

## 激光与光电子学进展

基于 SPA 和 PSO-LSSVM 的可见-近红外光谱  
水质 pH 值检测

李登珊, 李丽娜\*, 张认成

华侨大学机电及自动化学院, 福建 厦门 361021

**摘要** 为了提高可见-近红外(Vis-NIR)光谱法检测水质 pH 值的精度和稳定性,基于连续投影算法(SPA)和粒子群优化-最小二乘支持向量机(PSO-LSSVM)建立了多元校正模型。采集 60 个不同 pH 值水溶液样品的 Vis-NIR 光谱数据,运用 Savitzky-Golay 卷积平滑和标准正态变量变换对原始光谱数据进行预处理。基于 SPA 筛选的特征波长和 PSO 算法自动优化 LSSVM 的建模参数,建立多元非线性校正模型。结果表明,相比于其他对比模型,SPA-PSO-LSSVM 模型具有更高的精度与更优的稳定性,验证集的均误差方根为 0.67,决定系数为 0.91,剩余预测偏差为 3.10。

**关键词** 光谱学; 可见-近红外光谱; 连续投影算法; 粒子群优化; 最小二乘支持向量机

中图分类号 O657.33

文献标志码 A

DOI: 10.3788/LOP212915

Water Quality pH Value Determination for Visible-Near Infrared  
Spectroscopy Based on SPA and PSO-LSSVM

Li Dengshan, Li Lina\*, Zhang Rencheng

College of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, Fujian, China

**Abstract** To improve the detecting precision and robustness in the determination of water pH value using visible near-infrared (Vis-NIR) spectroscopy, a multivariate calibration model is constructed based on successive projections algorithm (SPA) and particle swarm optimization-least squares support vector machine (PSO-LSSVM). The Vis-NIR spectra data of 60 water samples with different pH values are collected, and the original spectral data are preprocessed by Savitzky-Golay smoothing and standard normal variate. Based on the characteristic wavelength of SPA screening and PSO algorithm, the modeling parameters of LSSVM are automatically optimized and a multivariate nonlinear calibration model is established. The results show that the SPA-PSO-LSSVM model has higher accuracy and stability than the comparison models. For the verification set, the root mean square error is 0.67, the coefficient of determination is 0.91, and the residual predictive deviation is 3.10.

**Key words** spectroscopy; visible near-infrared spectroscopy; successive projections algorithm; particle swarm optimization; least squares support vector machine

## 1 引言

pH 值是评价水质的重要指标之一,水质的 pH 值会影响人和动物的健康,过酸或过碱的水质具有很强的腐蚀能力,对环境也会产生巨大危害。目前,水质恶化已成为水环境建设中不可忽视的问题,因此,研究一种快速准确、可在线持续监测水质 pH 值的检测方法具有重要意义。

水质 pH 值的检测方法包括 pH 试纸法、pH 指示剂法、pH 计法等<sup>[1-4]</sup>。但常规方法操作过程复杂、费时,无法满足快速检测的需求。可见-近红外(Vis-NIR)光谱分析技术具有快速、无损和可实现在线检测等优点,已被广泛用于无损检测<sup>[5]</sup>。在可见-近红外区域,光谱的吸收主要来自于含氢基团(如 O—H、C—H 和 N—H)基本振动的合频和倍频。水质 pH 值的变化会影响水中 O—H 吸收带的形状和位置,因此,利用可见-近

收稿日期: 2021-11-09; 修回日期: 2021-12-02; 录用日期: 2021-12-13; 网络首发日期: 2021-12-23

基金项目: 国家自然科学基金青年科学基金(51805177)、华侨大学科研启动基金(11BS413)

通信作者: \*lilina@hqu.edu.cn

红外光谱检测水质 pH 值是可行的。杜艳红等<sup>[6]</sup>利用逐步多元线性回归 (MLR) 方法及可见-近红外光谱分析了水质的 pH 值, 获得了满意的预测精度, 但该方法检测的水质 pH 值范围较窄 (8.80~9.00), 无法满足实际应用的需求。在可见-近红外光谱分析中, 光谱数据中除了包含和待测成分相关的变量信息外, 还存在大量噪声和非信息变量。变量选择方法可以消除与待测成分无关的变量, 提高多元校正模型的精度和稳定性<sup>[7-11]</sup>。连续投影算法 (SPA) 采用向量空间中的简单投影和正向选择法获得最小共线性子集, 相比于其他基于智能优化算法和统计信息的波长优选方法, SPA 具有良好的稳定性和可重复性<sup>[12-13]</sup>。此外, 在水质 pH 值的可见-近红外光谱检测分析过程中, 光谱数据与待测成分间往往存在非线性关系, 线性建模方法不能完全反映待测成分与光谱数据间的关系。最小二乘支持向量机 (LSSVM) 是一种非线性建模方法, 可在高维空间中用较少的样本变量进行建模, 获得较为满意的预测精度<sup>[14]</sup>。但 LSSVM 需要优化正则化参数和核函数参数, 在参数寻优过程中需要耗费大量的时间。

