

相关向量机对废水处理系统出水水质的预测

刘鸿斌^{1,2} 宋留¹

(1. 南京林业大学江苏省林业资源高效加工利用协同创新中心, 江苏南京, 210037;

2. 华南理工大学制浆造纸工程国家重点实验室, 广东广州, 510640)

摘要: 准确预测出水水质对造纸废水处理过程具有重要意义, 为此笔者提出一种基于相关向量机(RVM)的软测量模型。首先, 利用偏最小二乘法(PLS)提取实际造纸废水处理过程数据的潜变量, 解决过程变量的共线性和高维度问题, 然后利用潜变量建立RVM预测模型。结果表明, 与RVM模型相比, 本文提出的PLS-RVM组合模型在对出水悬浮固形物(SS)的水质预测测试时, 均方根误差降低了7.76%, 决定系数提高了12.32%; 但对出水化学需氧量(COD)的预测测试效果提升并不明显。此外, PLS-RVM模型的预测效果较PLS-LSSVM模型有显著提升: 对出水SS的预测, 均方根误差降低了9.16%, 决定系数提高了15.29%; 对出水COD的预测结果中, 均方根误差降低了9.29%, 决定系数提高了18.34%。

关键词: 相关向量机; 降维方法; 支持向量机; 造纸废水处理; 软测量

中图分类号: TS736; X793

文献标识码: A

DOI: 10.11981/j.issn.1000-6842.2019.02.53

近年来, 软测量技术已成功地在精炼、化工、发电、食品加工、制浆造纸工业和城市与工业污染的监测等领域得到应用^[1]。软测量技术具备一系列优点: 替代昂贵的硬件设备, 降低生产成本; 良好的兼容性, 能与现有硬件传感器并行工作; 能够实时在线测量, 克服某些硬件传感器的滞后性, 可以实现更全面、及时的监测网络。在造纸废水处理的过程中, 一些难以测量或不易在线测量的重要参数, 如出水化学需氧量与出水悬浮固形物浓度等, 不仅是重要的出水指标, 也是必要的检测变量。为了提高造纸废水处理的达标率, 提升系统稳定性, 需要对这些重要的检测变量进行监测和控制, 但是传统的传感器测量方法普遍存在着价格昂贵、维护困难、稳定性差、使用寿命短等问题^[2]。针对这些问题, 软测量技术为实现污水处理过程出水水质的实时检测提供了一个很好的解决思路。

软测量技术的核心是建立输入变量与输出变量之间的软测量模型。其建模方法按照建模的机制划分为两大类^[3]: 一类是基于过程反应机理的机理模型, 该类模型通过分析过程的反应机理, 运用物理、化学和生物等基本定理来表述过程的内部规律, 建立过程模型, 但一般对象的反应机理都比较

复杂, 并且由于工况与环境的影响, 导致机理模型的预测性能大大降低; 另一类是数据驱动建模, 该方法是通过输入输出数据建立与过程外特性等价的模型, 因此不需要研究对象的内部规律, 只需获得足够多的数据即可建立对象的软测量模型。数据驱动建模有回归分析模型、人工智能模型、统计学习理论模型及概率核函数模型等软测量模型^[3]。回归分析建模方法主要有主元分析法(Principal Component Analysis, PCA)和偏最小二乘法(Partial Least Squares, PLS), 旨在寻找多个变量间的函数关系, 但局限于非线性较弱的对象。人工智能建模方法主要是人工神经网络, 汪瑶等^[4]采用人工神经网络建模预测出水水质, 但是神经网络易存在局部极小和泛化能力有限, 而且权值不易在线调整等缺点。黄银蓉等^[5]利用基于统计学习理论的最小二乘支持向量机(Least Square Support Vector Machine, LSSVM)对污水处理的出水指标进行建模预测, 但是支持向量机的核函数必须满足 Mercer 条件, 除核函数外, 惩罚因子与不敏感损失区域半径也会对模型的精度产生影响^[6]。另外, 支持向量基(SVM)的支持向量个数随着训练样本的增大呈线性增加, 可能会产生过拟合问题及增加计算时间。宋留等^[7]采用基于概

