

Application of Improved Multivariate Linear Regression Algorithm Based on Haar Wavelet Analysis in MBR

Ming Zhang¹, Chunqing Li¹, Tao Wang², Haitao Wang²

¹School of Computer Science and Software, Tianjin Polytechnic University, Tianjin

²School of Environmental and Chemical Engineering, Tianjin Polytechnic University, Tianjin

Email: 2395521076@qq.com, franklcq@163.com

Received: Aug. 25th, 2018; accepted: Sep. 7th, 2018; published: Sep. 14th, 2018

Abstract

When using the MBR system for wastewater treatment, it is often encountered that a large amount of sludge is trapped on the surface of the membrane. This phenomenon will cause membrane fouling and affect the effect of the MBR system in treating sewage. Therefore, the problem we need to solve is how to predict the MBR. The amount of sludge in the system is used to clean the sludge in time. In order to solve this problem, this paper uses the improved multiple linear regression algorithm based on Haar wavelet analysis to predict the apparent yield coefficient of sludge in MBR system. Through experiments, we compare the predicted results with the actual system results, and find that the algorithm achieves the expected effect of accurately predicting the apparent yield coefficient of the sludge, solves the problem of predicting the amount of sludge in the MBR system, and improves the processing of the MBR system. The sewage effect has a certain value.

Keywords

MBR, Multiple Linear Regression Algorithm, Apparent Yield Coefficient of Sludge, Haar Wavelet Analysis

基于Haar小波分析改进的多元线性回归算法在MBR中的应用

张明¹, 李春青¹, 王韬², 王海涛²

¹天津工业大学计算机科学与软件学院, 天津

²天津工业大学环境与化学工程学院, 天津

Email: 2395521076@qq.com, franklcq@163.com

收稿日期：2018年8月25日；录用日期：2018年9月7日；发布日期：2018年9月14日

摘要

用MBR系统进行污水处理时，经常会遇到有大量的污泥被截留在膜表面的现象，该现象会造成膜污染，影响MBR系统处理污水的效果，因此我们需要解决的问题就是如何预测MBR系统中的污泥量以便及时清理污泥。为了解决该问题，本文利用基于Haar小波分析改进的多元线性回归算法对MBR系统中的污泥表观产率系数进行预测。通过实验，我们将预测结果与实际系统结果做对比，发现该算法达到了能够准确预测污泥表观产率系数的预期效果，解决了预测MBR系统中污泥量的问题，对提高MBR系统处理污水效果具有一定的价值。

关键词

MBR，多元线性回归算法，污泥表观产率系数，Haar小波分析

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来污水处理成为热点话题之一，污水处理有很多种方法，而利用 MBR 技术进行污水处理是其中比较高效的方法之一[1]。MBR 技术是由膜分离技术和生物处理技术相结合而产生的技术，MBR 系统由各个膜组件连接组成，在 MBR 系统运行时污水流过各个膜组件，过滤掉大分子颗粒以此来达到过滤污水的目的[2]。但随着系统的运行会有大量的污泥被截留在膜表面，这些被截留的污泥会影响膜组件的寿命和处理污水的效果，所以我们要解决的问题就是如何预测 MBR 系统中的污泥量[3]。

到目前为止，预测 MBR 系统的模型有很多种，在本文中，我们采用 Haar 小波分析对多元线性回归算法改进的模型来预测污泥表观产率系数。多元线性回归算法非常适用于做数据预测分析，并且它的泛化能力比较好，但多元线性回归算法存在预测结果误差稍大的问题[4]，而 Haar 小波分析可以对训练样本数据做去噪预处理操作，能够提高预测结果的精度。研究发现，利用此模型可以准确地预测污泥表观产率系数，解决了预测 MBR 系统中污泥量的问题。

