

基于 WSVM 的软测量算法应用研究^{*}

鞠 旋, 刘春波, 潘 丰^{*}

(江南大学 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122)

摘要: 为了克服污水处理过程中随机干扰严重影响软测量效果的不足, 基于支持向量机(SVM)核方法和小波多分辨率理论, 提出了一种多分辨率小波核支持向量机的机器学习方法, 并应用于对污水处理过程水质参数的软测量中。利用支持向量回归机与参数特性混合建模, 实现了对 BOD 浓度与 COD 浓度的在线软测量。仿真结果表明: 应用该软测量技术能较好地克服污水处理过程中由随机干扰、强非线性、大时变、严重滞后等因素带来的一系列问题, 为基于软仪表的污水处理过程的参数在线检测和自动控制创造了条件。

关键词: 软测量; 小波核函数; 支持向量机; 污水处理

中图分类号: TP273; TH7

文献标志码: A

文章编号: 1001-4551(2011)04-0476-03

Soft sensor study and application based on WSVM

JU Xuan, LIU Chun-bo, PAN Feng

(Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education,
Jiangnan University, Wuxi 214122, China)

Abstract: In order to solve the problem associated with random disturbance in the process of soft sensor on wastewater disposal process, a new method called support vector machine(SVM) with multi-resolution wavelet kernel function was presented. A soft measure of BOD & COD on soft sensors model of wastewater treatment process control parameters was achieved by using a rare way of building a mixed model based on SVM and parameters' traits. The experimental results indicate that the soft measure can overcome a series of problems through the process of sewage disposal, such as random interfere, strong nonlinearity, big time change, serious lag, and so on. Therefore, this method will create conditions for getting parameters on-line and make wastewater disposal automatically.

Key words: soft measure; wavelet kernel function; support vector machine(SVM); sewage disposal

0 引 言

支持向量机最初是由 Vapnik^[1]提出的一种新兴的基于统计学习理论的学习机。支持向量机基本思想是把训练数据集从输入空间非线性地映射到一个高维特征空间(Hilbert 空间), 然后在此空间中求解凸优化问题(典型二次规划问题)。SVMs 的最优求解基于结构风险最小化思想, 因此比其他非线性函数逼近方法具有更强的泛化能力。

笔者在研究 SVM、软测量技术的基础上^[2-4], 提出

利用小波支持向量回归机与参数特性混合建模的方法, 实现了对污水处理过程中 BOD 浓度与 COD 浓度的软测量。

1 软测量技术的基本原理

软测量技术是依据某种最优化准则, 利用由辅助变量构成的可测信息通过软件计算实现对主导变量的测量^[5]。软仪表的核心是表征辅助变量和主导变量之间的数学关系的软测量模型。软测量技术原理基本框架图如图 1 所示。因此构造软仪表的本质就是如何

收稿日期: 2011-01-04

基金项目: 国家高技术研究发展计划(“863”计划)资助项目(2006AA020301-11)

作者简介: 鞠 旋(1983-), 男, 山东潍坊人, 主要从事控制理论与控制工程方面的研究. E-mail: juxuan_83@163.com

通信联系人: 潘 丰, 男, 教授, 博士生导师. E-mail: pan_feng_63@163.com

建立软测量模型,即数学建模问题。软测量模型注重的是通过辅助变量来获得对主导变量的最佳估计,而不是强调过程各输入/输出变量彼此之间的关系。软测量模型本质上是要完成由辅助变量构成的可测信息集 θ 到主导变量估计 Y 的映射,用数学公式表示即为: $Y=f(\theta)$ 。

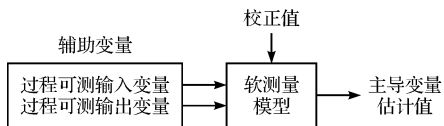


图1 软测量技术原理基本框架图

2 基于小波核函数的支持向量机(WSVM)

