	0.8998	0.6843
MAPE	7.13%	8.29%	8.35%	8.32%
σ	14.97/15.11	12.53/13.48	12.46/13.71	7.78/6.67
平均值	57.24/57.35	56.66/56.95	56.67/55.48	52.06/51.93
最大值	107.43/110.00	95.71/97.90	100.45/103.80	64.24/60.60
最小值	30.09/30.30	31.89/30.90	37.34/34.70	38.49/38.20

注:RMSE、AAE、R、MAPE 和 σ 分别表示均方根误差、绝对误差均值、相关系数、相对误差均值和标准剩余方差

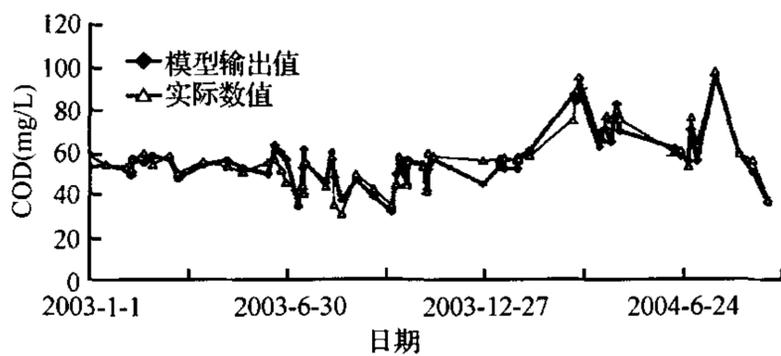


图1 检验样本的实际数值和模型输出值
Fig. 1 Target values and model output of the verification set data

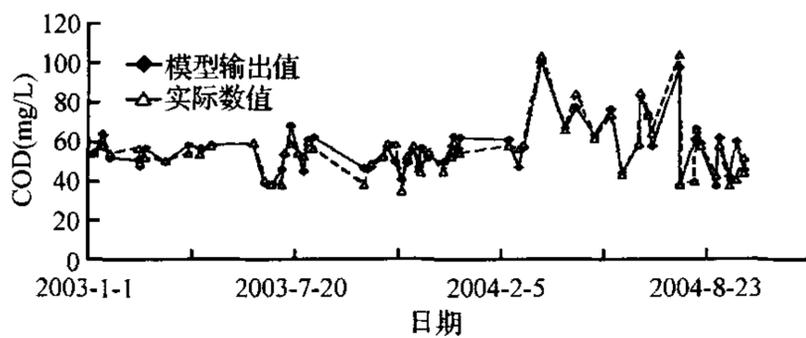


图2 测试样本的实际数值和模型输出值
Fig. 2 Target values and model output of the verification set data

4 活性污泥系统的运行控制

活性污泥系统可以用神经网络模型简洁地表示为:

$$y_{\text{COD-o}} = f(X, W_h, W_o) \quad (1)$$

式中: $y_{\text{COD-o}}$ ——出水 COD;

X ——输入变量;

W_h, W_o ——输入层与隐层和隐层与输出层之间的网络连接权值, 已通过前述训练得到。

因此, 针对一定的进水水质, 可以通过调用神经网络模型, 调整某一个或几个控制参数控制活性污泥系统出水 COD 的值, 同时实现对活性污泥系统的仿真和出水水质的预测。

本研究以调整曝气池内 MLSS 为例加以说明, 其他变量取实际运行数据的均值, 如表 2 所示, MLSS 的单变量有效取值范围虽然为 4300 ~ 6600 mg/L, 由于 $MLVSS/MLSS = 0.55 \sim 0.75$, 因此, 实际的 MLSS 有效范围为 4600 ~ 6200 mg/L。

调用建立的网络模型, 分别设定进水 COD 的值为 300、400、500、600、700、800、900 和 1000 mg/L, 将 MLSS 的最小值设定为 4600 mg/L, 最大值设定为 6200 mg/L, 取 8 等分, 运行程序得不同进水 COD、MLSS 时出水 COD 的值, 如图 3 所示。

表2 神经网络模型 10 个输入变量(指标)的均值
Table 2 Average values of ten input variables in neural network model

项目	均值
pH(进)	7.35
COD(进)(mg/L)	879
BOD ₅ (进)(mg/L)	297
SS(进)(mg/L)	599
NH ₃ -N(进)(mg/L)	31.70
TKN(进)(mg/L)	75.24
HRT1(h)	3.77
HRT2(h)	13.45
SV(%)	58.67
MLVSS(mg/L)	3438

由图 3 可知: ①就进水 COD、MLSS 而言, 进水 COD < 800 mg/L 时, 在整个 MLSS 范围内, 出水 COD 的值均小于 60 mg/L (国家二级排放标准); 进

