doi:10.3969/j.issn.1673-9833.2015.02.017

PSO-LIBSVM在污水水质建模中的应用

刘 帮，秦 斌，王 欣，朱万力

（湖南工业大学 电气与信息工程学院，湖南 株洲 412007）

摘 要：针对间歇式活性污泥法（SBR）复杂非线性等问题，常规神经网络建立的出水水质模型性能精度不高。采用支持向量机建立生化需氧量（BOD）软测量模型，并通过粒子群算法弥补支持向量机模型参数的不足。仿真结果表明，相对于BP神经网络、标准SVM模型，PSO-LIBSVM模型的误差小、精度高，降低了模型的复杂度并提高了其泛化能力，能达到较好的预测效果

关键词：LIBSVM；生化需氧量；支持向量机；粒子群算法

中图分类号：X703 **文献标志码**：A **文章编号**：1673-9833(2015)02-0089-05

Application of PSO-LIBSVM in Modeling of Sewage Water Quality

Liu Bang，Qin Bin，Wang Xin，Zhu Wanli

（School of Electrical and Information Engineering，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：Aiming at complex nonlinear problems in an sequencing batch type activated sludge process (SBR) and poor precision of sewage water quality model established by conventional neural network, applies an support vector machine to set up BOD soft measurement model, and improves the SVM parameter through particle swarm optimization. The simulation results show that compared with the BP neural network and standard SVM model, the PSO-LIBSVM has small error and high precision. It decreases the model complexity, improves its generalization ability, and achieves good prediction effect.

Keywords：LIBSVM；biochemical oxygen demand；support vector machine；particle swarm optimization

0 引言

水与人们的生活息息相关，是人类赖以生存的根本。随着经济的发展和城市化进程的加快，各种污水的排放量日趋加大，水环境污染不断加剧，给人们的身体健康带来非常严重的影响，造成了生态环境的恶化[1]。污水处理厂为治理水环境污染起到了一定的作用，但由于污水处理系统是一个高度非线性、强耦合、多变量和大滞后的复杂系统，其机理研究还不够成熟，关键水质参数不能实现在线测量，而污水处理效果的好坏依赖于对污水指标的精确测量，因此很难实现系统的闭环控制。对于一些重要水质测量指标[2-3]，如生化需氧量（biochemical oxygen demand，BOD）浓度，缺少成熟且经济的在线测量仪器。因此，对关键水质参数进行在线测量与优化控制变得十分迫切。文献[4]通过对污水处理工艺的机理模型进行非线性误差补偿分析，建立了基于BP神经网络的BOD软测量模型，能够较准确地对BOD参数进行估计。文献[5]采用3层前馈神经网络，设计了一种软硬件结合的水质参数软测量仪表，能实现对

收稿日期：2015-02-03

基金项目：国家自然科学基金资助项目（61074067，21106036），湖南省科技计划基金重点资助项目（2014FJ2018），湖南省自然科学基金资助项目（13JJ3110）

作者简介：刘 帮（1988-），男，湖南岳阳人，湖南工业大学硕士生，主要研究方向为复杂过程建模，集成优化控制，E-mail：476694465@qq.com

BOD水质参数的快速检测。由于神经网络固有的缺陷，如容易陷入局部极小、推广能力差等，导致其在实际应用中受到一定的约束。

支持向量机（support vector machine，SVM）是基于结构风险最小化原理，能使在小样本条件下的模型具有全局最优、最大泛化和推广能力；对于需要同时考虑诸多因素和条件的实际复杂问题有较强的适应性；相对于神经网络结构性的缺点，更能得到广泛地应用。但支持向量机存在核函数及其参数选择的问题[6]。

本文在现有研究成果的基础上，设计了一种粒子群优化：LIBSVM参数的BOD预测模型，能解决学习参数在一定程度上影响模型泛化能力的问题。通过建模后的仿真结果表明，该模型具有较好的BOD预估效果，推广性较强。