支持向量机的非线性处理能力都是通过“核映射”的方法实现的^[6]。已有的如 RBF 函数就是一种普遍使用的核函数,SVM 在软测量中表现出了良好的映像性能^[7-9]。但在应用中发现,对于目前已有的核函数,支持向量机不可能逼近某一 $L^2(\mathbf{R})$ (平方可积空间)子空间上的任意函数,因为已有的核函数通过平移不可能生成该子空间上的一组完备的基。这种基的不完备性导致了回归支持向量机不能逼近任意的目标函数。基于以上考虑,需要寻找一种核函数,这种核函数仅通过平移伸缩便能生成 $L^2(\mathbf{R})$ 空间上的一组完备的基。这样的函数在小波分解理论中已给出,就是小波函数。

小波函数 $\varphi(x)$ 通常由尺度函数 $\phi(x)$ 生成,它们之间满足如下的尺度方程:

$$\phi(x) = \sqrt{2} \sum_k h_k \phi(2x - k) \quad (1)$$

$$\varphi(x) = \sqrt{2} \sum_k g_k \phi(2x - k) \quad (2)$$

式中: $\{h_k\}$ 和 $\{g_k\}$ ——对离散的滤波器。

且它们之间存在下述的关系:

$$g_k = (-1)^{k-1} \bar{h}_{(k-1)} \quad (3)$$

式中: \bar{h} — h 的共轭。

参考小波分解理论和支持向量机核函数的构造要求^[10-12],经过分析这里选择的母小波函数为 Morlet 小波。因为该母小波函数是复数形式,取实数 Morlet 小波函数为:

$$\varphi(x) = \cos(\omega_0 x) e^{-x^2/2} \quad (4)$$

所以这里小波核函数的形式为:

$$k(x, x') = \prod_{i=1}^d \varphi\left(\frac{x_i - x'_i}{a_i}\right) = \prod_{i=1}^d \cos\left(\omega_0\left(\frac{x_i - x'_i}{a_i}\right)\right) \exp\left(-\frac{(x_i - x'_i)^2}{2a_i^2}\right) \quad (5)$$

3 函数仿真研究

已知函数 $\sin c(x - 10)$ 在均匀分布点 $x_k = -10 + 40/165 \cdot k, k=0, 1, \dots, 165$ 上的函数值为 $y = \sin c(x_k), k=0, 1, \dots, 165$ 。设总集合为: $T = \{(x_0, y_0), (x_1, y_1), \dots, (x_{165}, y_{165})\}$ 是根据单变量函数 $\sin c(x - 10)$ 在受到噪声的干扰下产生的,也就是说 x_0, x_1, \dots, x_{165} 是在 $[-10, 30]$ 上均匀分布的 166 个点,而 $y_k = \sin c(x_k) + \xi_k, k=0, 1, \dots, 165$ 。其中噪声 ξ_k 服从正态分布,且 $E\xi_k^2 = \sigma^2, E\xi_i = 0$, 这里 $\sigma^2 = 0.3, \omega_0 = 5$ 。实验中小波核 SVM 的平均预测误差为 1.60%, RBF 核 SVM 的平均预测误差为 3.04%。仿真图形如图 2 所示。

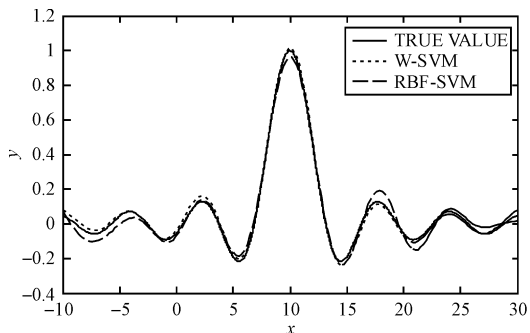


图2 小波核函数(W-SVM)与RBF核函数仿真结果

仿真结果表明,本研究构造的核函数支持向量机模型方法不但逼近精度高,而且泛化能力强。这说明了式(5)核函数的有效性和可行性。

4 WSVM 在污水处理软测量中的应用

工业污水由于成分复杂,有害污染物多,处理难度大,一直是企业发展壮大的瓶颈。由于污水处理过程的生产条件恶劣,随机干扰严重,具有强非线性、时变、大滞后等特点,建立的传统数学模型对水质关键参数在线监测效果不理想^[13-14]。本研究利用小波支持向量回归机与参数特性混合建模的方法,从而实现对 BOD 浓度与 COD 浓度的软测量。