收稿日期: 2018-03-21

基金项目: 制浆造纸工程国家重点实验室开放基金资助项目(201813); 南京林业大学高层次人才科研启动基金(GX1029)。

作者简介: 刘鸿斌, 男, 1981年生; 副教授; 主要研究方向: 复杂工业过程的建模、监测和控制。

E-mail: hongbinliu@njfu.edu.cn

率核函数的高斯过程模型预测出水水质,但是高斯过程受限于先验高斯分布的假设条件,实际过程的分布难以预知或不一定遵循高斯分布,因此受到限制。

相关向量机(Relevance Vector Machine, RVM)是 Tipping^[8]提出的一种基于贝叶斯框架的稀疏概率模型,是一种新的有监督学习方法。与 SVM 相比,RVM 核函数的选择不受 Mercer 条件的限制,可以根据需要构建核函数,并且 RVM 是机率式预测的。此外相较于基于概率核函数的高斯过程模型,RVM 避免了高斯过程局限于先验高斯分布的假设条件,且同样能实现概率式预测。刘遵雄等^[9]在解决电力负荷中期预测的问题上,提出了一种基于 RVM 的中期预测方法。韩中合等^[6]把 RVM 回归模型运用到工业生产中短期风速预测的问题上,取得了良好的预测效果。许玉格等^[10]利用活性污泥水处理仿真基准模型(BSM1)^[11]得到的污水仿真数据,建立出水水质的 RVM 软测量预测模型,并与 LSSVM 和神经网络(BP)模型比较,取得了比较满意的污水处理出水水质预测效果。

然而对于机器学习的回归问题,随着变量数量的增多,变量之间的相关性将会增强,数据的可视化也会变得难以实现。变量的相关性或共线性会导致模型预测结果的方差变大,预测精度下降。在这种情况下,要很好地解释模型中自变量对因变量的效应,是非常困难的。PCA 与 PLS 作为进一步简化建模难度的有效工具,不仅能避免共线性的负面影响,而且能减少数据的维度^[1]。近年来,PLS 除作为化学计量学中的多元统计工具外,也从诸多降维方法中脱颖而出。PCA 仅对自变量进行主成分提取,从而忽略了自变量对因变量系统的解释性。PLS 降维的目的是找出与因变量有较大协方差的自变量的不相关线性转换,即潜变量(Latent Variables)^[12]。降维得到的潜变量,不仅对因变量具有最大解释性,而且能反映因变量系统的信息,使得自变量对因变量的解释效应在模型中得到体现。

提高废水处理过程出水水质模型的精度已经成为当前研究的热点。不同地区造纸厂所产生废水的污染成分及特性会随着原材料、生产工艺和废水处理技术等影响因素的不同而有较大差异,基于数据驱动的造纸废水处理预测模型可以为制浆造纸废水处理过程的优化控制提供重要支撑。本文采用广东东莞一家造纸厂的废水处理数据,首先使用 PLS 方法对原始过程变量进行去共线性和降维处理,然后利用建立在贝叶

斯框架下的 RVM 模型建立造纸废水处理系统的 PLS-RVM 软测量预测模型。

1 废水处理过程建模

1.1 降维方法

PLS 降维是将过程变量矩阵 $X \in R^{n \times m}$ 与质量变量矩阵 $Y \in R^{n \times p}$ 投影到由少量潜变量(或得分矩阵) $T = [t_1, \dots, t_K] \in R^{n \times K}$ 所构成的低维空间中:

$$\begin{aligned} X &= TP^T + E \\ Y &= TQ^T + F \end{aligned} \quad (1)$$

式中, $P = [p_1, \dots, p_K]$ 与 $Q = [q_1, \dots, q_K]$ 分别表示 X 与 Y 的负载矩阵, E 与 F 分别表示 X 与 Y 的残差矩阵。即 PLS 将过程变量分解为两个斜交投影的子空间:得分子空间(质量相关子空间)与残差子空间(质量无关子空间)。由于 PLS 是一个迭代的过程,潜变量 T 不能直接由原始过程变量矩阵 X 得到,因此引入权重矩阵 $V = [v_1, \dots, v_K] = W(P^T W)^{-1}$,使其满足^[13]:

$$T = XV \quad (2)$$

式中, $W = [w_1, \dots, w_K]$ 是计算得分矩阵的权重矩阵。

1.2 RVM 原理

类似于 SVM, RVM 模型的函数预测输出^[14]:

$$t_n = y(x_n; \mathbf{w}) + \varepsilon_n = \sum_{i=1}^N w_i K(x_n; x_i) + \varepsilon_n \quad (3)$$

式中, ε_n 是独立的零均值,方差为 σ^2 的高斯噪声。 $K(x_n, x_i)$ 表示核函数, $\mathbf{w} = (w_0, \dots, w_n)^T$ 表示参数向量。假定目标值 t 独立分布,则 RVM 的似然概率估计为:

$$\begin{aligned} P(t | x_n, \mathbf{w}, \sigma^2) &= \prod_{i=1}^N N(\mathbf{w}^T \phi(x_n), \sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2} \|t - \Phi\mathbf{w}\|^2\right\} \end{aligned} \quad (4)$$

其中, Φ 是训练样本中的特征向量 x_i 代入核函数中得到的结构矩阵:

$$\Phi = \begin{pmatrix} 1 & K(x_1, x_1) & \cdots & K(x_1, x_N) \\ 1 & K(x_2, x_1) & \cdots & K(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_N, x_1) & \cdots & K(x_N, x_N) \end{pmatrix} \quad (5)$$