因此, 以 LSSVM 的预测均方根误差 (RMSEP) 作为目标函数, 通过多次迭代寻找全局最优解, 获得最优参数, 提高 LSSVM 的建模效率和预测精度。为了建立检测范围更广、预测精度更高的水质 pH 值可见-近红外光谱定量分析模型, 本文利用可见-近红外光谱分析技术对水质 pH 值进行定量分析研究。采集 60 个 pH 值范围为 2.71~12.22 的水溶液样品的可见-近红外光谱数据, 采用 Savitzky-Golay (S-G) 卷积平滑和标准正态变量 (SNV) 变换对原始光谱进行预处理, 利用 SPA 提取特征变量并结合粒子群优化 (PSO) 算法寻找最优的 LSSVM 建模参数, 建立水质 pH 值的 SPA-PSO-LSSVM 多元非线性校正模型, 为实现水质 pH 值的快速检测提供参考依据。

## 2 材料与方 法

### 2.1 仪器与试剂

实验使用的仪器: 光栅光谱仪为 FLAME-T-XR1-RS (美国海洋光学公司), 使用的光源为长寿命光源 HL-2000-LL, 光谱分辨率为 2860 pixel, 可检测波段为 400~1050 nm; pH-100 酸度计 (上海力辰仪器公司), 测量精度为  $\pm 0.01$ 。实验使用的试剂: 浓度为 0.1 mol/L 的 HCl 标准溶液 (深圳博林达科技有限公司); 浓度为 0.1 mol/L 的 NaOH 标准溶液 (深圳博林达科技有限公司); 蒸馏水。

### 2.2 实验方法

#### 2.2.1 样品制备

向蒸馏水中随机加入不同体积的 HCl 标准溶液或 NaOH 标准溶液, 共制备了 60 个不同 pH 值的样品, 用酸度计测量得到样品的 pH 值范围为 2.71~12.22, 平均值为 6.90, 标准差为 2.56。这表明样品的 pH 值分

布范围较广, 具有良好的代表性。

#### 2.2.2 光谱采集

将样品放在光程为 1 cm 的石英比色皿, 以空光谱仪为背景, 采集每个样品的光谱强度数据。实验时光谱仪参数设置如下: 平均次数为 200, 积分时间为 25 ms, 平滑次数为 3。用 Kennard-Stone (K-S) 法对光谱数据集进行分组, 其中, 校正集包含 40 个样品, 验证集包含 20 个样品。样品的 pH 值统计结果如表 1 所示。

表 1 样品的 pH 值统计结果  
Table 1 Statistical results of pH value for the samples

Sample	Sample number	Max value	Min value	Mean value	Standard deviation
Calibration set	40	12.22	2.71	7.56	2.52
Validation set	20	10.28	2.88	5.56	2.08

#### 2.2.3 模型建立与评价方法

光谱预处理和定量分析均使用 Matlab R2016b 软件实现: 为消除原始光谱中高频随机噪声、光散射等因素的影响, 用多项式阶数为 2、窗口参数为 7 的 S-G 卷积平滑和 SNV 变换对原始光谱数据进行预处理; 为提升预测模型的建模速度和精度, 采用 SPA 对波长变量进行筛选, 得到特征波长变量。采用 LSSVM 算法建立水质 pH 值预测模型, 结合 PSO 算法优化 LSSVM 的核参数和正则化参数, 保证模型的预测精度。

选用 RMSEP、决定系数 ( $R^2$ ) 和剩余预测偏差 (RPD) 作为模型预测精度评价指标。RMSEP 越小,  $R^2$  越接近 1, 表明模型的预测精度就越高。RPD 是真实值的标准差与预测均方根误差的比值: 当 RPD 在 2.0~2.5 范围时, 表明模型可以进行粗略的定量预测; 当 RPD 在 2.5~3.0 范围时, 表明模型具有较好的预测精度; 当 RPD 大于 3.0 时, 表明模型具有很好的预测精度。

### 2.3 连续投影算法

SPA 基于投影操作选择具有最少限度的冗余信息的变量子集, 使变量间的共线性达到最小<sup>[15]</sup>, 从而减少建模时的光谱波长变量数, 提高模型的预测精度和计算速度<sup>[16]</sup>。相比于其他基于智能优化算法或统计信息的波长优选方法, SPA 选择的特征波长具有良好的重复性。