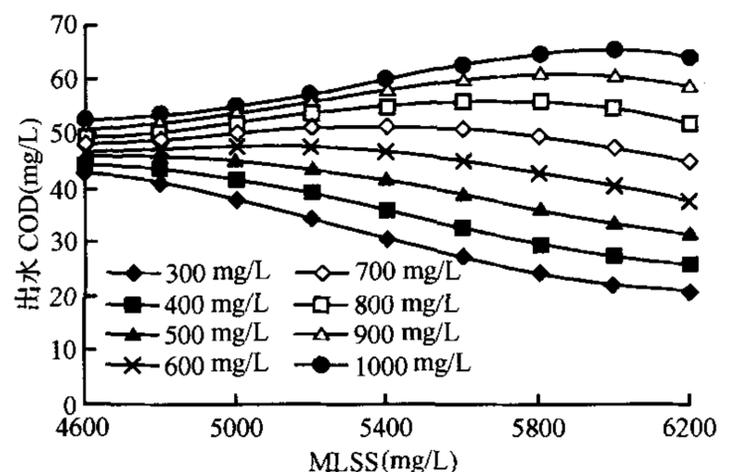


图3 不同进水 COD、MLSS 时的出水 COD 值
Fig. 3 Values of effluent COD with the various influent COD and MLSS

水 COD > 900 mg/L 时, MLSS < 5600 mg/L 时出水 COD 的值 < 60 mg/L; ②不同的进水 COD, 出水 COD 随 MLSS 的变化规律也不同, 当进水 COD < 600 mg/L 时, 出水 COD 随 MLSS 的增大而降低, 即较大的 MLSS 有利于降低出水 COD; 进水 COD > 700 mg/L 时, 出水 COD 随 MLSS 的增大先升高而后降低, 即较小的 MLSS 有利于降低出水 COD; ③在上述进水 COD 范围内, 出水 COD 都能达到设计的排放要求, 运行中究竟采用何种控制方案主要取决于经济性和便利性; ④对于相同的 MLSS, 进水 COD 越高, 则出水 COD 值也越大。

5 结束语

(1)对于高度非线性、工作机理不甚清楚的污水处理活性污泥系统,采用适用于非线性、黑箱系统建模的神经网络技术,理论上能取得较好的效果,具有较好的研究价值。

(2)针对 BP 网络模型建模,训练时如何避免进入局部极小点和出现“过训练”现象,并求得合理的神经网络结构是关键。本文提出了相应的建模基本原则和步骤:通过检验样本实时动态监控训练过程,如发现进入局部极小点和出现“过训练”现象立即停止,重新开始训练;采用逐步增加或减少节点的试凑法确定合理的网络结构,通过很多次随机改变网络初始连接权值以求得全局极小点邻域内的可行解。实例建模表明:遵循上述基本原则和步骤可以建立泛化能力较好的、有效的、可靠的活性污泥系统神经网络模型。

(3)可以很方便地利用建立的神经网络模型,实现对活性污泥系统运行情况的仿真和出水水质的预测,并可验证现有的活性污泥系统机理规律或发现新的机理规律,具有较好的实践意义。

(4)利用建立的神经网络模型,可根据不同的进水水质情况,制订相应的运行参数调整控制方案,以确保出水水质达到设计排放要求并实现活性污泥系统的优化控制;结合经济性模型,能确定污水厂的最佳运行方式,提高污水厂的科学管理水平,降低运行成本。

(5)神经网络建模属于数据驱动型建模技术,模型的可靠性除建模过程正确外,模型的有效性还取决于用于建模的数据的代表性和典型性的好坏和数量。因此,应针对运行工况(进水水质、流量、运行控制参数等)相差较大的不同情况,分别收集足

够多的数据,根据本文提出的建立 BP 模型的基本原则和步骤,建立可靠、有效的模型,从而建立一组完整的活性污泥系统神经网络 BP 模型,以适应于不同应用情况,这将是进一步研究的重要内容之一。

(本研究得到上海市松江污水处理厂有关领导和技术人员的大力支持和帮助,特致感谢。)

参 考 文 献

- [1] 张自杰,林荣忱,金儒霖. 排水工程(下册). 北京:中国建筑工业出版社,2001
- [2] 王文成. 神经网络及其在汽车工业中的应用. 北京:北京理工大学出版社,1998
- [3] 谢庆生,尹健,罗延科. 机械工程中的神经网络方法. 北京:机械工业出版社,2003
- [4] 阎平凡,张长水. 人工神经网络与模拟进化计算. 北京:清华大学出版社,2000
- [5] S. Krovvidy, W. Wee. A knowledge based neural network approach for waste water treatment system. IJCNN International Joint Conference on Neural Networks, 17 ~ 21 June 1990
- [6] 田禹,王宝贞,周定. BP 及 RBF 人工神经元网络对臭氧生物活性炭水处理系统建模的比较. 中国环境科学, 1998, 18(1):394 ~ 397
- [7] 楼文高. 活性污泥工艺神经网络模拟与控制研究[博士学位论文]. 上海:同济大学,2005
- [8] 董聪. 多层前向网络的全局最优化问题. 大自然探索, 1996, 15(58):27 ~ 31
- [9] Statsoft. Statistica Neural Networks. (Manual) Tulsa: Statsoft, Inc., 1999
- [10] 张乃尧,阎平凡. 神经网络与模糊控制. 北京:清华大学出版社,2000
- [11] 张青贵. 人工神经网络导论. 北京:中国水利水电出版社,2004