为避免出现在 SVM 中的支持向量过多而导致的过拟合问题,在 RVM 模型中需对权值 w_i 进行约束,使得其机率分布是在零周围的正态分布:

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{i=0}^N N(w_i | 0, \alpha_i^{-1}) \\ = \prod_{i=0}^N \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i w_i^2}{2}\right) \quad (6)$$

式中, 每个超参数 α_i 与其相对应的权值 w_i 相关, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_N)^T$ 。根据贝叶斯概率框架及似然估计分布, 权重的后验概率分布为:

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-\frac{N+1}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{w} - \mathbf{m})^T \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \mathbf{m})\right\} \quad (7)$$

式中, 后验协方差矩阵 $\boldsymbol{\Sigma} = (\sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{A})^{-1}$, 均值 $\mathbf{m} = \sigma^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{t}$, 矩阵 $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ 。由于大多数的 α_i 都是趋近于无穷的, 与之对应的后验分布协方差与均值则无限趋近于零:

$$\lim_{\alpha_i \rightarrow \infty} \boldsymbol{\Sigma} = \lim_{\alpha_i \rightarrow \infty} (\mathbf{A} + \sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} = 0 \Rightarrow \lim_{\alpha_i \rightarrow \infty} \mathbf{m} = 0 \quad (8)$$

即与 α_i 相对应的权值 w_i 等于 0, 因此相应的基函数 $\phi(x_i)$ 将从结构矩阵 $\boldsymbol{\Phi}$ 去除, 而保留下的非零权重所对应的特征向量也就是相关向量。

$$\text{对超参数进行优化 } \alpha_{\text{new}} = \frac{1 - \alpha_i \sum_{\hat{u}} \hat{u}}{m_i^2} = \frac{\gamma_i}{m_i^2},$$

$$\sigma_{\text{new}}^2 = \frac{\|\mathbf{t} - \boldsymbol{\Phi} \mathbf{m}\|^2}{N - \sum_i \gamma_i}。因此, 对于新的输入值 x' ,$$

模型的预测分布:

$$P(t_{\text{new}} | x', \boldsymbol{\alpha}, \sigma^2) = N(\mathbf{m}^T \boldsymbol{\phi}(x'), \sigma^2(x')) \quad (9)$$

式中, 均值 $\mathbf{m}^T \boldsymbol{\phi}(x')$ 为要求的估计值 t_{new} , 即:

$$t_{\text{new}} = \mathbf{m}^T \boldsymbol{\phi}(x') \quad (10)$$

1.3 评价指标

本文采用均方根误差 (Root Mean Square Error, RMSE) 和决定系数 (Determination Coefficient, R^2) 作为模型的评价指标, 来比较不同模型的预测精度。 R^2 和 RMSE 的表达式为:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (12)$$

式中, \hat{y}_i 是估计值, y_i 是真实值, \bar{y} 是平均值。 R^2 表示回归平方和占总离差平方和的比重, 比重越大, 则回归效果越好。即 R^2 越接近 1, 代表输入变量对

输出变量的解释能力越强, 对数据拟合的效果也越好。RMSE 是预测数据和原始数据对应点误差的平方和的均值的平方根, RMSE 越接近于 0, 代表该模型预测效果越佳。

2 基于 RVM 废水处理过程仿真预测模型

2.1 造纸废水处理数据

废水数据采集自广东东莞的一家造纸厂废水处理车间, 测量数据展示了造纸废水处理的好氧段工况。数据包含 170 个样本点, 8 个废水变量, 分别是进水化学需氧量 (Chemical Oxygen Demand, COD)、出水 COD、进水悬浮固形物 (Suspended Solids, SS)、出水 SS、溶解氧量 (DO)、流量 (Q)、温度 (T)、pH 值。选取 120 个样本数据作为训练集, 50 个数据作为测试集, 在 MATLAB 中分析处理该 170 个造纸废水数据。其中, 进水 COD、进水 SS、DO、 Q 、 T 、pH 值作为输入变量, 出水 COD 和出水 SS 作为输出变量构建模型。

2.2 RVM 预测模型

为了验证 RVM 对出水 COD 与出水 SS 的预测性能, 用相同的数据进行 LSSVM 建模预测。LSSVM 的回归模型如下式所示:

$$y(x_n) = \sum_{i=1}^N \varphi_i K(x_n, x_i) + \beta \quad (13)$$

其中, $K(x_n, x_i)$ 表示核函数, 本文采用径向基核函数 $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$, σ 为核宽度; φ 与 β 利用 PLS 求得其近似值。