1 支持向量回归机原理

回归问题的描述为：假设存在 L 个观测点，$T=\{(x_1,y_1),\cdots,(x_i,y_i),\cdots,(x_L,y_L)\}$ 独立同分布，其中 $x_i\in\mathbf{R}^d$，$y_i\in\mathbf{R}$，$i=1,2,\cdots,L$，是按照某个未知的 $\mathbf{R}^d\times\mathbf{R}$ 上的概率分布所产生。回归问题就是要找到一个最优决策函数 f 使其满足上面的假设，并且当有新的输入点时，可根据决策函数来推断其对应的输出。在SVM中引入不敏感损失函数，对应的回归问题可描述为

$$f(x_1)=\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x}+b, \tag{1}$$

式中：$\boldsymbol{w}\in\mathbf{R}^d$ 为权重向量；$b\in\mathbf{R}$ 为偏置项。

由于函数 f 拟合出的值与实际值之间存在误差 ε，并考虑到函数 f 拟合误差超过 ε 时的真实损失，引入松弛变量 ζ,ζ'，其优化问题为

$$\min\left(\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{L}(\zeta_i+\zeta_i')\right);$$
$$\text{s.t.}\begin{cases}y_i-f(x_i)\leqslant\varepsilon+\zeta_i,\\ f(x_t)-y_t\geqslant\varepsilon+\zeta_t',\\ \zeta_i,\zeta_i'\geqslant0,i=1,2,\cdots,L。\end{cases} \tag{2}$$

式中：$\|\boldsymbol{w}\|^2$ 表示置信范围；

$\sum_{i=1}^{L}(\zeta_i+\zeta_i')$ 表示经验风险；

C 表示正则化参数。

$\|\boldsymbol{w}\|^2$ 的作用是使拟合函数更平滑，加强拟合函数的推广能力；$\sum_{i=1}^{L}(\zeta_i+\zeta_i')$ 的作用是减小训练误差；C 的作用是控制对错分样本的惩罚程度。

为导出式（2）的对偶形式，并求解凸二次规划问题，引入拉格朗日（Lagrange）函数，则式（2）的对偶优化问题可描述为[7]

$$\min\left(\frac{1}{2}\sum_{i,j=1}^{L}(\alpha_i-\alpha_{i_i}^*)(\alpha_j-\alpha_{j_i}^*)\phi(x_i)\cdot\phi(x_j)+\varepsilon\sum_{i=1}^{L}(\alpha_i+\alpha_{i_i}^*)-y_i\sum_{i=1}^{L}(\alpha_i-\alpha_i^*)\right);$$
$$\text{s.t.}\begin{cases}\sum_{i=1}^{L}(\alpha_i-\alpha_{i_i}^*)=0,\\ 0\leqslant\alpha_i,\alpha_{i_i}^*\leqslant C,i=1,2,\cdots,L。\end{cases} \tag{3}$$

式中：a_i, a_i^* 为Lagrange乘子；ϕ 为非线性映射。

若直接对式（3）在特征空间进行回归，则会由于非线性映射 ϕ 的形式和参数不确定，及特征空间 Ω 的高维数，使 $\boldsymbol{w}$ 无法显式地表达。通过引入核函数，使原始输入向量隐式地向高维特征空间转换，由式（1）~（3）得表达式

$$\min\left(\frac{1}{2}\sum_{i,j=1}^{L}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x)+\varepsilon\sum_{i=1}^{L}(\alpha_i+\alpha_i^*)-\sum_{i=1}^{L}y_i(\alpha_i-\alpha_i^*)\right)。 \tag{4}$$

