

基于主元分析和神经网络的污水处理能耗分析

王丽娟

(西安工业大学 电信学院, 陕西 西安 710032)

摘要:目前水处理是一个高能耗行业,在污水处理厂的各种能耗中,污水处理曝气阶段的能耗是一个重要部分,几乎占到总能耗的60%。提供了一种简单有效的污水处理曝气能耗监测模型,该模型可以根据污水厂的入水水质参数及其他运行参数预测污水处理曝气能耗,有助于控制系统根据实时能耗调节控制策略。文中首先采用主元分析法(PCA)简化该模型的输入参数,其次采用神经网络模型对该模型的输入参数及输出参数进行映射。通过某污水厂的运行数据对该模型进行训练和测试,结果表明该模型性能学习精度高,有效预测污水处理曝气能耗。也表明污水处理曝气能耗不仅和入水流量、pH、BOD有关,也和曝气池温度有关。

关键词:污水处理厂;曝气能耗;化学需氧量;生物需氧量;BP网络

中图分类号: TP183

文献标识码: A

文章编号: 1673-629X(2011)03-0243-03

Analysis of Sewage Treatment Aeration Energy Consumption Based on PCA and BP Networks

WANG Li-juan

(Information School, Xi'an Technology University, Xi'an 710032, China)

Abstract: Sewage treatment plant is a high energy consumption industry by now. Sewage aeration energy consumption is the main of total sewage treatment plant energy consumption, nearly accounting for 60% of the total. Present a simple, powerful aeration energy consumption monitor model, which supplies the real time energy consumption through the influent quality parameters and other parameters. The model is helpful for advanced control system to adjust its control opinion. To make the model simple, PCA is used to choose the influent parameters affecting energy consumption strongly as few as possible. The model established by back-propagation network is trained and tested by sets of a sewage treatment plant operational data. The test result shows that the model works well with high efficiency and accuracy. The test result also shows aeration energy consumption is not only affected by influent flow, pH, biological oxygen demand (BOD), but also affected by temperature in aeration tank in the sewage treatment plant.

Key words: sewage treatment plant; aeration energy consumption; COD; BOD; BP network

0 引言

近年来随着水资源日益紧张和水环境日益被污染,污水处理厂开始受到关注。作为一个高能耗处理工艺过程,污水处理能耗可以分为三个部分,即预处理能耗、生化处理能耗和污泥处理能耗。其中,曝气能耗是最主要的能耗,占总量的60%^[1]。为降低污水处理能耗,研究者提出了许多的观点及方法,例如改进曝气设备、改进曝气工艺。除此之外,采用先进的控制系统对污水处理进行控制也是一种普遍的做法^[2-6]。在先进的控制系统中嵌入一个实时污水处理能耗监测模型是非常有益的,有助于控制系统实时调整控制策略^[7]。

许多污水能耗模型都是建立在水质参数的削减量或出水水质的基础上^[7],无法在污水处理过程中动态地预测能耗。

文中提出一种基于入水水质参数的污水处理曝气能耗监测模型,该模型使用主元分析法降低了输入参数的维度,结构简单;使用BP神经网络对在线数据进行学习,可以在污水处理过程中动态预测能耗,并且精度较高。文中研究的对象是太原某生活污水处理厂,其污水厂的排放符合国家标准。文中以曝气过程中,污水中DO浓度代表曝气能耗。

1 PCA简介

主元分析法是一种多变量统计方法,用于分析数据之间的线性关系,将原有的多个相关变量转化为少数几个相互独立的综合变量,在减少数据集变量维数的同时保留其绝大部分动态信息^[8]。

收稿日期:2010-07-05;修回日期:2010-10-17

基金项目:陕西省教育科研2009年度重点计划项目(09JK499)

作者简介:王丽娟(1980-),女,山西人,讲师,主要从事自动化、智能控制等方面研究。

已知 X 为 $n \times m$ 数据矩阵, n 为样本数, m 为变量个数, PCA 可将矩阵 X 分解为如下形式^[9]:

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E$$

其中 $k \leq \min(m, n)$, t_i 为得分 (score) 向量, 也称主元; p_i 为负荷 (loading) 向量, 各个得分向量之间是正交的, 各个负荷向量也是互相正交的, 同时每个负荷向量的长度都是 1。前 k 个负荷向量 p_i 定义为 X 的最大变化方向, t_i 是每个变量在 p_i 上的投影, E 为误差矩阵, 即每个测量值距离主要元素平面的正交距离。对矩阵 X 进行主元分析需要求出协方差矩阵 $X^T X$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ 及其对应的特征向量 p_1, p_2, \cdots, p_m 。这些特征向量即为矩阵的负荷向量。

最终的数据变量维数 k 可通过累积方差百分比 (Cumulative Percent Variance, CPV) 来确定, 计算公式:

$$CPV(k) = 100 \left[\frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^m \lambda_j} \right] \%$$

