

# 精细全光谱结合 GS-SVR 的复杂水体硝酸盐分析方法研究

雷会平<sup>1,2</sup>, 胡炳樑<sup>1</sup>, 于涛<sup>1\*</sup>, 刘嘉诚<sup>1</sup>, 李炜<sup>1,2</sup>, 王雪霁<sup>1</sup>, 邹妍<sup>3</sup>, 史倩<sup>3</sup>

1. 中国科学院西安光学精密机械研究所, 陕西 西安 710119
2. 中国科学院大学光电学院, 北京 100049
3. 山东省科学院海洋仪器仪表研究所, 山东 青岛 266000

**摘要** 硝酸盐是水质健康状态评价的一个关键要素。水体中高浓度的硝酸盐会导致生物多样性剧减以及生态系统的退化, 同时对人类的健康产生不可逆转的伤害。基于光学测量的水质在线监测是当前及未来水环境动态监测的发展趋势。相较于传统硝酸盐现场采样加实验室分析的测定方法, 具有操作便捷, 无需前处理, 检测效率高, 可靠性好且无污染等显著优点。由于实际水体组分的复杂性与多样性, 水体参数和吸光度二者并非呈现线性相关, 传统的单波长法, 双波长法, 偏小二乘法等线性回归预测模型已不适用。基于此, 提出一种精细全光谱结合可变速长网格搜索, 优化支持向量回归(GS-SVR)的水体硝酸盐分析方法。同时与陕西科技大学化学与化工学院合作, 采用标准的硝酸盐溶液, 铂-钴标准溶液, 福尔马肼标准混悬液根据实验要求配制了不同浓度梯度 94 组溶液样本。首先将采集到的透射率光谱数据完成吸光度转换, 并使用 Kennard-Stone 方法将 94 个溶液样本划分为 80 个训练集和 14 个测试集。其次使用改进的 GS 算法结合交叉验证, 通过多次迭代, 减小搜索范围、改变搜索步长对 SVR 进行参数寻优, 并将最优惩罚参数  $C$  和核函数宽度  $\sigma$  用于训练集中进行模型建立, 最后用所建立的模型对测试集进行浓度预测。并将预测效果与反向传播神经网络(BPNN), SVR, GS-SVR, 粒子群算法优化 SVR(PSO-SVR), 遗传算法优化 SVR(GA-SVR)的模型预测结果比较, 结果显示, 提出的算法模型相关系数  $R^2=0.9935$ , 预测均方根误差  $RMSEP=0.0435$ , 最优参数  $C$  和  $\sigma$  组合为(512, 0.0442), 平均训练时间为 13 s。相较于上述五种预测模型,  $R^2$  分别提高了 1.22%, 11.66%, 0.78%, 0.74% 和 0.77%, 训练效率分别提升 4.15 倍(BPNN), 8.30 倍(GS-SVR), 21.38 倍(PSO-SVR), 10.23 倍(GA-SVR)。模型的预测精度以及训练效率方面都取得了很大的提升, 为复杂水体硝酸盐浓度的快速实时在线监测提供了一种新的方法。同时, 该方法具备一定的普适性, 也适用于其他水质参数预测模型的建立。

**关键词** 精细全光谱; 硝酸盐; 网格搜索; 支持向量回归

**中图分类号:** X52 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)02-0372-07

## 引言

水是人类社会生产生活与持续发展必不可少的资源<sup>[1]</sup>。硝酸盐是水质健康状态评价的一个关键要素, 过高浓度的硝酸盐一旦流入水体环境就会刺激藻类快速繁殖, 致使水体恶化, 进而造成物种多样性剧减以及生态环境的破坏, 如果被人体误摄, 会引起“高铁血红蛋白症”。因此, 实现水体硝酸盐的高精度快速实时在线监测对于水务事业的安全保障与水

污染的预警防治极具重要的科学意义和社会价值。

基于光学测量的水质在线监测方法是当前及未来水环境动态监测的发展趋势<sup>[1]</sup>。相较于传统硝酸盐现场采样加实验室化学分析的测定方法, 其具有操作便捷, 无需前处理, 检测效率高, 重复性好且无二次污染等显著优点, 非常适合水环境中硝酸盐的快速在线监测, 在直饮水, 地表水, 废水的在线检测等方面具有广阔的发展前景。针对水体组份解析算法, 常见的有单波长法, 双波长法, 偏小二乘法(partial least square, PLS), 人工神经网络法(artificial neural net-