针对造纸废水处理过程中多变量、时变性、非线性和复杂性等特点, 本文选取高斯核函数作为 RVM 的核函数, 迭代次数 500。图 1 和图 2 分别表示 RVM 与 LSSVM 软测量模型对出水 COD 及出水 SS 的预测结果。在对出水 COD 预测时, RVM 模型的 RMSE 为 4.2658 mg/L, R^2 为 0.5684; 对出水 SS 预测时, RMSE 为 0.7565 mg/L, R^2 为 0.5478。与 LSSVM 模型相比, RVM 模型对出水 COD 预测结果的 RMSE 降低了 10.76%, R^2 提高了 24.08%; 在对出水 SS 的预测, RVM 模型的 RMSE 降低了 5.45%, R^2 提高了 10.85%。

2.3 基于降维的 RVM 预测模型

通过降维得到的潜变量, 不仅对因变量具有最大解释性, 而且能反映因变量系统的信息, 使得自变量对因变量的解释效应在模型中得到体现。表 1 为 PLS 对输入变量降维后的各个潜变量的累计贡献率。对于

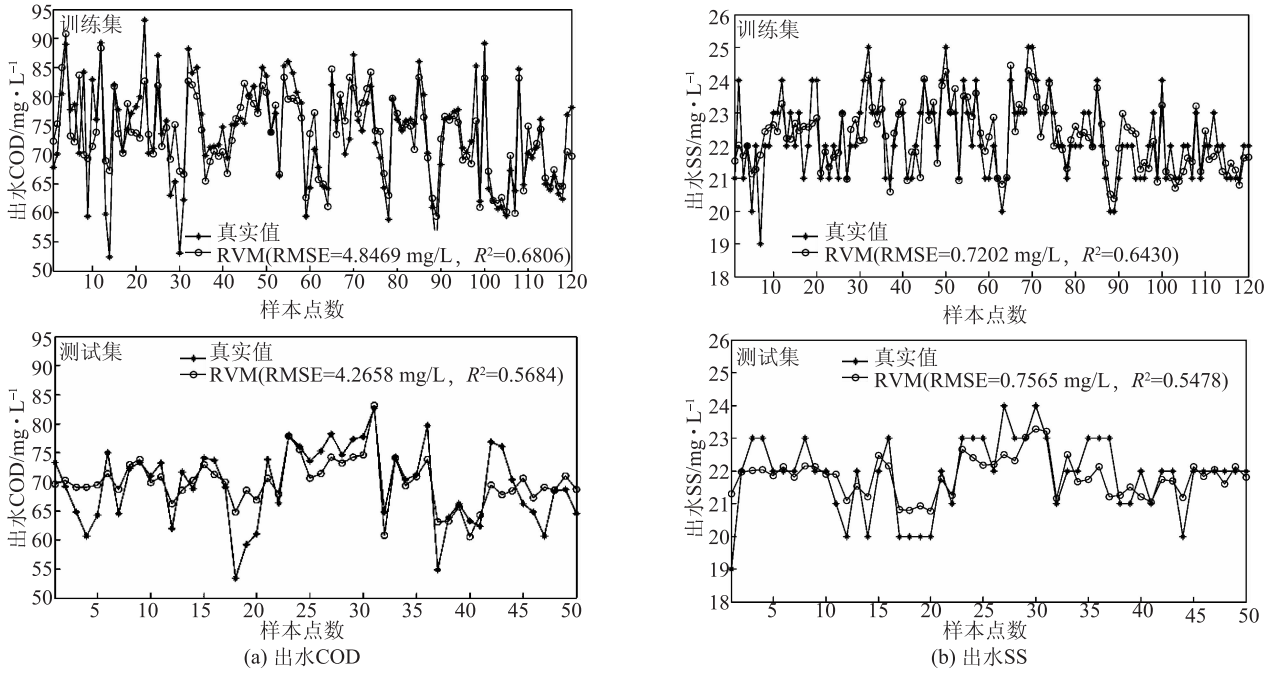


图 1 RVM 模型预测结果

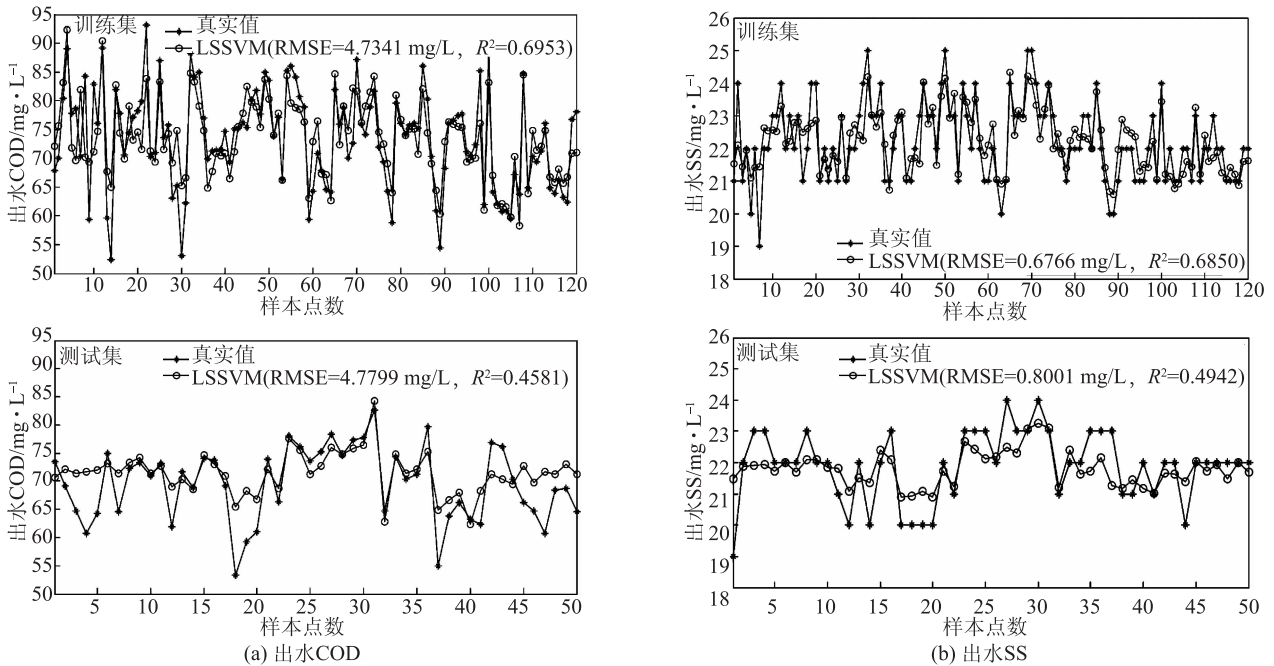


图 2 LSSVM 模型预测结果

PLS 潜变量个数的确定目前尚无统一的方法。表 1 显示, 潜变量 1~5 捕获了自变量 85% 以上的方差, 但是潜变量 4~6 对响应变量只占不到 0.5% 的方差, 响应变量累计贡献率基本不变, 说明后 3 个潜变量对响应变量基本无影响。考虑到随着变量的增多, 模型的复杂性及建模难度也会随之增大, 因此本文选取前 3 个潜变量作为模型的输入, 对出水 COD 与出水 SS 建立 PLS-RVM 模型, 并与 PLS-LSSVM 模型进行对比。