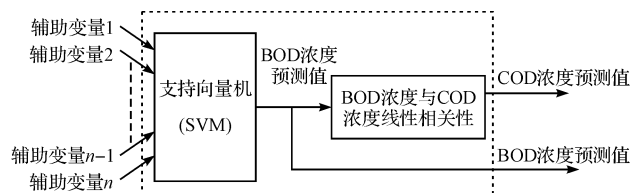


图3 基于SVM的软测量模型结构图

经过分析后,笔者选择进水的 COD 浓度、温度 T 、PH 值、流量 Q 以及污泥浓度 SS 、好氧反应区溶解氧 DO 浓度 6 个对污水处理效果有重要影响的变量作为模型的辅助变量,用来对出水水质参数 BOD 浓度、

COD 浓度在线预估^[15],软测量模型结构如图 3 所示。

根据厂方提供的系统正常运行时采集的 200 组数据,本研究选择其中进水的 COD 浓度、温度 T 、 PH 值、流量 Q 、污泥浓度 SS 、好氧反应区溶解氧 DO 浓度数据组成模型输入向量 X ,出水的 BOD 浓度和 COD 浓度作为模型的输出向量 Y_1 和 Y_2 ,即:

$$X = [PH, T, Q, DO, COD, SS] \quad (6)$$

$$Y_1 = [BOD] \quad (7)$$

$$Y_2 = [COD] \quad (8)$$

模型的输入向量 X 实际上是一个 200×6 的矩阵,输出向量 Y_1 和 Y_2 都是 200×1 的矩阵。且两个输出向量之间满足理论线性关系,即:

$$Y_2 = k \cdot Y_1 + b \quad (9)$$

在确定了重要辅助变量后,需要对其进行预处理和尺度变换。尺度变换常常将它们变换为 $[-1, 1]$ 或 $[0, 1]$ 的范围。本设计利用公式:

$$D'_i = \frac{D_i - D_{\min}}{D_{\max} - D_{\min}} \quad (10)$$

模型将数据变化为 $[0, 1]$ 范围内,其中 D'_i 为变换后得到的归一化值, D_i 为实际值, D_{\min} 为实际值最小值, D_{\max} 为实际值最大值。

这些工作将使得数据对于支持向量机更容易学习和训练。经过小波支持向量机处理后,将得到的数据经过反归一化处理(如下式所示)后,就得到了真正的处理结果。

$$D_i = D'_i \cdot (D_{\max} - D_{\min}) + D_{\min} \quad (11)$$

随机选择处理好的 200 组数据中的 150 组作为训练样本,剩余的 50 组数据作为测试样本。利用训练样本训练出一个计算速度快、拟合效果比较好的支持向量机模型后,预估 BOD 的浓度。

相关性的计算公式如下所示:

$$M_i = a \cdot N_i + b \quad (12)$$

$$\Delta = \min \|M_i - M_j\| (i \neq j) \quad (13)$$

$$\frac{\partial \Delta}{\partial a} = \sum_{i=1}^n (a \cdot x_i + b - y_i) \cdot x_i = 0 \quad (14)$$

$$\frac{\partial \Delta}{\partial b} = \sum_{i=1}^n (a \cdot x_i + b - y_i) = 0 \quad (15)$$

$$a = \frac{\sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i / n}{\sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i / n} \quad (16)$$

$$b = \frac{\sum_{i=1}^n y_i \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \cdot y_i \cdot \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i / n} \quad (17)$$

将训练样本出水端 BOD、COD 的 150 组数据用于计算相关性公式(12)的系数 a 与 b ,其中 M 代表 COD 数据矩阵, N 代表 BOD 数据矩阵。根据最小二乘法原则,使得公式(13)的 Δ 最小,求解偏微分方程式(14)、式(15)得到系数 a 、 b 的计算式(16)和式(17)。

根据 BOD 浓度与 COD 浓度的线性相关性,预估出 COD 浓度,完成混合软测量模型的建立。最后用 50 组测试样本,测试建立的混合模型的泛化性和预测精度。预估 BOD 浓度、COD 浓度测试效果如图 4、图 5 所示。