设光谱矩阵  $\mathbf{X}$  为  $N$  行  $p$  列矩阵。其中:  $N$  为样本数;  $p$  为光谱波长变量数;  $\mathbf{x}_k \in \mathbf{R}^N$  为  $\mathbf{X}$  的第  $k$  列向量。SPA 的具体步骤如下:

1) 初始化: 设定要提取的波长数  $M$ , 在首次迭代 ( $m=1$ ) 前, 任选  $\mathbf{X}_{N \times p}$  的第  $j$  列  $\mathbf{x}_j$  记为  $\mathbf{x}_{k(0)}$ ,  $k(0)=j, j \in 1, \dots, p$ 。

2) 将没有被选入的列向量位置集合记为  $S = \{j, 1 \leq j \leq p, j \notin \{k(0), \dots, k(m-1)\}\}$ 。

3) 分别计算剩余列向量  $\mathbf{x}_j (j \in S)$  与当前列向量

$\mathbf{x}_{k(m-1)}$  的投影  $\mathbf{P}_{x_j} = \mathbf{x}_j - [\mathbf{x}_j^T \mathbf{x}_{k(m-1)}] \mathbf{x}_{k(m-1)} [\mathbf{x}_{k(m-1)}^T \mathbf{x}_{k(m-1)}]^{-1}$ ,  $j \in S$ 。

4) 选择投影最大值对应的波长点的序号, 令  $k(m) = \arg(\max_j \|\mathbf{P}_{x_j}\|)$ ,  $j \in S$ 。

5) 令  $\mathbf{x}_j = \mathbf{P}_{x_j}$ ,  $j \in S$ ,  $m = m + 1$ : 若  $m < M$ , 则返回步骤 2); 否则, 结束算法。

SPA 最终提取出的波长变量组合为  $\{k(m), m = 0, \dots, M - 1\}$ 。对于每一对  $\mathbf{x}_j$  和  $M$  确定的变量子集, 利用多元线性回归建立多元校正模型, 并计算验证集的 RMSEP, 最小 RMSEP 值对应的波长变量组合即为 SPA 最终的波长选择结果。需要注意的是, 在 SPA 中,  $k(0)$  和  $M$  的取值范围十分关键,  $k(0)$  的取值范围为  $1 \sim \rho$ : 当光谱波长变量数  $\rho$  小于校正集样本数  $N$  时,  $M$  应小于波长变量数  $\rho$ ; 当光谱波长变量数  $\rho$  大于校正集样本数  $N$  时,  $M$  应小于校正集样本数  $N$ 。

## 2.4 PSO-LSSVM 回归算法

LSSVM 利用等式约束替代 SVM 的不等式约束, 用线性方程组替代二次规划问题得到支持向量, 可以在高维空间中用较少的样本变量进行学习, 很好地解决小样本、高维数、非线性和局部最小等方面的问题<sup>[14]</sup>。此外, 径向基函数 (RBF) 只有一个参数  $\delta^2$ , 大幅减少了计算复杂性。因此, 采用 RBF 作为 LSSVM 算法的核函数, 则 LSSVM 模型需要优化的参数有两个, 分别是正则化参数  $\gamma$  和核参数  $\delta^2$ 。

确定 LSSVM 模型的正则化参数  $\gamma$  和核参数  $\delta^2$  一直是提高 LSSVM 模型的预测精度和泛化性能的关键问题<sup>[17]</sup>。常用的参数优化方法主要有网格搜索法、 $k$  折交叉验证法、遗传算法和 PSO 算法等。PSO 算法是一种全局寻优算法, 其原理简单, 搜索效率高, 可用于解决非线性极值问题。因此, 采用 PSO 算法对

LSSVM 的正则化参数  $\gamma$  和核参数  $\delta^2$  进行优化, 具体步骤如下:

1) 选定校正集样本和验证集样本, 将正则化参数  $\gamma$  和核参数  $\delta^2$  的选择范围均设置为  $0 \sim 10^6$ , 最大迭代次数为 200, 期望误差精度为 0.65。

2) 设定粒子群粒子个数为 20, 学习因子  $c_1 = c_2 = 1.5$ , 速度最大值  $v_{\max} = 20$ , 速度最小值  $v_{\min} = -20$ , 初始权重为 1.2, 终止权重为 0.8, 随机初始化每个粒子。

3) 进行 LSSVM 模型训练。

4) 计算验证集的 RMSEP 并将其作为每个粒子的适应度值, 根据每个粒子的适应度值更新每个粒子的速度和位置、历史最优位置以及群体的全局最优位置。

5) 判断算法是否满足终止条件: 若参数满足 PSO 训练的终止条件 (达到最大迭代次数或模型训练误差小于等于 0.65), 则终止 PSO 训练, 输出最终的正则化参数  $\gamma$  和核参数  $\delta^2$ , 建立 LSSVM 模型; 否则, 返回步骤 3) 重新进行迭代。