表 1 PLS 各潜在变量贡献百分比

潜变量	贡献率	累计贡献率	对响应变量累计贡献率
1	25.27	25.27	50.02
2	16.81	42.08	58.40
3	12.28	54.36	62.04
4	20.38	74.74	62.23
5	15.71	90.45	62.27
6	9.55	100.00	62.27

图 3 和图 4 分别表示 PLS-RVM 与 PLS-LSSVM 模型预测的出水 COD 及出水 SS 的预测值与测量真实值。对于出水 COD, PLS-LSSVM 模型预测结果的 RMSE 为 4.6311 mg/L, R^2 为 0.4913; 对于出水 SS, PLS-LSSVM 模型的 RMSE 为 0.7682 mg/L, R^2 为 0.5337。相较于 PLS-LSSVM 模型, 在对出水 COD 的预测, PLS-RVM 模型的 RMSE 降低了 9.29%, R^2 提高了 18.34%; 在对出水 SS 的预测, PLS-RVM 模型的 RMSE 降低了 9.16%, R^2 提高了 15.29%。

2.4 结果分析

表 2 显示的是 4 种模型的预测性能对比结果, 由此可以看出, 基于 PLS 降维的模型其建模预测效果要优于未降维模型的预测效果。对于出水 COD 模型, PLS-RVM 训练模型的 RMSE 为 3.6953 mg/L, R^2 为 0.8143, 相较于 RVM 模型, RMSE 降低了 23.76%, R^2 提高了 19.64%; 模型测试结果的 RMSE 为 4.2010 mg/L, R^2 为 0.5814, 相较于 RVM 模型, RMSE 与 R^2 无明显优势提升。对于出水 SS 模型, PLS-RVM 训练

