

# 基于遗传算法的 BOD 神经网络软测量

田 奕<sup>1,2</sup>, 乔俊飞<sup>1</sup>

(1. 北京工业大学 电子信息与控制工程学院, 北京 100022;

2. 华北科技学院 电子信息工程系, 北京 101601)

**摘要:**针对污水处理过程中关键水质参数无法在线监测的问题,提出基于遗传算法和 BP 神经网络相结合的污水水质软测量方法,该方法采用遗传算法优化神经网络结构和权、阈值分布,再用 BP 算法对神经网络进行训练,得到最优的建模网络。仿真结果表明该方法可以避免单独使用 BP 网络容易陷入局部最小的问题,并能加快全局收敛速度,对水质参数 BOD (生化需氧量)预测实时性好、稳定性高、精度高,可用于污水水质的在线预测。

**关键词:**污水处理;软测量;遗传算法;神经网络

**中图分类号:** TP183

**文献标识码:** A

**文章编号:** 1673-629X(2009)03-0127-03

## NN Soft - Measuring for BOD Predict Based on GA

TIAN Yi<sup>1,2</sup>, QIAO Jun-fei<sup>1</sup>

(1. School of Electronic Information & Control Engineering, Beijing Univ. of Tech., Beijing 100022, China;

2. Department of Electronic & Information Science Engineering, North China

Institute of Science and Technology, Beijing 101601, China)

**Abstract:** Considering that on-line information of some essential wastewater parameters is inaccessible in monitoring and controlling wastewater treatment processes, a soft-measuring technique applied to wastewater quality measurement is put forward based on genetic algorithm (GA) and BP neural networks. This method applies genetic algorithm to optimize structure, weights and thresholds of the neural network, then uses BP algorithm to train the neural network in order to get to the superior network. The simulation results show that the new method can avoid getting into the part minimum problem caused by using BP neural network alone and accelerate the overall converging speed. This model, which is of good real-time property, good stability and high precision, can be applied to on-line predict wastewater BOD.

**Key words:** wastewater treatment; soft-measuring; genetic algorithm; neural network

## 0 引言

随着现代工业的迅猛发展,污水处理成为环境保护领域的一个重要课题。生化需氧量 BOD (Biochemical Oxygen Demand) 是评价污水处理性能的关键指标,也是污水处理自动控制的重要依据,能否对 BOD 进行实时监测已成为提高治污质量的关键。BOD 无法用传感器直接检测,目前污水处理厂大多通过采样离线人工化验的方法得到。这种测定方法繁琐、测定周期较长(5天),测出的结果严重滞后,不能及时反映实际情况,从而限制了污水处理闭环控制系统的使用<sup>[1]</sup>。

为了解决直接测量法存在的问题,近年来在过程控制和检测领域出现了一种新技术——软测量技术,它采用间接测量的思路,利用易于获取的其它测量信息,通过模型计算来实现对被测变量的估计<sup>[2]</sup>。污水处理过程生产条件恶劣,随机干扰严重,具有强非线性、大时变、严重滞后的特点,而且水污染控制系统是一个多变量、多目标、多层次的包含海量信息的复杂系统,各种水质参数之间存在强烈耦合和关联,采用传统的机理建模方法是行不通的。因此文中采用遗传算法和 BP 神经网络相结合的软测量技术,利用可在线测量的过程参数,建立 BOD 的软测量模型,以实现在线估计。

## 1 软测量模型

建立污水水质软测量模型实际上是复杂系统的辨识问题,以软测量模型的估计值作为被控输出反馈信

收稿日期:2008-06-11

基金项目:国家自然科学基金(60674066)

作者简介:田奕(1980-),女,河北石家庄人,讲师,硕士,主要研究领域为模式识别与智能控制;乔俊飞,博士后,教授,主要研究领域为模式识别、智能优化与控制、非线性系统等。

号的控制系统原理图见图 1, 软测量模型的设计思路如图 1 中的软测量框图所示。从图中可以看出, 针对污水处理过程设计一个软测量模型可分为以下几个步骤: 初步确定辅助变量、辅助变量的精选、现场数据采集与数据预处理、建立软测量模型、离线训练模型、模型的在线校正。