2. 线性回归算法和 Haar 小波分析算法

2.1. 一元线性回归算法

一元线性回归算法又叫直线拟合算法，一元线性回归算法是研究一个自变量与一个因变量之间相关关系的算法，它是研究两个变量相关关系的最简单模型[5]。在一元线性回归模型中，如果 x_i 与 y_i 存在相关关系，则建立一元线性回归模型($y_i = \alpha + \beta x_i + \gamma_i, i = 1, 2, \dots, n$)和一元线性回归方程($\hat{y} = a + bx$)。其中 $\alpha + \beta x_i$ 表示 x_i 与 y_i 之间的相关关系， γ_i 表示外界影响因素的总和，也叫做随机误差； b 为回归直线的斜率， a 为回归直线的截距。在求解 a, b 时，我们用最小二乘法来计算 a, b ，即 $Q(\alpha, \beta) = \sum \gamma_i^2 = \min$ ，经计算可得到 a, b 如公式(1)所示。

$$\begin{cases} a = \bar{x} - b\bar{x} \\ b = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2} \end{cases} \quad (1)$$

在计算出一元线性回归方程后，为了验证所建立一元线性回归方程的正确性，还需要计算误差。在一元线性回归算法中，我们可以利用拟合度的大小来衡量误差的大小，拟合度表示实际数据点在回归直线 $\hat{y} = a + bx$ 周围的紧凑程度，紧凑程度越大，则误差就会越小，反之则相应的误差就会越大。

2.2. 多元线性回归算法

多元线性回归算法又叫做重回归算法或复回归算法，它是研究多个自变量与一个因变量之间相关关系的算法[6]。多元线性回归算法和一元线性回归算法的原理类似，如果这些变量之间存在相关关系，则建立多元线性回归模型(公式(2))和多元线性回归方程(公式(3))。

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_{m-1} x_{i,m-1} + \gamma \quad (2)$$

$$\hat{y} = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_{m-1} x_{i,m-1}, i = 1, 2, \dots, n \quad (3)$$

令 $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_{m-1})^T$, $\hat{\beta} = (b_0, b_1, \dots, b_{m-1})^T$, $\gamma = (\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_{m-1})^T$ ，其中 β 是未知参数向量， γ 为误差向量；令 $Y = (y_1, y_2, \dots, y_n)^T$, X 为方程(2)的设计矩阵。把方程(2)写成矩阵的形式： $Y = X\beta + \gamma$ ，然后我们要做的就是求解 $\hat{\beta}$ 了。同一元线性回归算法一样，利用最小二乘算法来求解向量 $\hat{\beta}$ ，经过计算得到 $\hat{\beta} = (X^T X)^{-1} X^T Y$ 。

在多元线性回归算法中计算误差时，同样可以利用拟合度的大小来衡量误差的大小，在这里我们利用相关系数 R 的大小来衡量拟合程度，其中 $0 \leq R \leq 1$ 。当相关系数 R 趋向于 1 时，那么说明算法的拟合度高，相应的误差就越小，反之算法的拟合度低，相应的误差就越大。

2.3. Haar 小波分析的分解与重构算法

在 Haar 小波分析中，我们假设 V_j 是具有形式为 $\phi(2^j x - k), k \in Z, x \in R$ 的所有函数组成的空间， W_j 是具有形式为 $\sum_{k \in Z} d_k \psi(2^j x - k), d_k \in R$ 的所有函数组成的空间[7]。如果要处理的信号中含有噪声，那么就要对信号进行分解来除去噪声，即除去信号中属于 V_{j+1} 但不属于 V_j 的信号，把信号分解为 $V_{j+1} = V_j \oplus W_j$ 的形式。如此迭代下去得到 $V_{j+1} = V_0 \oplus W_0 \oplus W_1 \oplus \dots \oplus W_j$ ，即 $f_{j+1} = f_0 + w_0 + \dots + w_j$ ($w_j = \sum_{k \in Z} d_k^j \psi(2^j x - k) \in W_j$, $f_j = \sum_{k \in Z} c_k^j \phi(2^j x - k) \in V_j$)，这样就把信号分解成了若干个分量之和，以便于对信号去噪，图 1 更加清晰的描述了 Haar 小波分解过程[8]。