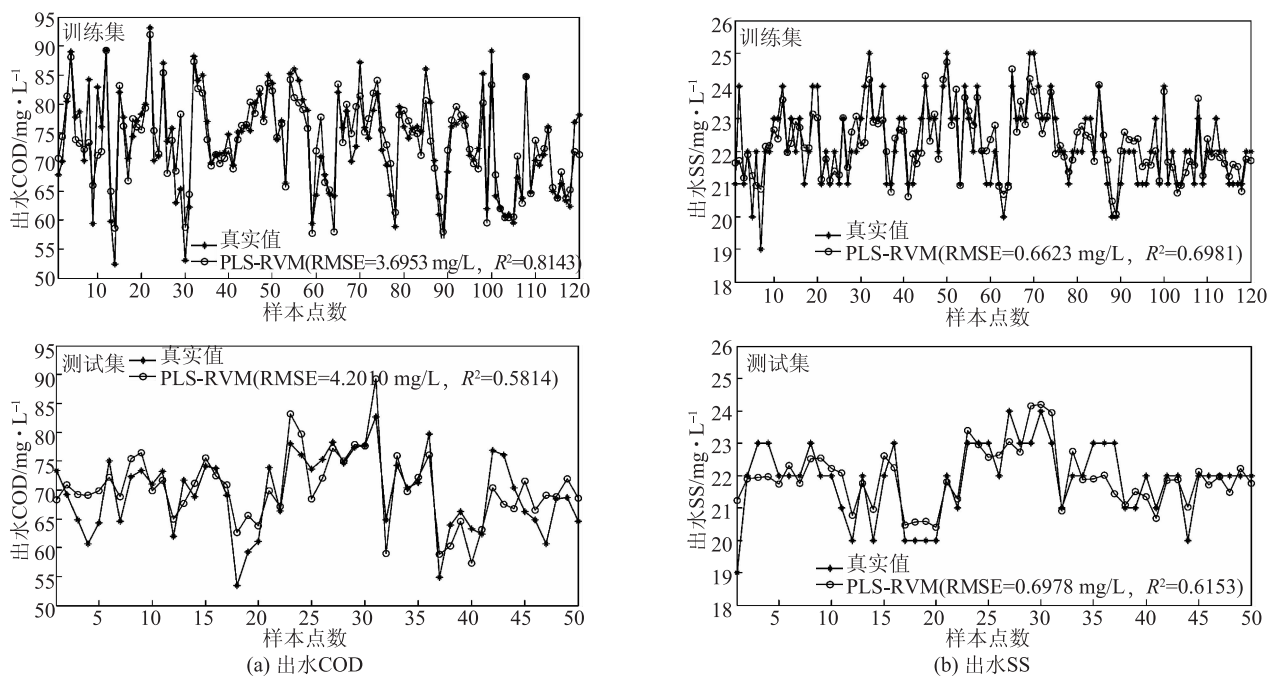


图 3 PLS-RVM 模型预测结果

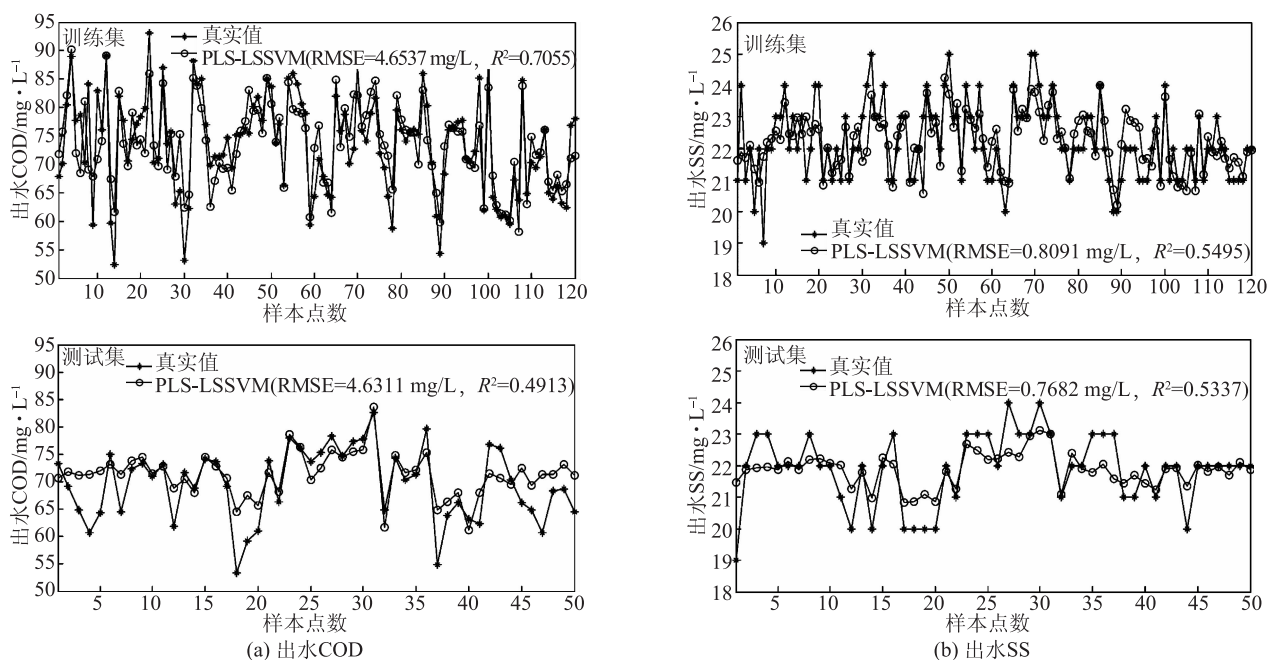


图 4 PLS-LSSVM 模型预测结果

模型的 RMSE 为 0.6623 mg/L, R^2 为 0.6981, 相较于 RVM 模型, RMSE 降低了 8.04%, R^2 提高了 8.57%; 在模型测试时的 RMSE 为 0.6978 mg/L, R^2 为 0.6153, 相较于 RVM 模型测试结果, RMSE 降低了 7.76%, R^2 提高了 12.32%。此外, 在模型训练时, 两种出水水质的 RVM 模型的训练结果的 RMSE 与 R^2 要略差于 LSSVM 模型, 但是在模型预测测试时, 其 RMSE 与 R^2 要优于 LSSVM 模型。而 PLS-RVM 模型无论是在训练拟合还是预测测试阶段, 其建模预测效果都要优于 PLS-LSSVM 模型。说明在对数据进行降维后, 模型的建模难度和预测精度得到优化, 相比于 LSSVM 模型, RVM 模型的泛化能力更强, 对样本的测试能力优于 LSSVM 模型。

表 2 4 种模型的预测结果

模型		出水 COD		出水 SS	
		RMSE/mg·L ⁻¹	R^2	RMSE/mg·L ⁻¹	R^2
LSSVM	训练集	4.7341	0.6953	0.6766	0.6850