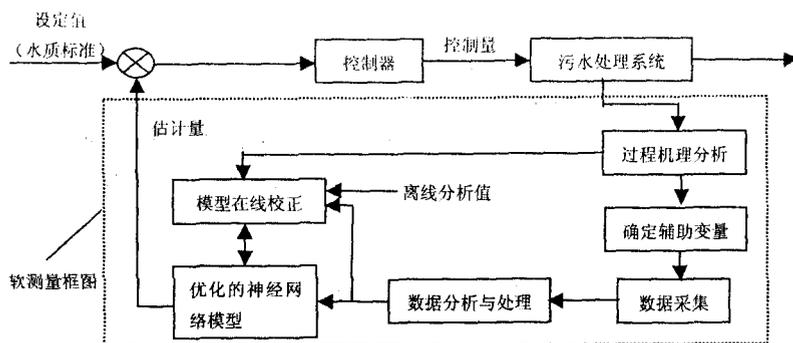


图 1 污水处理控制系统原理图

### 1.1 初步确定辅助变量

通过对污水处理过程及主导变量进行机理分析, 在可测变量集中, 按照灵敏性、特异性、过程适用性、精确性、鲁棒性的原则初步选择所有与主导变量相关的原始辅助变量。

### 1.2 辅助变量的精选

辅助变量的选择范围是对象的可测变量集, 要通过辅助变量的精选, 实现从海量数据中挑选出建立软测量模型所必需的信息。常用的是采用机理分析和数学分析相结合的方法。

### 1.3 现场数据采集与数据预处理

在对变量的数据进行采集时, 采集样本的空间要尽量覆盖整个操作范围, 且注意选择的每一个样本在样本空间内要有一定的代表性, 要选择合适的样本量, 样本数据要均匀。软仪表的性能很大程度上依赖于过程测量数据的准确性和有效性, 因此对获得的数据进行预处理是十分必要的。测量数据预处理包括测量数据的误差处理和数据变换两部分。在本课题中, 主要进行异常数据的剔除, 采用的是统计假设检验法中的拉依达准则 ( $3\sigma$  准则)。之后再对数据进行归一化处理, 使各变量的均值为 0, 标准差为 1, 进而消除由于不同特征因子量纲不同和数量级不同所带来的影响。

### 1.4 建立软测量模型并加以训练

软测量技术的核心是建立对象的数学模型。软测量建模有两种基本方法: 机理建模和辨识建模。机理建模是通过对被测对象的机理分析, 找出不可测主导变量和可测辅助变量之间的关系, 以数学表达式的形式进行计算。从理论上来说, 机理模型是最精确的模型, 然而它要求对被测对象的内部特性完全了解。由

于污水处理过程的复杂性, 往往难以完全通过机理分析得到软测量模型。因此, 基于机理分析的方法建模非常困难。辨识建模是直接利用输入输出数据所提供的信息建立数学模型, 这种方法可以在最少限度先验知识和假设的条件下建模。神经网络方法是软测量技术中的常用方法, 将神经网络用于建立污水指标监测

软测量模型, 与传统的建模方法相比, 它需要的先验知识较少, 而且避免了复杂棘手的模型结构辨识问题, 可以很好地描述实际对象的特性。文中采用遗传算法和 BP 神经网络相结合的软测量技术<sup>[3,4]</sup>, 利用可在线测量的过程参数, 建立 BOD 的软测量模型, 以实现在线预测。

BP 算法是应用最广泛的神经网络训练算法, 但这种算法要求目标函数连续可导, 在局部搜索时比较成功。其主要缺点是收敛

速度慢, 容易陷入局部最优, 而且对于较大的搜索空间、多峰值和不可微函数也不能搜索到全局最小点。GA 算法采用整体搜索策略, 优化计算不依赖传统的梯度信息, 更适合解决高度复杂的非线性问题, 搜索过程中不易陷入局部最优, 更易找到全局最优, 但容易出现计算速度低、早熟收敛和缺乏局部优化能力等缺点。因此将两者结合起来取长补短可实现高效全局的参数优化。具体做法是先由遗传算法优化出神经网络结构和最优权、阈值分布范围, 再由 BP 算法继续训练, 在这个最优范围中寻找网络最优解, 从而得到最优的建模网络。这样做一方面克服了梯度法对初始值的依赖和局部收敛问题, 另一方面也克服了单纯遗传算法所特有的盲目性和概率性问题, 提高了整体的搜索效率。