由式（3）计算得出最终回归估计函数为

$$f(x)=\sum_{x_i\in SV}(\alpha_i-\alpha_i^*)K(x_i,x)+b, \tag{5}$$

式中：SV 为训练样本集对应的支持向量的集合；

$K(x_i,x)$ 为核函数；

$$b=\frac{1}{L_{SV}}\left\{\sum_{0<\alpha_i<C}\left[y_i-\sum_{x_i\in SV}(\alpha_i-\alpha_i^*)K(x_i,x)-\varepsilon\right]+\sum_{0<\alpha_i^*<C}\left[y_i-\sum_{x_i\in SV}(\alpha_i-\alpha_i^*)K(x_i,x)+\varepsilon\right]\right., \tag{6}$$

其中 L_{SV} 为支持向量集所含元素的个数。

2 PSO 基本原理

支持向量机模型的性能取决于选择合适的核函数类型、核函数参数、惩罚系数等。因此，为获得高效的预测估计模型，需要对这些参数进行合理选择，确定最优参数。

粒子群优化算法（particle swarm optimization，PSO）[8]是一种基于群体协作的随机搜索算法，通过群体中个体间的合作与信息共享来指导完成寻优。PSO算法具有编程方便，结构简单，易于实现，搜索速度快，收敛能力强等特点，通常被用于在复杂环境中求解最优问题。PSO算法求解最优问题时，将每一只飞行的鸟都当作优化问题的一个潜在解，食物

的位置则对应于优化问题的全局最优解。在搜索空间中，这些鸟称为“粒子”，每个粒子都有自己的位置和速度两个特征，分别用$X_{id}=(x_{i1},x_{i2},\cdots,x_{id})$和$V_{id}=(v_{i1},v_{i2},\cdots,v_{id})$表示，其中$i$=1, 2, …，$n$，$n$为种群数量，$d$为搜索空间的维数。粒子在搜索空间中的位置坐标所对应的目标函数值为粒子的适应度值。$P_i=(p_{i1},p_{i2},\cdots,p_{id})$表示第$i$个粒子经历的最好位置；$P_g=(p_{g1},p_{g2},\cdots,p_{gd})$为群体经历的最好位置。在每次迭代过程中，速度和位置更新公式如下：

$$v_{id}^{k+1}=w*v_{id}^{k}+c_1*rand_1^k()*\left(p_{id}^{k}-x_{id}^{k}\right)+c_2*rand_2^k()*\left(p_{gd}^{k}-x_{gd}^{k}\right),\quad(7)$$

$$x_{id}^{k+1}=x_{id}^{k}+v_{id}^{k+1}。\quad(8)$$

式（7）~（8）中：w为惯性权值；

c_1, c_2为加速常数；

$rand_1$, $rand_2$为区间[0,1]上的随机分布函数。

为了防止粒子在寻优中脱离搜索空间的可能性，通常限定$v_{id}\in[-v_{\max},v_{\max}]$，其中$v_{\max}$为粒子最大飞行速度。$w$起权衡全局与局部搜索的作用，为使算法的收敛速度加快，对w进行处理，即随着迭代的进行，线性减少w的值[9]，且有

$$w=w_{\max}-iter*(w_{\max}-w_{\min})/iter_{\max},\quad(9)$$

式中：$iter$, $iter_{\max}$分别为当前和最大迭代次数；

$w_{\max}$, $w_{\min}$分别为最大和最小惯性权值。

3　基于PSO-LIBSVM生化需氧量建模

3.1　建模过程

由于径向基核函数（radial basis function，RBF）学习效果好，模型精度高。因此，本文采用RBF作为LIBSVM的核函数，其表达式为

$$K(x_i,x)=\exp\left\{-\frac{\|x-x_i\|^2}{2\sigma^2}\right\},\quad(10)$$

式中σ为核函数参数。

本文对参数C和σ同时进行优化。

PSO-LIBSVM预测模型的算法步骤如下：

Step1　通过经验公式（式（11））明确C, σ的搜索区间；初始化种群，包括学习因子、种群规模、迭代次数的初始化；随机设置粒子的位置和速度；

$$C\in\left[d\cdot10^{-1.5},d\cdot10^{2.5}\right],\sigma\in\left[0.1\sqrt{d},\sqrt{d}\right],\quad(11)$$