## 3 结果与讨论

### 3.1 不同 pH 值样品的原始光谱及预处理光谱

图 1(a) 为不同 pH 值样品的原始光谱, 图 1(b) 为蒸馏水 (pH=7.12) 的光谱。可以发现, 不同 pH 值样品的光谱与蒸馏水的光谱整体趋势基本一致, 但由于样品的 pH 值存在差异, 相应的光谱强度值也存在差异。此外, 从原始光谱强度数据中无法直接观察到特征峰。文献<sup>[18]</sup>的研究表明, 970 nm 附近的特征峰对应水的  $2v_1 + v_3$  组合跃迁, 840 nm 附近的特征峰对应水的  $2v_1 + v_2 + v_3$  组合跃迁, 750 nm 附近的特征峰对应水的  $3v_1 + v_3$  组合跃迁。其中,  $v_1$  为 O—H 的对称拉伸,  $v_2$  为 O—H 的弯曲模式,  $v_3$  为 O—H 的反对称拉伸。

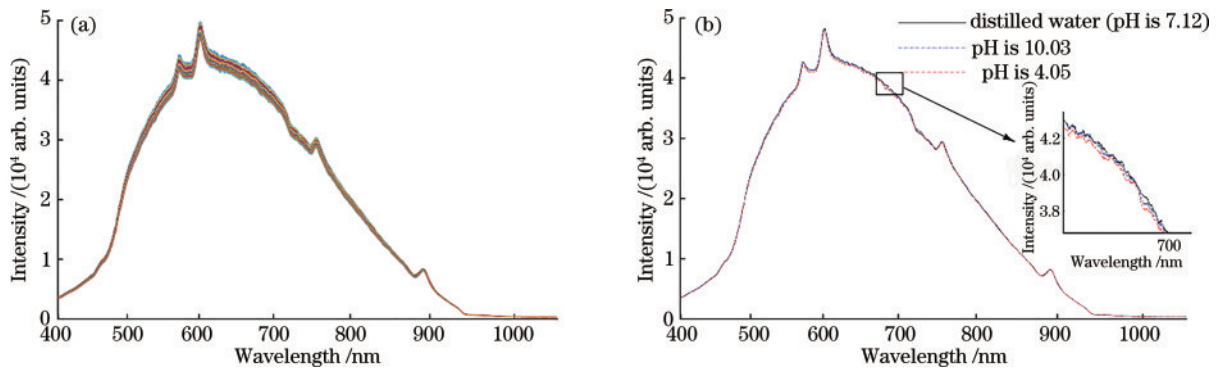


图 1 不同 pH 值样品的实验光谱。(a) 原始光谱; (b) 蒸馏水的光谱

Fig. 1 Experimental spectra of samples with different pH values. (a) Original spectra; (b) spectra of distilled water

由于光谱采集过程受高频随机噪声、光散射等多种因素的影响, 原始光谱包含的噪声较多。此外, 光谱强度值的范围过宽, 不利于校正模型的训练。因此, 先用多项式阶数为 2、窗口参数为 7 的 S-G 卷积平滑对原始光谱进行预处理, 然后采用 SNV 变换对光谱进一步