通过对污水处理过程的机理分析可知与 BOD 耦合和关联最大的几个辅助变量是 COD、SS、DO 和 PH。因此确定神经网络输入节点为 4 个, 隐含层节点数由遗传算法优化得到<sup>[5]</sup>, 输出节点为 1 个, 建立一个 MISO 系统模型。网络优化流程如图 2 所示。

#### (1) 初始化。

采用实数编码方案, 码串由四部分组成: 隐层节点数、各层单元之间的连接权值、隐层单元阈值、输出层单元阈值。由于在编码中, 隐节点个数不定, 所以网络优化时码串长度不定, 会给遗传操作带来不便。为保持个体长度的一致性, 保证交叉、变异时子代个体的完整性, 取码串的最大可能长度进行编码。随机产生一组结构和权值各不相同的单隐层前馈网络, 构成初始种群, 初始权值和阈值在  $[-1, 1]$  之间选择。种群规模  $N = 50$ , 设隐节点数为  $H$ , 根据经验  $H \in [4, 10]$ , 码串

长度  $L = (p + q + 1) \times H_{\max} + q + 1 = 62$ , 式中  $p, q$  分别为输入层、输出层单元数。

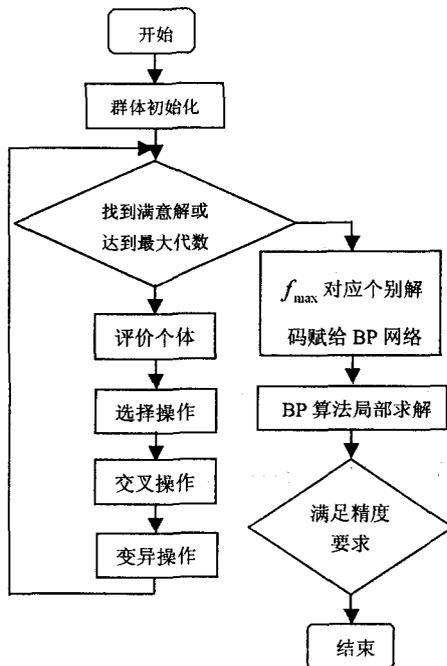


图2 GA-BP网络优化方法流程图

(2) 计算网络误差。

对于给定样本输入, 计算每个网络输出与给定样本输出之间的误差值, 确定目标函数  $E(n) = \frac{1}{N} \sum_{k=1}^N (y(k) - \bar{y}(k))^2$ , 式中,  $n$  表示种群中的个体,  $n \in [1, N]$ ;  $y(k), \bar{y}(k)$  表示第  $k$  个样本输入时, 输出节点的期望输出和网络实际输出。

(3) 定义适应度函数。

将目标函数映射成适应度函数  $f, f(n) = 1/E(n)$ 。

(4) 进行遗传操作。

根据适应度值进行遗传操作, 为保证个体的多样性和提高算法的全局收敛性能, 选择采用小生境法、最优选择机制与轮盘赌法相结合的方法进行。采用自适应算术交叉, 非一致变异、交叉和变异概率分别由式(1)和式(2)调整计算。设定遗传代数  $T = 100$ , 在全局范围内得到一组适应度最大的个体, 即可反映出最优的网络结构和权阈值分布。

$$p_c = \begin{cases} p_{c_1} \frac{f_{\max} - f}{f_{\max} - f_{\text{avg}}} + p_{c_2} \frac{f - f_{\text{avg}}}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}} \\ p_{c_1}, & f < f_{\text{avg}} \end{cases} \quad (1)$$

式中  $p_{c_1} = 0.9, p_{c_2} = 0.6$ 。

$$p_m = \begin{cases} p_{m_1} \frac{f_{\max} - f}{f_{\max} - f_{\text{avg}}} + p_{m_2} \frac{f - f_{\text{avg}}}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}} \\ p_{m_1}, & f < f_{\text{avg}} \end{cases} \quad (2)$$