Haar 小波重构算法是分解算法的逆过程，即 $f = f_0 + w_0 + w_1 + \dots + w_{j-1}$ ，其中 f_0 和 w_i 为分量， $f_0(x) = \sum_{k \in Z} c_k^0 \phi(x - k) \in V_0$, $w_n(x) = \sum_{k \in Z} d_k^n \psi(2^n x - k) \in W_n$ 。图 2 更加详细地描述了 Haar 小波重构过程，

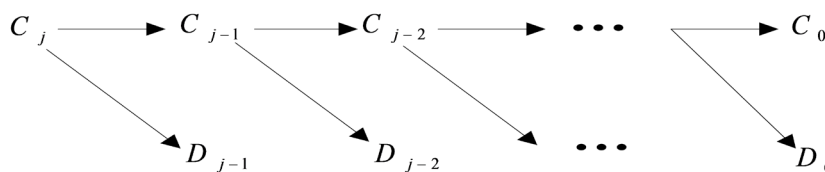


Figure 1. Schematic diagram of Haar wavelet decomposition algorithm
图 1. Haar 小波分解算法示意图

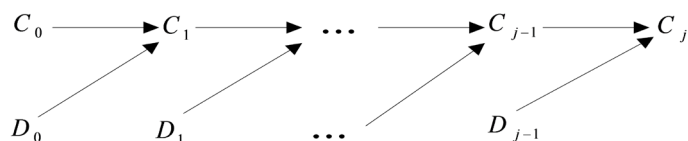


Figure 2. Schematic diagram of Haar wavelet reconstruction algorithm
图 2. Haar 小波重构算法示意图

本文利用 Haar 小波分解算法与重构算法的特性来对多元线性回归模型进行优化[9]。

3. 多元线性回归算法在 MBR 中的应用

3.1. 多元线性回归算法对污泥表观产率系数的计算

在 MBR 系统中, 我们可以利用污泥表观产率系数 Y_o 的大小来衡量系统中的污泥量, 影响 Y_o 的主要因素有 X (微生物浓度), SRT (污泥龄), HRT (水力停留时间), $c_i - c_e$ (进出水 COD 浓度差) 和 $c_i - c_{sup}$ (进水与上清液 COD 浓度差)。在实验中, 为了计算方便我们用 Y_o^{-1} 来表示 Y_o 的逆。

我们创建图 3 所示的模型来预测污泥表观产率系数 Y_o 。首先生成 1000 条样本数据和理论的 Y_o^{-1} 来训练多元线性回归模型(Y_o^{-1} 的计算公式是如公式(4)所示), 把这 1000 条数据组成训练数据集矩阵 X 和矩阵 Y , 其中 1000 条序列 $(X, SRT, HRT, c_i - c_e, c_i - c_{sup})$ 组成的样本数据存放在矩阵 X 中, 1000 条 (Y_o^{-1}) 组成的样本数据存放在矩阵 Y 中, 然后利用最小二乘法计算出 $\hat{\beta}$, 最后再计算得出多元线性回归方程, 以此来创建图 3 所示的模型。

$$Y_o^{-1} = \frac{SRT}{X} \left[\frac{c_i - c_e}{HRT} - \frac{c_i - c_{sup}}{SRT} \right] \quad (4)$$

3.2. 实验分析

在实验中, 我们选取了某污水处理厂稳定状态下的数据和图 3 所示的模型来进行测试, 把用于测试的实验数据分组, 对每组实验数据分别用图 3 所示的模型来预测污泥表观产率系数并计算拟合度, 然后将预测结果和实际系统结果做对比, 对比结果如表 1 和图 4 所示, 观察实验结果表明该模型的精度较高。拟合度情况如图 5 所示, 研究发现该模型计算所得到的拟合度都很接近 1, 表明该预测模型的拟合度较高。

4. Haar 小波分析对多元线性回归算法的改进

4.1. 用 Haar 小波分析来优化多元线性回归算法

前面用到的多元线性回归算法需要生成一个训练数据集矩阵 X 来训练模型, 但这样得到的模型在做预测分析时误差稍大, 这是因为训练数据集矩阵 X 中的数据分布不均匀, 所以我们要对训练数据集矩阵 X 做去噪预处理, 使数据分布均匀, 以此来达到提高模型泛化性和减小误差的目的。