式中d为搜索空间的维数。

Step2　根据目标函数计算所有粒子的适应度值，本文以SVM训练样本得到的模型作为适应度值。

Step3　将适应度值与每个粒子经历过的最好位置进行比较，如果更好，则全局最优位置P_g被当前粒子的最优位置所替代。

Step4　对每个粒子当前的个体位置P_i与全局位置P_g进行比较，如果更好，更新P_g。

Step5　按照式（7）和（8）更新粒子当前的速度和位置。

Step6　判断算法是否满足终止条件（目标函数误差达到预先设定的收敛精度，或者算法的循环次数达到用户设定的最大迭代次数）。若不满足，则返回Step2继续寻优；反之，转到Step7。

Step7　输出最优解C, σ并代入SVM模型，重新训练学习，得到较优的SVM预测模型。

根据上文描述的算法对BOD进行建模。通过对污水处理工艺及影响污水处理效果的因素进行分析，结合现场操作人员的建议，概括出12个过程参数作为模型的输入变量，主要包括：生化需氧量、曝气池酸碱度、曝气池溶解氧溶度、化学需氧量、总氮、氨氮、总磷、悬浮固体浓度、混合液固体浓度、氧化还原电位、进水水量、温度，出水BOD质量浓度作为模型输出。由于各变量有不同的工程单位，并且各变量的数量级不同，如果直接采用原始数据计算，会降低算法的精度并造成计算的不稳定。因此，采用归一化方法对数据进行预处理。选取污水处理厂提供150组历史数据，对数据样本进行预处理，随机划分100组学习样本和50组检测样本，建立基于PSO-LIBSVM生化需氧量的预估模型

3.2　仿真参数设置及结果

参数的设置分为2部分。第一部分是支持向量机参数设置：RBF核参数$\sigma\in$（0.1,1.5）；惩罚系数$C\in$（0.01,700）。第二部分为PSO算法参数设置：粒子维度(σ,C)，代表解空间为二维；最大迭代次数为50；种群数量为10；取c_1=1.7，c_2=1.5；惯性权值$w_{\max}$=0.9，$w_{\min}$=0.1。经过PSO寻优后的最佳参数为C=3.927 3，σ=2.236 1。

以平均相对误差和均方根误来评价模型性能的优劣，其表达式为

$$MAPE=\frac{1}{L}\sum_{i=1}^{L}\left|\frac{y_i-\hat{y}_i}{y_i}\right|,\quad(12)$$