预处理。预处理后的光谱如图 2 所示。可以发现, 预处理后光谱变得平滑, 同时消除了随机噪声。

### 3.2 特征波长优选

水样的原始光谱数据量大, 光谱共线性严重, 导致多元校正建模运算时间长、精度低。SPA 基于投影操

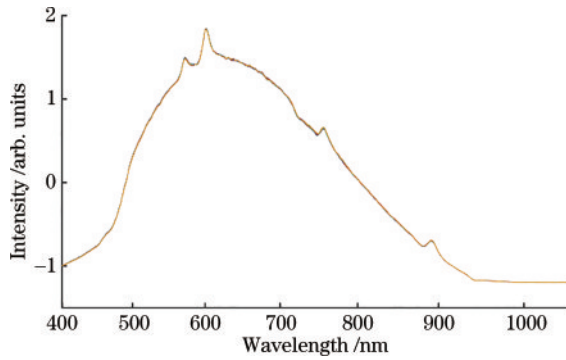


图 2 基于 S-G 平滑和 SNV 的光谱预处理结果

Fig. 2 Spectral preprocessing results based on S-G smoothing and SNV

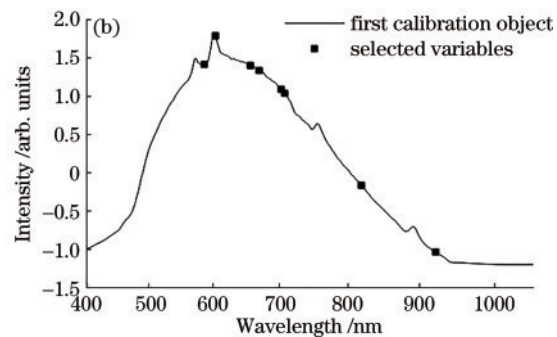
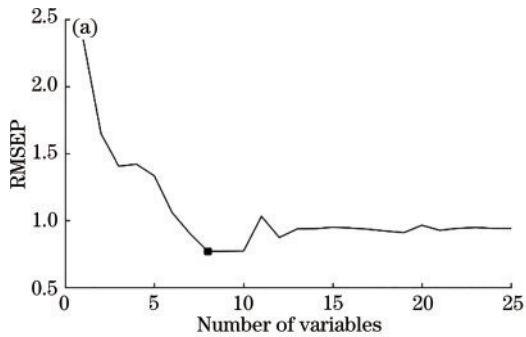


图 3 基于 SPA 的特征波长筛选。(a) 验证集的 RMSEP; (b) 特征波长分布

Fig. 3 Feature wavelength screening based on SPA. (a) RMSEP of the validation set; (b) characteristic wavelength distribution

### 3.3 水质 pH 值光谱分析多元校正模型的建立

#### 3.3.1 线性多元校正模型的建立

将预处理后的全波长数据集和 SPA 优选出的特征波长数据集的校正集分别作为 PLS 的输入数据, 建立 PLS 模型, 验证集的 RMSEP 随主因子数的变化曲线如图 4 所示。可以发现: 当主因子数为 8 时, 全谱 PLS 模型的 RMSEP 取得最小值; 当主因子数为 4 时, SPA-PLS 模型的 RMSEP 取得最小值。因此, 全谱 PLS 和 SPA-PLS 模型的最佳主因子数分别为 8 和 4。基于该结论分别建立水质 pH 值的全谱 PLS 和 SPA-PLS 多元校正模型。利用特征波长建立水质 pH 值的 SPA-MLR 多元校正模型。

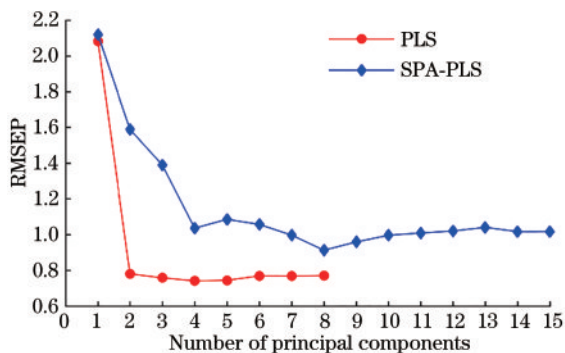


图 4 RMSEP 随主因子数变化曲线

Fig. 4 Variation curve of RMSEP with the number of principal components

作选择具有最少限度的冗余信息的变量子集, 使变量间的共线性达到最小<sup>[19]</sup>。水质 pH 值检测实验的光谱数据经过预处理后, 采用 SPA 对光谱特征波长进行筛选, 设置 SPA 提取的波长变量数  $M$  范围为 1~25, 利用 SPA 筛选的波长变量分别建立多元线性回归校正模型, 并计算各模型验证集的 RMSEP 值。当 SPA 选取不同数目的特征波长时, 验证集的 RMSEP 值的变化曲线如图 3(a) 所示。可以发现, 提取的特征波长数为 8 时, 验证集的 RMSEP 取得最小值, 为 0.7704。因此, SPA 最终选择了 8 个特征波长, 特征波长的分布如图 3(b) 所示, 按照重要性排序依次为 818.87、922.06、708.34、587.02、604.32、703.25、657.04、670.42 nm。