式中  $p_{m_1} = 0.1, p_{m_2} = 0.001$ 。

(5) 进行优化。

由遗传算法全局寻优得到的最优个体可知网络的最佳隐节点个数为  $H = 5.831 \approx 6$ , 即最优网络结构为  $4-6-1$ , 码串的有效长度为  $(4+1+1) \times 6 + 1 + 1 = 38$ , 将有效码串解码赋给 BP 网络, 作为网络初始权阈值。用基于  $L-M$  规则的 BP 算法对遗传算法优化过的权阈值进行训练, 局部求解, 以精确地得到网络结构的参数(权值、阈值)。网络的隐含层采用 Sigmoid 函数, 输出层采用线性函数。神经网络的训练参数为: 训练次数  $\text{epochs} = 2000$ , 训练目标  $\text{goal} = 0.002$ , 学习速率  $\text{lr} = 0.01$ 。

2 仿真试验研究

实验数据来源于某污水处理厂水质监测分析报表。将给定的输入输出样本集分为训练样本和测试样本, 采用 MATLAB 对 GA-BP 神经网络软测量模型进行仿真实验。将数据预处理剔除异常数据和标准化之后剩下的 66 组数据样本分为两部分: 其中 46 组数据用作训练样本, 其余 20 组数据作为校验样本。图 3 为 GA-BP 算法的出水水质 BOD 拟合情况图, 图 4 为 GA-BP 算法的出水水质 BOD 预测情况图 (X 轴: 输入样本点, Y 轴: 生化需氧量 BOD(mg/L))。通过仿真结果可以看出, 应用遗传算法优化神经网络达到了很好的效果, 由 GA-BP 算法构造的神经网络软测量模型具有较高的预测精度。

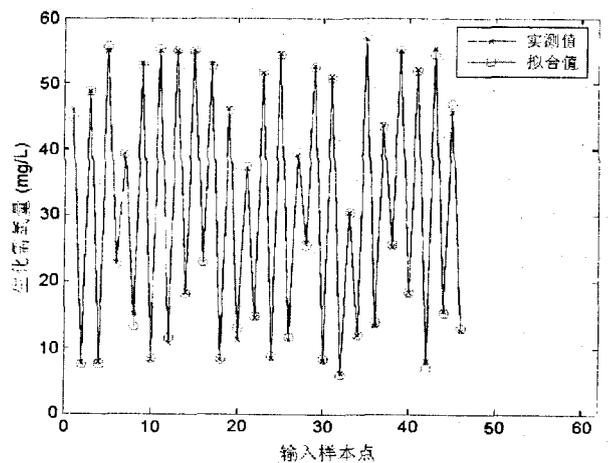


图3 BOD拟合情况

3 结束语

通过对污水处理过程的机理分析, 选择与 BOD 耦合和关联最大的几个易测参量 COD、SS、DO 和 PH, 运用软测量技术, 采用 GA-BP 网络建立了出水水质 BOD 软测量模型。通过仿真研究证明了该模型的有

(下转第 133 页)

(60,45,0.8)、(60,30,0.9)。这里假设三个质量指标的权重分别为0.3,0.3,0.4。

依据公式(1)、(2)可得:

对 Adv1:  $V_{11} = 1, V_{12} = 1, V_{13} = 1$ ;

对 Adv2:  $V_{21} = 44/45, V_{22} = 11/16, V_{23} = 0$ ;

对 Adv3:  $V_{31} = 44/45, V_{32} = 7/8, V_{33} = 1$ ;

所以  $QoSMatch(Adv1, Req) = Score(s_1) = 1 \times 0.3 + 1 \times 0.3 + 1 \times 0.4 = 1$ ;  $QoSMatch(Adv2, Req) = Score(s_2) = 44/45 \times 0.3 + 11/16 \times 0.3 + 0 \times 0.4 = 0.4996$ ;  $QoSMatch(Adv3, Req) = Score(s_3) = 44/45 \times 0.3 + 7/8 \times 0.3 + 1 \times 0.4 = 0.9558$ 。

假设用户设定的服务质量相似度阈值为0.75,则返回的满足用户需求的服务列表次序为 Adv1、Adv3。

#### 4 实验结果与分析

为了测试和验证匹配方法,实验中提供了图形化界面,通过指定对各个阶段的最小匹配等级,然后按下 <Match> 按钮开始进行服务匹配,当期望得到的匹配等级都小于服务匹配方法推出的匹配等级时,返回 true,即满足所期望的匹配结果。