收稿日期: 2020-01-06, 修订日期: 2020-04-20

基金项目: 中国科学院战略性先导科技专项(A类)(XDA23040101), 国家重点研发计划项目(2017YFC1403700), 陕西省重点研发计划项目(2019SF-254), 山东省科学院青年基金项目(2018QN0036)资助

作者简介: 雷会平, 1993年生, 中国科学院西安光学精密机械研究所硕士研究生 e-mail: leihuiping2017@opt.cn

\* 通讯作者 e-mail: yutao@opt.ac.cn

work, ANN), 支持向量机法 (support vector machine, SVM)。Jean Causse<sup>[2]</sup>和 Pons<sup>[3]</sup>等采用连续光谱二阶导数测定淡水中硝酸盐和溶解有机碳, 并在多个淡水样品中验证了该方法的可靠性。宓云翥等<sup>[4]</sup>依次采用 PLS, SVM 和 LM-反向传播神经网络 (back propagation neural network, BPNN) 建立吸光度与化学需氧量 (chemical oxygen demand, COD) 间的浓度模型, 结果验证 SVM, LM-BPNN 等回归模型的预测效果显然优于 PLS。倪雪春等将支持向量回归机 (support vector regression, SVR) 运用于混合重叠的光谱分析中, 实现了硝酸盐含量的浓度预测。王莉丽等<sup>[5]</sup>研制了一种新型的海水硝酸盐检测系统, 结合 PLS 法进行硝酸盐浓度与吸收光谱间的相关性分析, 实现了海水硝酸盐的在线检测。杨琼<sup>[6]</sup>等结合近红外光谱以及后向间隔偏最小二乘 (back interval partial least square, BiPLS) 法对生活废水中的 COD 回归建模, 结果显示预测效果优于 PLS 建模。刘思乡等利用主成分分析对原始光谱数据降维, 然后使用局部加权线性回归建模, 获得了较好的预测结果。陈颖等<sup>[7]</sup>提出了一种基于光谱的混合回归预测模型, 包括双波长法与最小二乘支持向量机法, 将溶液样本分为高浓度样本和低浓度样本两部分, 验证结果良好。

但是, 实际水体组分更加复杂与多元化, 水体参数和吸光度二者并非呈现线性相关, 使用线性回归模型难以达到满意的预测效果, 所以必须建立非线性的预测模型。ANN 建模预测难以克服预测结果的不稳定性, 而 SVM 因其在小样本、非线性、高维问题中的优势获得广泛关注。本文结合精细全光谱技术, 提出一种基于改进的网格搜索 (grid search, GS) 方法优化 SVR 以期完成建模预测, 并与传统预测模型的结果加以对比, 结果验证, 提出的算法在预测精度以及训练效率方面都有显著提升。

## 1 基本原理

### 1.1 支持向量回归思想

SVM 基于结构风险最小化原理, 可以有效地应对实际应用情境下出现的小样本、非线性和高维问题<sup>[8]</sup>, 具有良好的性能和预测精度。SVR 的基本思路是把原始数据  $x$  非线性映射到高维特征空间, 然后在此空间中解决线性回归问题。SVM 的回归函数为

$$f(x) = w\varphi(x) + b \quad (1)$$

其中  $\varphi(x)$  为特征空间, 是权重向量,  $b$  是偏差项。系数和  $b$  则由风险函数最小化来估计

$$R(C) = \frac{1}{2} \|w\|^2 + C \frac{1}{l} \sum_{i=1}^l L_\epsilon(y_i, f(x_i)) \quad (2)$$

其中  $C$  为测量经验风险的代价函数。  $\frac{1}{2} \|w\|^2$  是正则化项。

$L_\epsilon(y_i, f(x_i))$  被称为  $\epsilon$ -不敏感损失函数, 定义如式(3)

$$L_\epsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \epsilon & |y_i - f(x_i)| \geq \epsilon \\ 0 & |y_i - f(x_i)| < \epsilon \end{cases} \quad (3)$$