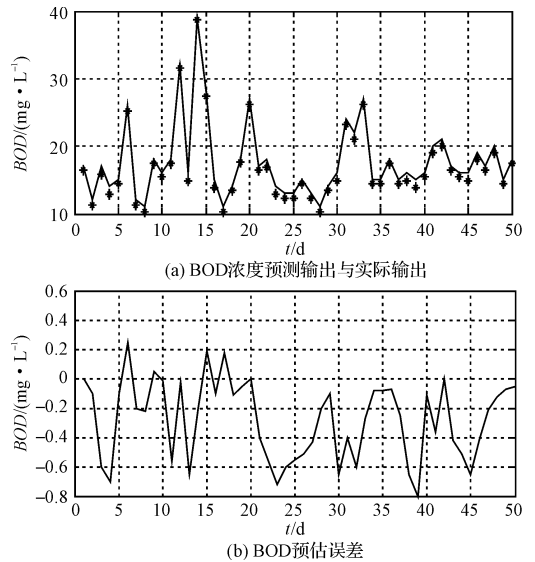


图 4 预估 BOD 浓度测试效果

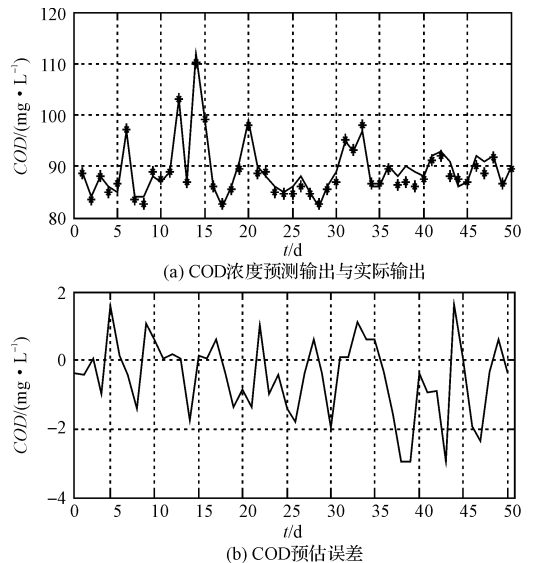


图 5 利用 BOD 与 COD 相关性预估 COD 浓度测试效果

从图中可以看出,该软测量模型在误差范围之内很好的完成了对 BOD 浓度、COD 浓度的预测。BOD 预测的最大相对误差是 4.81%,预测最小相对误差 0%,预测平均相对误差 3.32%;COD 预测的最大相对误差是 3.93%,预测最小相对误差 0%,预测平均相对误差 2.33%。

(下转第 492 页)

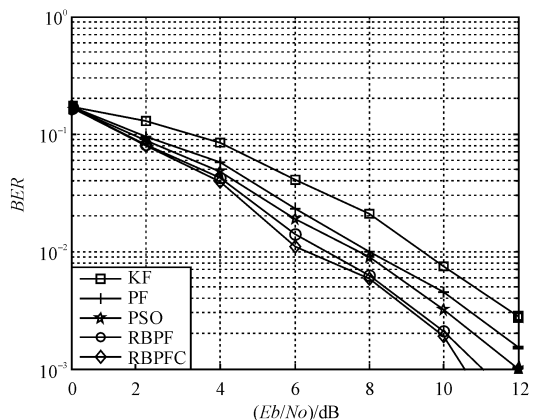


图 3 $M_r = 3$ 时误比特率曲线图

3 结束语

在本研究中,笔者探讨了多径信道数目和状态都为时变的 MIMO-OFDM 信道估计,由于传统的信道估计算法假设多径分量的数目为常数,对时变信道估计性能不佳。本研究通过应用随机集理论,构造了多径分量数目时变的信道系统,不过,虽然本研究对通信中信道估计的随机集理论应用进行了研究,但是还有许多其他的通信信号处理问题尚未解决。

参考文献 (References) :

[1] ANGELOSANTE E B, MARCO L. Sequential estimation of

OFDM channel[J]. **IEEE Proceeding of Signal Processing**,2007,53(8):3238-3243.

[2] ANGELOSANTE E B, MARCO L. Sequential estimation of multipath MIMO-OFDM channels[J]. **IEEE Transactions on Signal Processing**,2009,53(8):3167-3181.