用这种方法, 根据实际需要选定的 CPV 的期望值 (如 85%) 确定应当保留的主元个数。

2 数据预处理

文中所建立的能耗预测模型是一个基于大量运行数据的智能模型, 对于基于数据的模型来说, 必须保证数据的精确性, 需进行预处理^[7], 文中采用了两种方法来实现这一点。首先采用一种聚类方法来消除异常值^[10]。在这种方法中, 每个数据对象与距其最近的对象距离满足 2 倍最小平均距离时, 该数据对象与距其最近的对象被划分为一类。而所谓最小平均距离的计算公式如下:

给定一组 d 维空间的数据对象 y_1, y_2, \cdots, y_n , 则

$$MMD = \frac{1}{n} \sum_{i=1}^n \min_{j \neq i} \left[\left(\sum_{k=1}^d (y_{ik} - y_{jk})^2 \right)^{1/2} \right]$$

如果一个数据对象与距其最近的对象距离大于 2 倍最小平均距离时, 则该数据对象被认为是异常值。这一算法成功地将异常值与正常值区分开了。此外, 文中还采用了移动平均滤波的方法来减小噪声。数据的部分处理结果如图 1, 2 所示。

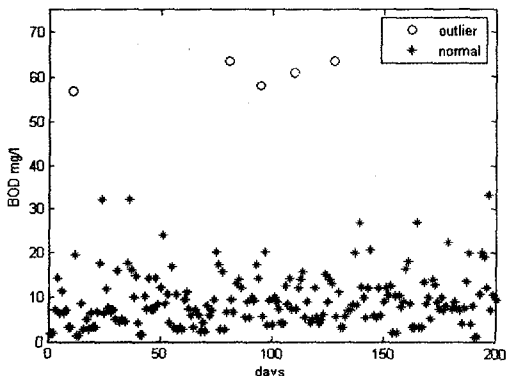


图 1 部分 BOD 数据

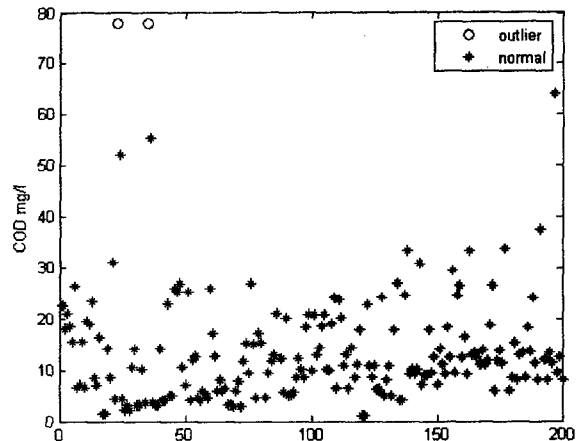


图 2 部分 COD 数据

3 基于 PCA 和 BP 神经网络能耗预测模型

根据污水生化处理工艺能耗对入水水质参数响应有差异的特点, 利用 PCA 主元分析法选取特征参数, 利用 BP 神经网络模拟人体大脑以非线性拟合的方法分析所得特征参数, 建立一个能耗预测模型。具体步骤如下:

1) 原始样本 $X_{n \times m}$ 的标准化: 将选取的过程变量首先经过标准化处理以消除量纲不同带来的不合理影响, $X_i^* = (X_i - E(X_i)) / (\text{var}(X_i))^{1/2}$ 。

2) 计算主成分值^[9]: 采用矩阵简化奇异值分解方法计算始样本的主成分值。设标准化后的样本为 $X_{n \times m}^*$, 则 $X_i^* = U \Sigma V^T$, 式中 $U_{n \times n}$ 和 $V_{m \times m}$ 为奇异向量构成的两个正交矩阵, Σ 为由奇异值构成的对角方阵, 即 $\Sigma = \text{diag}(\delta_1, \delta_2, \cdots, \delta_n)$ 。

计算每个主元的方差: $\lambda_i = \delta_i^2 / (n-1)$

计算总的方差: $\lambda = \sum_{i=1}^n \lambda_i = \sum_{i=1}^n \delta_i^2 / (n-1)$

计算累积方差百分比: $CPV(k) = \frac{\sum_{i=1}^k \delta_i^2 / (n-1)}{\sum_{i=1}^n \delta_i^2 / (n-1)}$

3) 找出所有累积方差百分比大于给定 CPV 值的主元个数 k , 从而求出主元, 降低输入变量维数。

4) 给定 BP 神经网络的隐含层数及其神经元个数。

5) 利用测试集检验, 返回上步, 直到检验结束。

6) 最终确定 BP 神经网络的隐含层数及其神经元个数, 建立能耗预测模型, 并用该模型预测。

主成分分析结果^[11]见表 1。从中可以看出, BOD、入水流量、pH 值、SS、COD 这 5 个参数 84.968% 的信息可以用三个主要成分表示出来。

依据表 2 选出与主成分 1 相关性强的变量 BOD 以及与主成分 2 相关性强的变量入水流量、与主成分

3 相关性强的变量 PH, 以及曝气池的温度来作为 BP 网络的输入, 以 DO 浓度作为 BP 网络输出, 对神经网络进行训练^[12,13]。本研究使用三层神经网络, 网络隐含层神经元的传递函数采用 s 型正切函数 tansig, 输出层传递函数采用 s 型对数函数 logsig, 并设定网络的训练函数为 trainlm。第一个隐含层含有 15 个神经元, 第二个隐含层含有 8 个神经元, 其结构见图 3。选择 70 个样本训练该网络, 并使用训练后的网络对其余样本进行测试, 结果见表 3。