#### 3.3.2 非线性多元校正模型的建立

建立 LSSVM 模型时, 需要考虑 3 个关键问题, 即最优输入子集、合适的核函数以及最优的模型参数<sup>[14]</sup>。首先, 将全波长数据集和 SPA 优选出的特征波长数据集的校正集样本分别作为最优输入子集, 然后采用径向基核函数作为 LSSVM 的内核函数。利用 PSO 算法优化 LSSVM 的正则化参数  $\gamma$  和核参数  $\delta^2$ , 优化过程如图 5 所示。可以发现: 基于全谱的 PSO-LSSVM 模型的参数优化过程中, 验证集的 RMSEP 值最初保持稳定, 随后下降至 0.8799 后保持稳定并达到全局最优; 基于特征波长建立的 SPA-PSO-LSSVM 模型的优化过程中验证集 RMSEP 值不断下降, 到达 0.6681 后保持稳定。最终得到最优 PSO-LSSVM 模型的  $\gamma=30033.60$ 、 $\delta^2=2256.18$ , 最优 SPA-PSO-LSSVM 模型的  $\gamma=11059.35$ 、 $\delta^2=1324.38$ 。根据优化结果建立的 PSO-LSSVM 和 SPA-PSO-LSSVM 多元校正模型及其验证结果, 如表 2 所示。可以发现: 全谱 PLS 模型的 RMSEP 为 0.91,  $R^2$  为 0.82, RPD 为 2.28, 表明该模型可以进行粗略的定量预测; SPA-PLS 模型的 RMSEP 为 0.74,  $R^2$  为 0.87, RPD 为 2.81, 表明该模型具有较好的预测能力与稳定性; SPA-MLR 模型的 RMSEP 为 0.77,  $R^2$  为 0.88, RPD 为 2.70, 表明该模型具有较好的预测能力与稳定性; PSO-LSSVM 模型的 RMSEP 为 0.88,  $R^2$  为 0.83, RPD 为 2.36, 表明该模型可以进行粗略的定量预测。

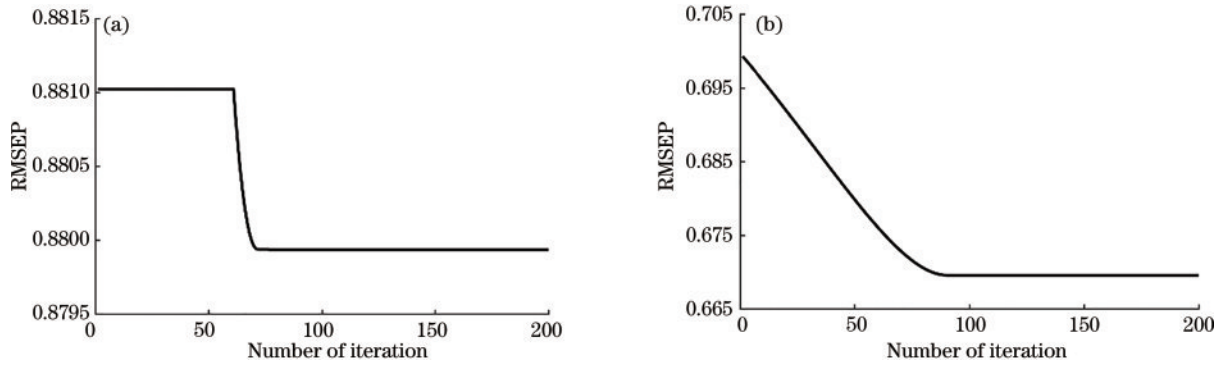


图 5 PSO 优化 LSSVM 参数的过程。(a) PSO-LSSVM; (b) SPA-PSO-LSSVM

Fig. 5 Process of PSO optimizing LSSVM parameters. (a) PSO-LSSVM; (b) SPA-PSO-LSSVM

表 2 不同建模方法的参数和结果

Table 2 Parameters and results of different modeling methods

Mode	Number of wavelengths	Components number	$\gamma$	$\delta^2$	RMSEP	$R^2$	RPD
PLS	2860	8	—	—	0.91	0.82	2.28
SPA-PLS	8	4	—	—	0.74	0.87	2.81
SPA-MLR	8	—	—	—	0.77	0.88	2.70
PSO-LSSVM	2860	—	30033.60	2256.18	0.88	0.83	2.36
SPA-PSO-LSSVM	8	—	11059.35	1324.38	0.67	0.91	3.10

基于特征波长建立的 SPA-PSO-LSSVM 模型的预测集实测值和预测值散点图如图 6 所示。可以发现,该模型的 RMSEP 为 0.67,  $R^2$  为 0.91, RPD 为 3.10, 大部分样本点接近理想回归线, 具有很好的预测能力和稳定性。

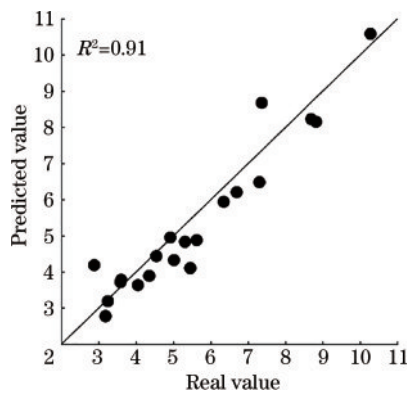


图 6 SPA-PSO-LSSVM 模型对于验证集的预测值散点图

Fig. 6 Scatter plot of the predicted values of the SPA-PSO-LSSVM model for the validation set