[3] ANGELOSANTE E B, MARCO L. Multiuser detection in dynamic environment part I : user identification and data detection[J]. **IEEE Transactions on Information Theory**,2007,53(9):3158-3170.

[4] ANGELOSANTE E B, MARCO L. Multiuser detection in dynamic environment part II : joint user identification and parameter estimation[J]. **IEEE Transactions on Information Theory**,2009,53(5):2365-2374.

[5] ANGELOSANTE E B, MARCO L. Some applications of FISST to wireless communications [C]//11th International Conference on Information Fusion Proceeding,2008:1-6.

[6] KHAIRY EI B, ALNEYADI H M. Comparison of the behavior of MMSE detection scheme for DS-CDMA and MC-CDMA [C]//Second IFTP International Conference,2005:490-495.

[7] BARHUMI I, LEUS G, MOONEN M. Optimal training design for MIMO-OFDM systems in mobile wireless channels [J]. **IEEE Transactions on Signal Processing**,2003,51(6):1615-1624.

[8] MAHLER R. Statistical Multisource Multi-target Information Fusion. Boston[M]. MA:Artech House,2007.

[9] REYNOLDS C W. Flocks herds and schools; a distributed behavioral model[J]. **Computer Graphics**,1987,21(4):131-139.

[10] SHI Y, EBERHART R C. Parameter selection in particle swarm optimization [C]//Proceedings of the 7th Annual Conference on Evolution Computation,1998:591-601.

[编辑:张 翔]

(上接第 478 页)

5 结束语

在工业控制过程中,很多重要但却无法直接测量的变量可以通过软测量技术实现对其的监控。软测量技术需要借助大量与主导变量有关的辅助变量,软测量技术的关键在于建立合适的数学模型。笔者在对核函数研究的基础上,基于支持向量机的条件和小波多分辨率分解技术,利用 Morlet 小波函数构造出再生核函数,并将这种核函数用于一维函数和污水处理实例的仿真研究中。该软测量模型可对出水水质实现预估,从而可加强测量的实时性,为在线控制污水在 UASB 反应器、AA/O 好氧处理区的停留时间提供一定的依据。

参考文献 (References) :

[1] VAPNIK V N. The Nature of Statistical Learning Theory [M]. 2nd ed. New York: Springer,1999.

[2] 李海青,黄志尧. 软测量技术原理及应用[M]. 北京:化学工业出版社,2000.

[3] 邓乃扬,田英杰. 数据挖掘中的新方法—支持向量机 [M]. 北京:科学出版社,2004.

[4] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000,126(11):6-10.

[5] 朱学峰. 软测量技术及其应用[J]. 华南理工大学学报:自然科学版,2002,30(11):11-18.

[6] [英]克里斯特安尼. 支持向量机导论[M]. 李国正,译. 北京:电子工业出版社,2004.

[7] FENG Rui, ZHANG Yan-zhu, SHAO Hui-he. A novel soft sensor modeling approach based on least squares support machines[J]. **High Technology Letters**,2004,10(4):39-42.

[8] 俞金寿. 基于支持向量机的数据建模在软测量中的理论研究[J]. 华东理工大学学报,2003,28(1):8-10.

[9] 李东晖,杜树新,吴铁军. 监控污水处理过程运行状态的加权支持向量机方法[J]. 化工自动化及仪表,2005,32(1):18-20.

[10] MICCHELLI C, PONTIL M. Learning the kernel function via regularization[J]. **Journal of Machine Learning Research**,2005(6):1099-1125.

[11] SMOLA B S A J. Learning with Kernels[M]. Cambridge, MA: The MIT Press,2001.

[12] WILLIAMS P, LI S, FENG J F. Scaling the kernel function to improve performance of the support vector machine [C]//Lecture Notes in Computer Science, Advances in Neural Net-works: Proceedings Second International Symposium on Neural Networks,2005:831-836.

[13] 周正立,张 悦. 污水生物处理应用技术与工程实例 [M]. 北京:化学工业出版社,2006.

[14] 高艳玲. 污水处理生物新技术[M]. 北京:中国建筑工业出版社,2006.

[15] 廖青州,宁 艳. BOD 与 COD 的理论相关性及其应用 [J]. 广西化工,2006,29(12):19-25. [编辑:张 翔]