	测试集	4.7799	0.4581	0.8001	0.4942
RVM	训练集	4.8469	0.6806	0.7202	0.6430
	测试集	4.2658	0.5684	0.7565	0.5478
PLS-LSSVM	训练集	4.6537	0.7055	0.8091	0.5495
	测试集	4.6311	0.4913	0.7682	0.5337
PLS-RVM	训练集	3.6953	0.8143	0.6623	0.6981
	测试集	4.2010	0.5814	0.6978	0.6153

3 结论

本文提出了一种对造纸废水出水水质建模预测的软测量模型。预测模型首先通过偏最小二乘法 (PLS) 对原始变量投影提取潜变量达到对数据降维的目的, 然后采用相关向量机 (RVM) 模型对造纸废水处理过程的出水水质进行建模和预测。对出水 COD 的预测模型, PLS-RVM 建模训练的均方根误差 (RMSE) 最小, 为 3.6953 mg/L, 相较于 RVM 模型降低了 23.76%; 但是预测结果的 RMSE 为 4.2010 mg/L, 相较于 RVM 模型并无明显优势提升。对于出水 SS 的预测模型, 同样 PLS-RVM 建模训练的 RMSE 最小, 为 0.6623 mg/L, 相较于 RVM 模型降低了 8.04%; 其预测结果的 RMSE 为 0.6978 mg/L, 相较于 RVM 模型降低了 7.76%。结果表明, 本文提出的 PLS-RVM 组合模型能够降低模型建立的复杂度并提高预测模型的泛化能力。