### 3.4 多元校正方法的验证结果

SPA 从原始光谱的 2860 个波长变量中选择了 8 个特征波长, 波长变量数缩减为原来的 0.28%。相比全谱 PLS 模型, SPA-PLS 模型的主因子数从 8 个减少为 4 个, RMSEP 降低了 18.68%,  $R^2$  和 RPD 分别增加了 6.09% 和 23.24%。这表明 SPA 可以简化多元校正模型的建模复杂度, 提高模型预测精度, 是一种有效的特征波长选择方法。

对比不同模型的预测结果发现: 相比于 SPA-PLS

模型, SPA-PSO-LSSVM 模型的 RMSEP 降低了 9.46%,  $R^2$  和 PRD 分别增加了 4.59% 和 10.32%; 相比于 SPA-MLR 模型, SPA-PSO-LSSVM 模型的 RMSEP 降低了 12.98%,  $R^2$  和 PRD 分别增加了 3.41% 和 14.81%; 相比于 PSO-LSSVM 模型, SPA-PSO-LSSVM 模型的 RMSEP 降低了 23.86%,  $R^2$  和 RPD 分别增加了 9.64% 和 31.36%。这表明相比于 PLS、SPA-PLS、SPA-MLR 和 PSO-LSSVM 等模型, SPA-PSO-LSSVM 模型具有更高的预测精度。

相比于人工经验选取的 LSSVM 模型参数, SPA-PSO-LSSVM 模型利用 PSO 自动寻找 LSSVM 的全局最优参数, 提高了建模速度。相比于 PSO-LSSVM, SPA-PSO-LSSVM 利用特征波长建模, PSO 寻找 LSSVM 最优参数所需的时间更短, 预测精度也有所提升。这表明 SPA-PSO-LSSVM 可以提高 LSSVM 模型的建模效率和预测精度, 是一种行之有效的非线性多元校正方法。

综上所述, 相比于 PLS、SPA-PLS、SPA-MLR 和 PSO-LSSVM 模型, SPA-PSO-LSSVM 模型的预测精度最高。SPA-PSO-LSSVM 模型的 RMSEP 为 0.67,  $R^2$  为 0.91, RPD 为 3.10, 满足水质 pH 值检测的实际应用精度要求。

## 4 结 论

采用可见-近红外光谱分析技术, 对水质 pH 值范围为 2.71~12.22 的 60 个样品的 pH 值进行定量分析, 建立了基于特征波长的 SPA-PSO-LSSVM 非线性多元校正模型。结果表明, SPA 从 2860 个波长变量中选

择 8 个特征波长,降低了多元校正模型的计算复杂度,提高了模型预测精度。基于用 PSO 算法自动寻找 LSSVM 的全局最优参数可提高 LSSVM 的建模效率。SPA-PSO-LSSVM 模型验证集的 RMSEP 为 0.67,  $R^2$  为 0.91, PRD 为 3.10, 相比于已有模型,该模型能检测的 pH 值范围更广,预测精度更高,满足水质 pH 值检测的实际应用需求。综上所述,可见-近红外光谱结合 SPA-PSO-LSSVM 可以实现水质 pH 值的快速检测,为水质 pH 值的快速检测提供了参考依据。