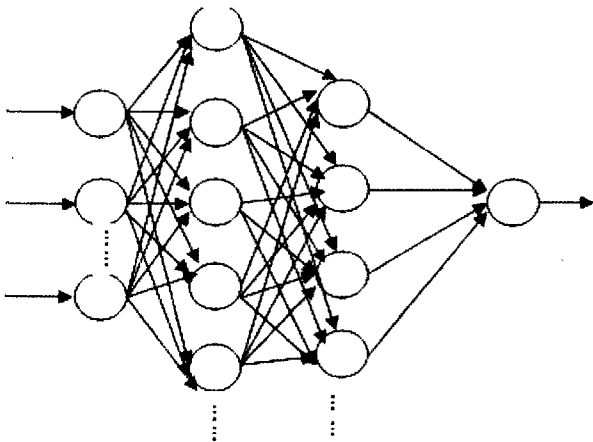


图 3 污水处理能耗模型神经网络结构

表 1 主成分分析结果

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.054	41.072	41.072
2	1.342	26.830	67.902
3	0.811	16.228	84.131
4	0.603	12.064	96.195
5	0.190	3.805	100.000

表 2 各变量与主成分的相关性

	Component		
	1	2	3
V1	-0.149	0.817	0.238
V2	0.276	-0.655	0.647
V3	0.538	0.490	0.488
V4	0.914	-0.027	-0.221
V5	0.912	0.070	-0.224

表 3 实验结果

Trained samples	MSE	Test samples	MSE
70	0.0002	30	0.0012
		30	0.000187
		30	0.00095
		60	0.00020

理入水水质参数的降维时, 效果良好。用降低维度后的入水水质参数建立一个 BP 网络模型, 可以有效预测污水处理曝气能耗。且该模型的学习精度高、学习速度快。结果也表明, 污水处理曝气能耗不仅和入水量、BOD、pH 值有关, 也和曝气池温度有关。

参考文献:

[1] Wang M H, Wang L K. Mathematical modeling of electrical energy consumption and heating requirements by municipal wastewater treatment plants[J]. Journal of Environmental Sciences, 1979, 24: 23-26.

[2] Olsson G, Andrews J F. Dissolved oxygen control in the activated sludge process[J]. Wat. Sci. Tech., 1981, 13(10): 341-347.

[3] Charpentier J, Florentz M, David G. Oxidation-reduction potential (ORP) regulation: a way to optimize pollution removal and energy savings in the low load activated sludge process [J]. Wat. Sci. Tech., 1987, 19: 645-655.

[4] Charpentier J, Godarth, Martin G, et al. Oxidation-reduction potential (ORP) regulation as a way to optimize aeration and C, N and P removal: experimental basis and various full-scale examples [J]. Wat. Sci. Tech., 1989, 21: 1209-1223.

[5] Ferrer J, Rodrigo M A, Seco A, et al. Energy saving in the aeration process by fuzzy logic control[J]. Wat. Sci. Technol., 1998, 38(3): 209-217.

[6] Bongards M, Ebel A, Hilmer T. Predictive control of wastewater works by neural networks[C]//Soft. Comp. Ind. Appl. Proc. Biannual World Autom. Cong. Sevilla, Spain: [s. n.], 2004: 397-402.

[7] 姚红星. 一类神经网络数据预处理的两种方法的比较[J]. 微机发展(现更名: 计算机技术与发展), 2003, 13(6): 77-79.

[8] 贾群. PCA-BP 算法模块化设计的编程实现[J]. 计算机技术与发展, 2008, 18(12): 98-101.

[9] 郑小霞, 钱锋. 基于 PCA 和最小二乘支持向量机的软测量建模[J]. 系统仿真学报, 2006, 18(3): 739-741.

[10] Yin P Y, Chen L H. A new non iterative approach for clustering[J]. Pattern Recognition Letters, 1994, 15(2): 125-133.

[11] 毛丽娟. 基于 SPSS 和 BP 网络的锌产量预测模型[J]. 计算机测量与控制, 2008, 16(8): 1116-1118.

[12] 丁晓贵, 刘桂江. 基于 PIDNN 的污水处理系统参数辨识研究[J]. 计算机技术与发展, 2008, 18(5): 200-202.

[13] 田奕, 乔俊飞. 基于遗传算法的 BOD 神经网络软测量[J]. 计算机技术与发展, 2009, 19(3): 127-133.

4 结束语

实验结果表明, PCA 主成分分析法应用于污水处