对训练数据集矩阵 X 去噪有三个步骤, 第一步对矩阵 X 进行 Haar 小波分解, 在这里我们只分解一层就够了; 第二步对分解后的各分量数据做阈值量化处理(也就是去噪处理), 常用的阈值函数有硬阈值和软阈值函数, 这里我选用硬阈值函数($\eta(w) = wI(|w| > T)$); 第三步对处理后的各小波分量进行 Haar 小波重构, 重新获得训练数据集矩阵 X' , 这样就使得训练数据集矩阵 X' 中的数据分布更加均匀了, 然后再利用 X' 来做多元线性回归算法, 在实验中我们利用图 6 所示的模型来预测 Y_o^{-1} 。

4.2. 实验分析

在实验中, 我们首先建立图 6 所示的模型, 然后选取某污水处理厂稳定状态下的数据并利用该模型

进行实验。经过实验测得的拟合度情况如图 7 所示，对比图 5 我们发现改进后的多元线性回归算法的拟合度都在 0.9 以上，比改进前的拟合度更高，误差更小，达到了利用 Haar 小波分析改进的多元线性回归

Table 1. Table of apparent yield coefficient of sludge under steady state (30 degrees Celsius)
表 1. 稳定状态下污泥表观产率系数表(30°C)

X	SRT	HRT	$c_i - c_e$	$c_i - c_{sup}$	Y_0^{-1} 实际值	Y_0^{-1} 预测值
2432.8	9.4	5.2	269.6	293.3	4.9195	4.9904
2042.1	8.5	6	246.1	248.6	4.2727	4.1097
2433.8	6.9	6.8	216.9	295.1	2.2947	2.2890
2005.9	5.4	6.3	248.7	235.8	2.6880	2.5248
2230.7	5.4	6.8	282.5	201.4	2.5034	2.5350

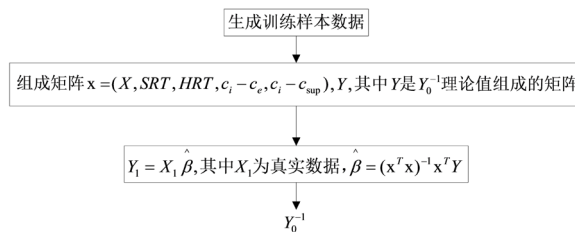


Figure 3. Sludge apparent yield coefficient prediction model
图 3. 污泥表观产率系数预测模型

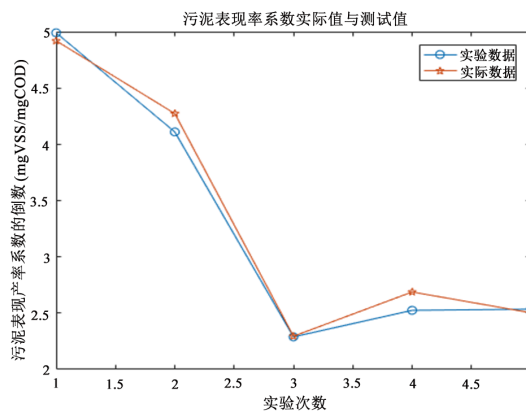


Figure 4. Relationship between actual value of sludge apparent yield coefficient and test value
图 4. 污泥表观产率系数的实际值与测试值关系图