式(3)中, 若预测误差小于  $\epsilon$ , 则损失为零, 否则损失超出范

围。引入两个正松弛变量  $\xi$  和  $\xi^*$  表示从实际值到相应边界值的距离。将式(2)转化为式(4)

$$\varphi(w, \xi, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi + \xi^*) \quad (4)$$

约束条件为

$$\begin{cases} y_i - [w\varphi(x)] - b \leq \epsilon + \xi_i & \xi_i < 0 \\ [w\varphi(x)] + b - y_i \leq \epsilon + \xi_i^* & \xi_i^* \geq 0 \end{cases} \quad (5)$$

该优化问题通过式(6)求解

$$H(\partial, \partial^*) = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\partial_i - \partial_i^*)(\partial_j - \partial_j^*) K(x_i, x_j) + \sum_{i=1}^l y_i (\partial_i - \partial_i^*) - \epsilon \sum_{i=1}^l (\partial_i + \partial_i^*) \quad (6)$$

约束条件为

$$\sum_{i=1}^l (\partial_i - \partial_i^*) = 0 \quad \partial_i, \partial_i^* \in [0, C] \quad (7)$$

其中  $\partial_i$  和  $\partial_i^*$  是 Lagrange 乘子。通过计算, 获得回归超平面的最优权向量, 即

$$w^* = \sum_{i=1}^l (\partial_i - \partial_i^*) K(x_i, x) \quad (8)$$

因此, 回归函数为

$$f(x) = \sum_{i=1}^l (\partial_i - \partial_i^*) K(x_i, x) + b \quad (9)$$

式(9)中,  $K(x_i, x)$  为核函数, 其值为  $K(x_i, x) = \Phi(x_i) \Phi(x)$ 。在 SVM 中, 径向基核函数 (RBF)  $K(x_i, x) = \exp(-\|x_i - x\| / 2\sigma^2)$  仅须确定一个变量, 便于参数优化。此外, 其构造的 SVR 具有良好的非线性预测性能。因此, 将 RBF 应用于 SVR 中。

### 1.2 改进的网格搜索法

据实际应用情况, 参数设置和核函数的选择对 SVR 的学习和泛化性能影响很大<sup>[9]</sup>。网格搜索 (grid search, GS) 是一种直接的参数优化算法, 用于确定 SVR 的最佳参数值<sup>[10-11]</sup>。通过设置搜索中的上下界 (搜索间隔) 和跳跃间隔的适当值, 使得优化算法遍历网格中的每一个参数组合, 并采用交叉验证技术作为性能度量找到最优参数组合。实际操作中往往设置较大的搜索范围和较小的搜索间隔, 因此, 参数寻优的过程会占用大量时间, 导致模型效率低下。本文提出了多次变步长的网格搜索算法对 SVR 模型参数进行调整, 具体操作步骤如下:

(1) 初始化参数搜索范围及搜索步长, 在初始的网格搜索范围中, 将预先设置的参数组合搜索步长扩大  $T$  倍, 进行第一次参数寻优。

(2) 依据图 1 的参数搜索空间确定流程, 将  $C$  与  $\sigma$  的搜索范围缩小, 并且缩小搜索步长  $F$  倍, 继续第二次参数寻优。具体流程如图 1。

(3) 依据图 1 的参数空间确定流程, 把  $C$  与  $\sigma$  的搜索空间再次缩小, 同时把它的步长缩小  $S$  倍, 继续进行参数寻优。其中  $T=FS$ 。