通过实验分析,该匹配算法提高了服务的查准率和查全率。

与普通的基于属性值的服务查找相比,文中的服务查找增加了语义信息,使得找到的服务与实际需要的服务更加符合,从而增加了查准率。

文中首先在本体的语义表达部分引入了等价和子类关系,使得在匹配中相关本体中的实体可以匹配,增加了可能匹配服务的范围,使得服务的查全率有所提高。

#### 5 结束语

介绍了语义 Web 服务的描述语言 OWL-S,设计并实现了一个基于普适计算的服务发现系统。针对 OWL-S 描述的服务的发布和查询过程进行了介绍,重点介绍了服务匹配算法。通过定义服务发布者和请求者共同遵守的本体和服务的语义描述,达到可以实现语义层次的匹配,从而为服务与服务之间的自动组合和协作奠定了基础,使得动态地利用网络中的资源,即时组合出新的应用成为可能。

#### 参考文献:

- [1] 徐光祐,史元春,谢伟凯. 普适计算[J]. 计算机学报,2003,26(9):1042-1050.
- [2] OWL Web Ontology Language Guide[R/OL]. 2003. W3C Candidate Recommendation 18 August 2003. <http://www.w3.org/TR/2003/CR-owl-guide-20030818/>.
- [3] Martin D. OWL-S: Semantic Markup for Web Services[R/OL]. 2004-02. Technical report, Daml consortium, <http://www.daml.org/services/owl-s/1.0/owl-s.pdf>, WRSP Primer Working Draft 0.3.
- [4] 赵军,基于 OWL-S 的 Web 服务发现系统的研究和实现[J]. 计算机技术与发展,2006,16(10):163-166.
- [5] Kopena J, Regli W. DAMLJessKB: A tool for reasoning with the semantic web[J]. In IEEE Intelligent Systems,2003,18:74-77.
- [6] Mokhtar S B. EASY: Efficient semantic Service discovery in pervasive computing environments with QoS and context support[J]. J. Syst. Software,2007,10:9-11.
- [7] 仲梅. 基于语义的 Web 服务发现[D]. 镇江:江苏大学,2006:35-37.

(上接第 129 页)

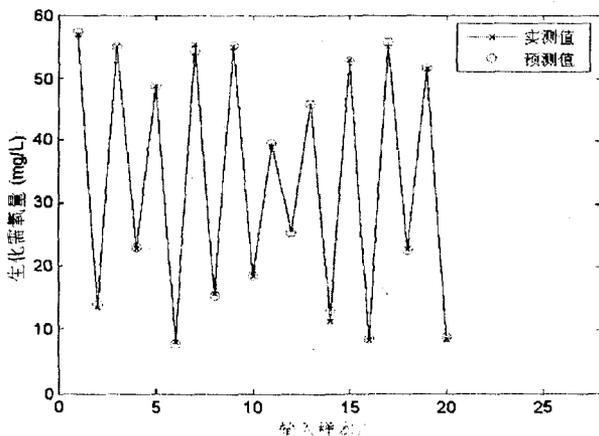


图4 BOD预测情况

效性,能实现较高的预测精度。该模型的实现对进一步推动污水处理过程实时闭环控制的实现和开发其他

工业过程的具有智能优化算法的软测量仪表具有一定的启发意义。

#### 参考文献:

- [1] 邹家庆. 工业废水处理技术[M]. 北京:化学工业出版社,2003.
- [2] 于静江,周春晖. 过程控制中的软测量技术[J]. 控制理论与应用,1996,13(2):137-144.
- [3] 欧阳林群. 基于 GA-BP 的个股超短线预测应用研究[J]. 微计算机信息,2007,1(2):306-307.
- [4] Whitley D, Starkweather T, Bogart C. Genetic algorithms and neural networks: Optimizing connections and connectivity[J]. Parallel Compute,1990,4(3):347-361.
- [5] 张庆红,程国建. 基于遗传算法的神经网络性能优化[J]. 计算机技术与发展,2007,17(12):125-127.