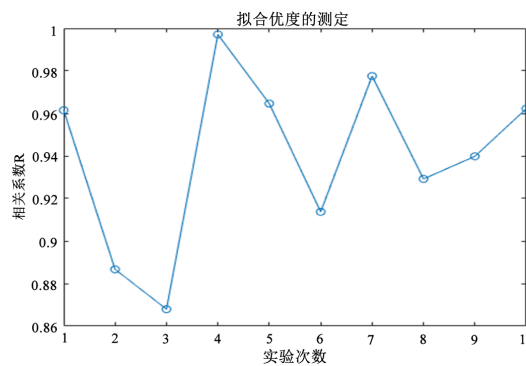


Figure 5. Fit test experiment
图 5. 拟合度测定实验图

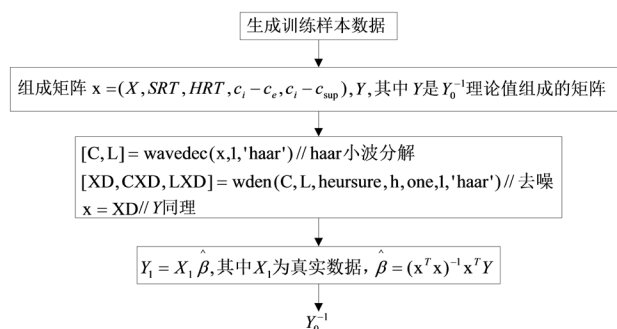


Figure 6. Improved prediction model of apparent yield coefficient of sludge

图 6. 改进后的污泥表观产率系数预测模型

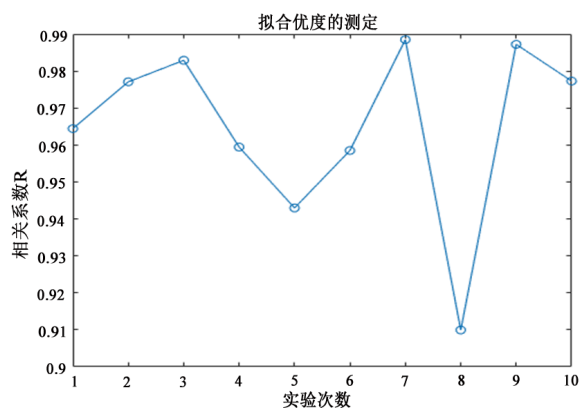


Figure 7. Improved fitness of the predicted model

图 7. 改进后预测模型的拟合度

算法对污泥表观产率系数做准确预测的目的。

5. 结论

本文讨论了在稳定状态下 Haar 小波分析改进的多元线性回归算法对污泥表观产率系数的预测, 在研究过程中我们分别利用多元线性回归算法和经 Haar 小波分析改进后的多元线性回归算法来预测污泥表观产率系数。图 5 所示的实验结果表明多元线性回归算法预测结果的误差并不是很大, 但是我们希望误差更小, 因此我们在多元线性回归算法中引入 Haar 小波分析, 研究结果表明经 Haar 小波分析改进后的多元线性回归算法预测结果的精度更高。将该算法应用在 MBR 膜污染仿真预测中, 可以快速准确的预测出污泥表观产率系数, 对 MBR 领域的研究具有一定的理论和实践意义。

基金项目

国家自然科学基金项目(51378350); 国家自然科学基金青年科学基金资助项目(50808130); 国家自然科学基金青年科学基金项目(21506159)。

参考文献

- [1] 计根良, 郑宏林, 周勇. MBR 系统运行条件对膜污染影响研究[J]. 水处理技术, 2014(6): 23-79.
- [2] 钟毓. 膜生物反应器工艺的优缺点及膜组件设计[J]. 中国环保产业, 2010(9): 50-51.
- [3] 韩永萍, 等. MBR 膜污染的形成及其影响因素研究进展[J]. 膜科学与技术, 2013(1): 102-110.
- [4] 李俊德. 多元线性回归的预测和控制[J]. 河南师范大学学报(自然科学版), 1986(3): 3-108.

-
- [5] 陶志强. 数理统计及计算机技术在我区铸造生产中的应用[J]. 曲靖师专学报, 1989(3): 31-88.
- [6] 陈永胜, 宋立新. 多元线性回归建模以及 SPSS 软件求解[J]. 通化师范学院学报, 2007(12): 77-112.
- [7] 李世雄. 小波变换及其应用(续三) [J]. 高等数学研究, 2003(2): 44-91.
- [8] 覃太贵. 快速小波变换在信号去噪中的应用[J]. 三峡大学学报(自然科学版), 2004(5): 89-143.
- [9] Dai, T. (2014) Wavelet Analysis of One Order Linear Stochastic System. *Journal of Mathematical and Computational Science*, **1**, 21-57.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org