$$RMSE=\sqrt{\frac{1}{L}\sum_{i=1}^{L}(y_i-\hat{y}_i)^2},\quad(13)$$

式（12）~（13）中：i为样本数；

y_i为BOD的实际值；

$\hat{y}_i$为BOD的测量值。

为了作进一步比较，分别采用BP神经网络和标准SVM算法对BOD进行建模仿真。其中标准SVM采

用的是RBF核函数；C, σ分别取值1和1.29；BP网络的参数为：输入神经元数为12，输出神经元数为1，隐层神经元数为13，训练函数取traingdx函数。

对应的仿真结果如图1~3，3种模型的预测性能见表1。

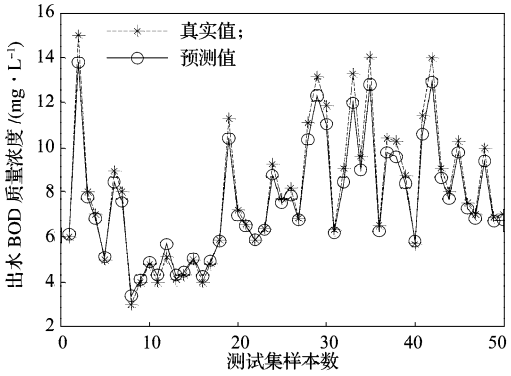

图1 PSO-LIBSVM模型预测输出

Fig. 1 The PSO-LIBSVM model prediction output

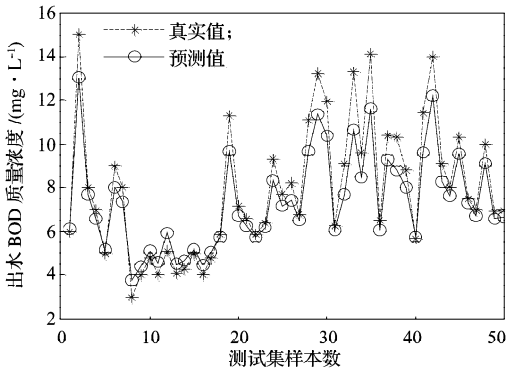

图2 标准SVM预测输出

Fig. 2 The standard SVM prediction output

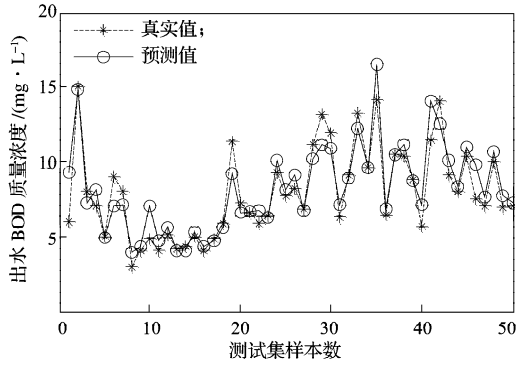

图3 BP神经网络预测输出

Fig. 3 The BPNN prediction output

表1 PSO-LIBSVM与标准SVM、BP网络预测性能比较

Table 1 The prediction performance of PSO-LIBSVM compared with standard SVM and BPNN

算法	指标		
	支持向量数	平均相对误差	均方根误差
PSO-LIBSVM	27	0.046 4	0.033 0
标准SVM	40	0.090 5	0.063 5
BP神经网络		0.110 9	0.107 2

通过分析图1~3和表1可知，在拟合精度方面，PSO-LIBSVM的平均相对误差和均方根误差最小，分别为0.046 4和0.033 0；标准SVM次之，BP神经网络的误差最大，因此PSO-LIBSVM拟合精度最高。在泛化性能方面，由于BP神经网络是基于经验风险最小化的结构，其泛化能力不及标准SVM和PSO-LIBSVM；而支持向量机的泛化性在一定程度上依赖其参数的选取，经过PSO优化过的LIBSVM比标准SVM使用了较少的支持向量，增强了支持向量机解的稀疏型，降低了模型的复杂度，因此PSO-LIBSVM泛化性最强。

4 结语

本文以为Matlab 2013a和Libsvm 3.1工具箱为平台，把粒子群算法和支持向量回归机相结合，建立了出水BOD的PSO-LIBSVM预测模型，并且与标准SVM模型、BP神经网络模型的预测效果进行了对比，从平均相对误差、均方根误差等几个性能指标进行了分析。结果表明，本文提出的模型预估效果最好，泛化性能最强，更符合实际问题的需求。

参考文献：

[1] 刘海云，张玉华，孙维锋. 鹤岗市城市污水处理发展现状及对策探讨[J]. 环境科学与管理，2007，32(7)：27-29.
Liu Haiyun，Zhang Yuhua，Sun Weifeng. The Dirty Water in City in Hegang City in Crane Handles Development Present Condition and Counterplan Study[J]. Environmental Science and Management，2007，32(7)：27-29.

[2] 徐方舟. 污水处理控制系统设计及其软测量的研究[D].无锡：江南大学，2011.
Xu Fangzhou. Design of Wastewater Treatment Control System and Research of Soft Sensing Technique[D]. Wuxi Jiangnan University，2011.