(4) 将(3)中获取的最优参数组合代入 SVR 模型, 建立训练集吸光度光谱数据与硝酸盐浓度的数学模型。

(5) 将测试集应用于步骤(4)获取的回归模型完成预测,

输出测量结果。

通过以上操作，将原本的搜索空间减小并多次改变步长，简化了计算，大幅提升了寻优效率。

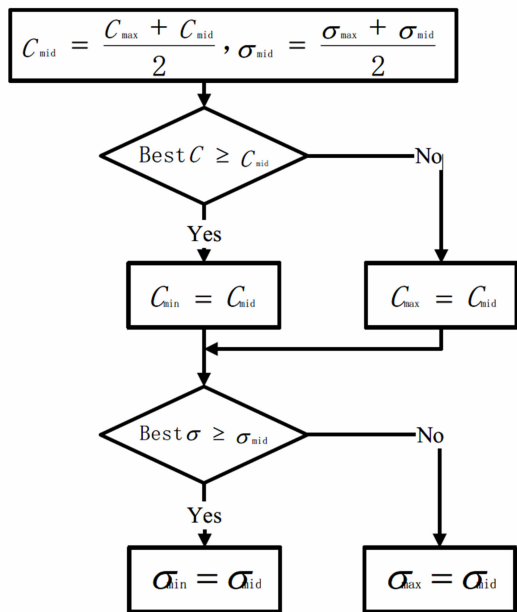


图 1 参数搜索空间的确定

Fig. 1 Determination of parameter search space

## 2 实验部分

### 2.1 数据获取

使用自研光谱仪采集溶液的原始透射光谱数据，并将其转换为吸光度光谱数据，然后基于可变步长的网格搜索法进行模型参数寻优，将获得的最优参数组合应用于 SVR 模型，建立训练集吸光度光谱数据与硝酸盐浓度间的数学模型，然后利用该数学模型完成测试集的浓度预测。总体流程图如图 2 所示。

### 2.2 样本溶液

溶液样本的配制在陕西科技大学化学与化工学院实验室内完成。采用标准的硝酸盐溶液，铂-钴标准溶液，福尔马肼标准混悬液根据实验要求按照不同的浓度梯度配置了 94 组不同浓度的溶液样本。硝酸盐、浊度、色度的浓度量程分别为  $7\sim 15\text{ mg}\cdot\text{L}^{-1}$ ， $0.5\sim 5\text{ NTU}$ ， $7\sim 15$  度(铂-钴色度单位)，浓度间隔分别为  $1\text{ mg}\cdot\text{L}^{-1}$ ， $0.5\text{ NTU}$ ，1 度。每一个待测溶液样本的三种组份浓度配比为随机生成。

### 2.3 装置

实验装置采用光源，光源衰减器，比色皿，光谱仪和计算机组成的集成系统获取被测物质的透射光谱。样本溶液光谱数据的采集装置选取了项目组自主研发的双光路主动校正连续光谱仪，原理如图 3 所示。该装置使用双光路主动校正结合连续光谱精细获取的方法，相对于传统的单光路连续长时间作业的系统误差校正进行了去除。自研仪器的光谱范围： $185\sim 1100\text{ nm}$ ，光谱分辨率： $1.5\text{ nm}$ ，闪耀波长： $250\text{ nm}$ ，共含 3 648 个像素点。