本文所述非线性建模方法适用于表征变量间的静态关系, 对于造纸废水处理系统这种动态特性很强的

过程, 如何将非线性方法与动态建模方法 (如多元时间序列分析法) 相结合来提高模型的预测精度, 是今后研究的方向。

参考文献

- [1] Fortuna L, Graziani S, Rizzo A, et al. Soft sensors for monitoring and control of industrial processes [M]. Springer London, 2007, 128.
- [2] Huang D P, Liu Y Q, Li Y. Soft sensor research and its application in wastewater treatment[J]. CIESC Journal, 2011, 62(1): 1. 黄道平, 刘乙奇, 李 艳. 软测量在污水处理过程中的研究与应用 [J]. 化工学报, 2011, 62(1): 1.
- [3] Cao P F, Luo X L. Modeling of soft sensor for chemical process[J]. Ciesc Journal, 2013, 64(3): 788. 曹鹏飞, 罗雄麟. 化工过程软测量建模方法研究进展 [J]. 化工学报, 2013, 64(3): 788.
- [4] Wang Y, Xu L, Yin W Z, et al. Soft sensor modeling of papermaking treatment processes based on ANN and LSSVR [J]. Transactions of China Pulp and Paper, 2017, 32(1): 50. 汪 瑶, 徐 亮, 殷文志, 等. 基于 ANN 和 LSSVR 的造纸废水处理过程软测量建模 [J]. 中国造纸学报, 2017, 32(1): 50.
- [5] Huang Y R, Zhang S D. Online soft measurement for wastewater treatment based on MIMO least squares support vector machine [J]. Automation & Instrumentation, 2010(4): 15. 黄银蓉, 张绍德. MIMO 最小二乘支持向量机污水处理在线软测量研究 [J]. 自动化与仪器仪表, 2010(4): 15.
- [6] Han Z H, Li Q J, Yuan Y M, et al. RVM based short-term wind speed prediction model [J]. Journal of Electric Power Science and Technology, 2017, 32(3): 38. 韩中合, 李秋菊, 苑一鸣, 等. 基于相关向量机的短期风速预测模型 [J]. 电力科学与技术学报, 2017, 32(3): 38.
- [7] Song L, Yang C, Zhang H, et al. Soft sensor modeling of papermaking wastewater treatment process based on Gaussian process regression [J]. China Environmental Science, 2018, 38(7): 2564. 宋 留, 杨 冲, 张 辉, 等. 基于高斯过程回归的造纸废水处理过程软测量建模 [J]. 中国环境科学, 2018, 38(7): 2564.
- [8] Tipping M E. Sparse Bayesian Learning and Relevance Vector Machine [J]. Journal of Machine Learning Research, 2001, 1(3): 211.
- [9] Liu Z X, Zhang D Y, Sun Q D, et al. Mid-term electric load prediction based on the relevant vector machine [J]. Journal of XI'AN Jiaotong University, 2004, 38(10): 1005. 刘遵雄, 张德运, 孙钦东, 等. 基于相关向量机的电力负荷中期预测 [J]. 西安交通大学学报, 2004, 38(10): 1005.
- [10] Xu Y G, Cao T, Luo F. Wastewater effluent quality prediction model based on relevance vector machine [J]. Journal of South China University of Technology (Natural Science Edition), 2014, 42(5): 103. 许玉格, 曹 涛, 罗 飞. 基于相关向量机的污水处理出水水质预测模型 [J]. 华南理工大学学报 (自然科学版), 2014, 42(5): 103.
- [11] Liu D W, Shen W H. Introduction of the benchmark simulation

- model-1 for wastewater treatment [J]. *Guangzhou Environmental Sciences*, 2007, 22(1): 11.
- 刘大伟, 沈文浩. 废水处理仿真基准模型 BSM1 简介 [J]. *广州环境科学*, 2007, 22(1): 11.
- [12] Dong Y, Qin S J. Dynamic latent variable analytics for process operations and control [J]. *Computers & Chemical Engineering*, 2018, 114: 69.
- [13] Peng K X, Ma L, Zhang K. Review of Quality-related Fault Detection and Diagnosis Techniques for Complex Industrial Processes [J]. *ACTA Automatica Sinica*, 2017, 43(3): 349.
- 彭开香, 马亮, 张凯. 复杂工业过程质量相关的故障检测与诊断技术综述 [J]. *自动化学报*, 2017, 43(3): 349.
- [14] Wong P K, Xu Q, Vong C M, et al. Rate-dependent hysteresis modeling and control of a piezostage using online support vector machine and relevance vector machine [J]. *IEEE Transactions on Industrial Electronics*, 2012, 59(4): 1988.

Prediction of Effluent Quality in Wastewater Treatment Systems Using Relevance Vector Machine

LIU Hongbin^{1,2,*} SONG Liu¹

(1. *Jiangsu Co-Innovation Center of Efficient Processing and Utilization of Forest Resources, Nanjing Forestry University, Nanjing, Jiangsu Province, 210037*; 2. *State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou, Guangdong Province, 510640*)

(* E-mail: hongbinliu@njfu.edu.cn)

Abstract: Predicting effluent quality precisely is important for papermaking wastewater treatment processes. In this paper, a soft sensor model based on relevance vector machine (RVM) was proposed to predict the effluent quality in a papermaking wastewater treatment plant. The proposed method included two steps. In the first step, partial least squares (PLS) method was used to extract the latent variables of the papermaking process data, which could solve the problems of collinearity and high dimensionality in the process variables. A RVM model using the extracted latent variables was developed in the second step. The results showed that the prediction performance of the PLS-RVM was better than that of the RVM for the prediction of the effluent quality in a papermaking wastewater treatment plant. With regard to the prediction of effluent suspension solids (SS), the root mean square error (RMSE) was decreased by 7.76% and the determination coefficient (R^2) was increased by 12.32%; however, the prediction results were not improved significantly for the prediction of effluent chemical oxygen demand (COD). In addition, the prediction performance of the PLS-RVM was better than that of the PLS-LSSVM model for both the effluent SS and the effluent COD. In terms of the prediction of the effluent SS, the RMSE was decreased by 9.16% and the R^2 was increased by 15.29%. In terms of the prediction of the effluent COD, the RMSE was decreased by 9.29% and the R^2 was increased by 18.34%.

Keywords: relevance vector machine; dimensionality reduction; support vector machine; papermaking wastewater treatment process; soft sensor

(责任编辑:刘振华)