[3] 罗腾飞. 基于改进BP神经网络的污水处理出水指标预测[D]. 呼和浩特：内蒙古农业大学，2012.
Luo Tengfei. Predictive Indicators of Water in Waste Water Treatment Based on Improved BP Artificial Neural Network [D]. Hohhot: Inner Mongolia agricultural University，2012.

[4] 王小艺，刘载文，苏 震. 基于神经网络的污水BOD软测量补偿方法研究[J]. 系统仿真学报，2009，21(增刊2)：83-85.
Wang Xiaoyi，Liu Zaiwen，Su Zhen. Research on BOD Soft-Sensing Compensation Method Based on Neural Network [J]. Journal of System Simulation，2009，21(S2)：83-85.

[5] 乔俊飞，郭 楠，韩红桂. 基于神经网络的BOD参数软测量仪表的设计[J]. 计算机与应用化学，2013，30(10)：1219-1222.
Qiao Junfei，Guo Nan，Han Honggui. Design of Soft

Measurement Instrument for BOD Parameters Based on Neural Network[J]. Computers and Applied Chemistry，2013，30(10)：1219−1222.

[6] 付元元，任　东. 支持向量机中核函数及其参数选择研究[J]. 科技创新导报，2010 (9)：6−7.
Fu Yuanyuan，Ren Dong. Study on the Kernel Function and Its Parameter Selection of SVM[J]. Science and Technology Innovation Herald，2010 (9)：6−7.

[7] 周　威，金以慧. 利用模糊次梯度算法求解拉格朗日松弛对偶问题[J]. 控制与决策，2004，19(11)：1213−1217.
Zhou Wei，Jin Yihui. Fuzzy Subgradient Algorithm for Solving Lagrangian Relaxation Dual Problem[J]. Control and Decision，2004，19(11)：1213−1217.

[8] 董　芳. 粒子群算法研究及其在动态优化中的应用[D]. 杭州：浙江大学，2014.
Dong Fang. Researches on Particle Swarm Optimizer and Its Applications in Dynamic Optimization[D]. Hangzhou：Zhejiang University，2014.

[9] 冯辉宗，彭　丹，袁荣棣. 基于PSO-SVM的发动机故障诊断研究[J]. 计算机测量与控制，2014，22(2)：355−357.
Feng Huizong，Peng Dan，Yuan Rongdi. Research on Automobile Engine Fault Diagnosis Based on PSO-SVM [J]. Computer Measurement & Control，2014，22(2)：355−357.

（责任编辑：邓光辉）

（上接第63页）

Detection and Imaging. [S. l.]：International Society for Optics and Photonics，2013. doi：10.1117/12.2032938.

[13] 董　博，陶忠祥，苏伍各. 基于SIFT的红外与可见光图像配准方法[J]. 火力与指挥控制，2011，36(11)：168−171.
Dong Bo，Tao Zhongxiang，Su Wuge. IR and Visible Images Registration Approach Based on SIFT[J]. Fire Control and Command Control，2011，36(11)：168−171.

[14] Maes F，Collignon A，Vandermeulen D，et al. Multimodality Image Registration by Maximization of Mutual Information[J]. IEEE Transactions on Medical Imaging，1997，16(2)：187−198.

[15] 葛　杰，曹晨晨，李　光. 基于机器视觉的图像形状特征提取方法研究进展[J]. 包装学报，2015，7(1)：54−60.
Ge Jie，Cao Chenchen，Li Guang. Research Progress in Shape Feature Extraction Methods Based on Machine Vision [J]. Packaging Journal，2015，7(1)：54−60.

[16] 胡永祥，汤井田，蒋　鸿. 利用高维互信息的多模态医学图像配准[J]. 计算机工程与应用，2007，43(24)：242−245.
Hu Yongxiang，Tang Jingtian，Jiang Hong. Multi-Modality Medical Image Registration Using High Dimension Mutual Information[J]. Computer Engineering and Applications，2007，43(24)：242−245.

（责任编辑：邓　彬）