### 参 考 文 献

- [1] Kim H, Lim H K, Cho S, et al. An anthracene appended guanidine derivative as water soluble fluorescence sensor for high pH values and water content measurements[J]. *Journal of Photochemistry and Photobiology A: Chemistry*, 2019, 383: 112023.
- [2] Zhang X X, Zhang W J, Tang S K. L-Tryptophan functionalized graphene quantum dots as a fluorescence indicator for pH detection in real water[J]. *Journal of Photochemistry and Photobiology A: Chemistry*, 2019, 372: 71-77.
- [3] Rasouli Z, Abdollahi H, Maeder M. Generalized indicator-based determination of solution pH[J]. *Analytica Chimica Acta*, 2020, 1109: 90-97.
- [4] Shibata M, Kato M, Iwamoto Y, et al. Potentiometric determination of pH values of dilute sulfuric acid solutions with glass combination electrode equipped with ionic liquid salt bridge[J]. *Journal of Electroanalytical Chemistry*, 2013, 705: 81-85.
- [5] 张娟, 原帅, 张骏. 基于小波变换-遗传算法-偏最小二乘的草莓糖度检测研究[J]. *分析科学学报*, 2020, 36(1): 111-115. Zhang J, Yuan S, Zhang J. Detection of sugar degree in strawberry based on wavelet transform-genetic algorithm-partial least squares[J]. *Journal of Analytical Science*, 2020, 36(1): 111-115.
- [6] 杜艳红, 张伟玉, 杨仁杰, 等. 基于可见-近红外光谱的水质 pH 值分析[J]. *湖北农业科学*, 2012, 51(3): 612-614, 620. Du Y H, Zhang W Y, Yang R J, et al. Analysis on pH in water by visible-near infrared spectroscopy[J]. *Hubei Agricultural Sciences*, 2012, 51(3): 612-614, 620.
- [7] 莫欣欣, 孙通, 刘津, 等. 近红外光谱结合变量选择方法定性检测食用油中高效氟吡甲禾灵残留[J]. *分析试验室*, 2018, 37(2): 125-130. Mo X X, Sun T, Liu J, et al. Qualitative detection of haloxyfop-P-methyl residue in edible oil by near infrared spectroscopy combined with variable selection method[J]. *Chinese Journal of Analysis Laboratory*, 2018, 37(2): 125-130.
- [8] Yun Y H, Wang W T, Deng B C, et al. Using variable combination population analysis for variable selection in multivariate calibration[J]. *Analytica Chimica Acta*, 2015, 862: 14-23.
- [9] 赵环, 宦克为, 石晓光, 等. 基于自加权变量组合集群分析法的近红外光谱变量选择方法研究[J]. *分析化学*, 2018, 46(1): 136-142. Zhao H, Huan K W, Shi X G, et al. A variable selection method of near infrared spectroscopy based on automatic weighting variable combination population analysis[J]. *Chinese Journal of Analytical Chemistry*, 2018, 46(1): 136-142.
- [10] Xu S, Lu B, Baldea M, et al. An improved variable selection method for support vector regression in NIR spectral modeling[J]. *Journal of Process Control*, 2018, 67: 83-93.
- [11] Jiang H, Xu W D, Ding Y H, et al. Quantitative analysis of yeast fermentation process using Raman spectroscopy: comparison of CARS and VCPA for variable selection[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020, 228: 117781.
- [12] Yun Y H, Li H D, Deng B C, et al. An overview of variable selection methods in multivariate analysis of near-infrared spectra[J]. *TrAC Trends in Analytical Chemistry*, 2019, 113: 102-115.
- [13] 章海亮, 罗微, 刘雪梅, 等. 应用遗传算法结合连续投影算法近红外光谱检测土壤有机质研究[J]. *光谱学与光谱分析*, 2017, 37(2): 584-587. Zhang H L, Luo W, Liu X M, et al. Measurement of soil organic matter with near infrared spectroscopy combined with genetic algorithm and successive projection algorithm[J]. *Spectroscopy and Spectral Analysis*, 2017, 37(2): 584-587.
- [14] Liu F, He Y, Wang L. Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy [J]. *Analytica Chimica Acta*, 2008, 610(2): 196-204.
- [15] Soares S F C, Gomes A A, Araujo M C U, et al. The successive projections algorithm[J]. *TrAC Trends in Analytical Chemistry*, 2013, 42: 84-98.
- [16] 陈远哲, 王巧华, 高升, 等. 基于近红外光谱的淡水鱼贮藏期质构品质的无损检测模型[J]. *激光与光电子学进展*, 2021, 58(12): 1230001. Chen Y Z, Wang Q H, Gao S, et al. Nondestructive testing model for textural quality of freshwater fish in storage using near-infrared spectroscopy[J]. *Laser & Optoelectronics Progress*, 2021, 58(12): 1230001.
- [17] 郭志明, 陈立平, 黄文倩, 等. 近红外光谱结合 GA-LSSVR 分析烟草尼古丁含量[J]. *激光与光电子学进展*, 2012, 49(2): 021201. Guo Z M, Chen L P, Huang W Q, et al. Application of genetic algorithm-least squares support vector regression with near infrared spectroscopy for prediction of nicotine content in tobacco[J]. *Laser & Optoelectronics Progress*, 2012, 49(2): 021201.
- [18] Huang Y Q, Lin M S, Cavinato A G, et al. Influence of temperature on the measurement of NaCl content of aqueous solution by short-wavelength near infrared spectroscopy (SW-NIR)[J]. *Sensing and Instrumentation for Food Quality and Safety*, 2007, 1(3): 91-97.
- [19] 李冠稳, 高小红, 肖能文, 等. 特征变量选择和回归方法相结合的土壤有机质含量估算[J]. *光学学报*, 2019, 39(9): 0930002. Li G W, Gao X H, Xiao N W, et al. Estimation of soil organic matter content based on characteristic variable selection and regression methods[J]. *Acta Optica Sinica*, 2019, 39(9): 0930002.