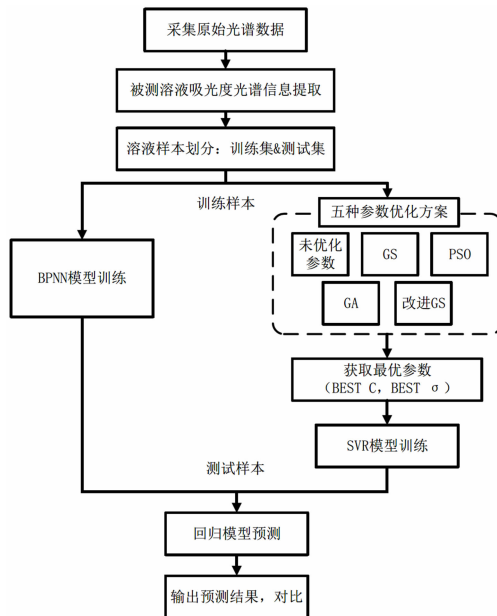


图 2 实验整体流程

Fig. 2 Overall flow of the experiment

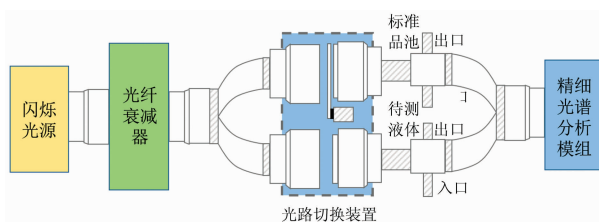


图 3 光谱法主要测量技术原理图

Fig. 3 Main measurement technology schematic diagram of spectrometric method

将被测溶液的原始透射光谱数据转换为吸光度光谱数据，所有样本溶液的吸光光谱图如图 4 所示。

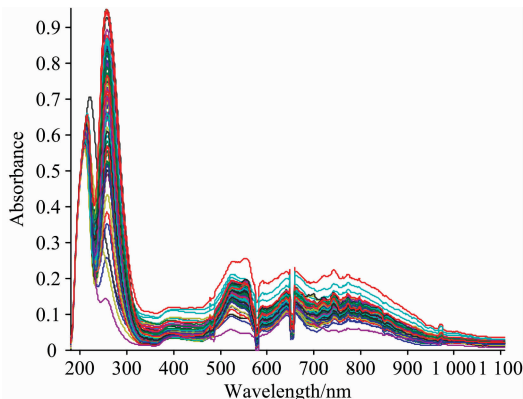


图 4 94 个不同浓度溶液样本的吸收光谱

Fig. 4 Absorption spectrum of 94 samples with different concentrations

### 2.4 样本分类

对待测溶液的吸光度数据进行建模预测之前，使用

Kennard-Stone 算法将其划分为训练集和测试集，其划分原则是把已选取的样本点与其余样本点间的最小直线距离最大化。样本  $i$  与样本  $j$  间的欧氏距离定义如式(10)

$$d_{ij} = \sqrt{\sum_{k=1}^p (r_{ik} - r_{jk})^2} \quad (10)$$

式(10)中， $r_{ik}$ 与 $r_{jk}$ 分别为样本 $i$ 和样本 $j$ 在波长 $k$ 处的吸光度， $p$ 为待测样品溶液光谱的波段数。通过上述算法 94 个溶液样本被划分为 80 个训练样本与 14 个测试样本。

### 3 结果与讨论

#### 3.1 改进 GS-SVR 的参数寻优

改进的 GS 法通过多次减小参数的搜索空间以及修正步长大幅提高了参数寻优速度，配合交叉验证法进行训练集的建模，将获得的最优的  $C$  和  $\sigma$  参数再用于测试样本的浓度预测。惩罚参数  $C$  取值范围设为  $[2^{-9}, 2^9]$ ， $\sigma$  取值范围设为  $[2^{-10}, 2^{10}]$ ，二者的搜索步长初始设为 0.5，依次将其扩大 10 倍，缩小 2 倍，缩小 5 倍，选择 RBF 核函数，得到三次参数寻优的均方误差的 3D 视图如图 5，图 6，图 7 所示。

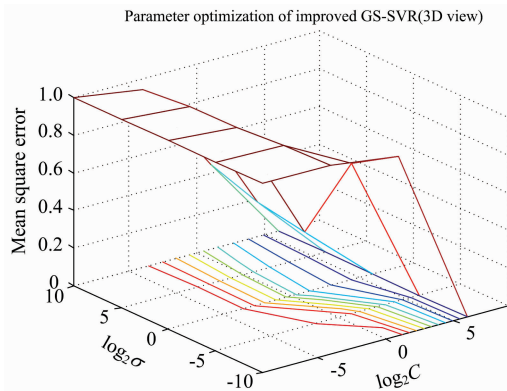


图 5 改进 GS-SVR 第一次参数寻优 3D 视图

Fig. 5 3D view for the first parameter optimization of Improved GS-SVR

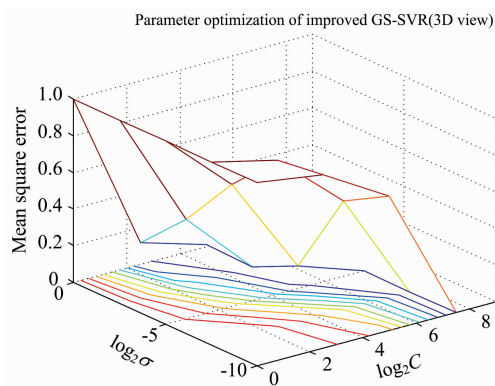


图 6 改进 GS-SVR 第二次参数寻优 3D 视图

Fig. 6 3D view for the second parameter optimization of Improved GS-SVR

会缩小，降低了参数选择的计算量，克服了传统方法非常耗时的缺陷。

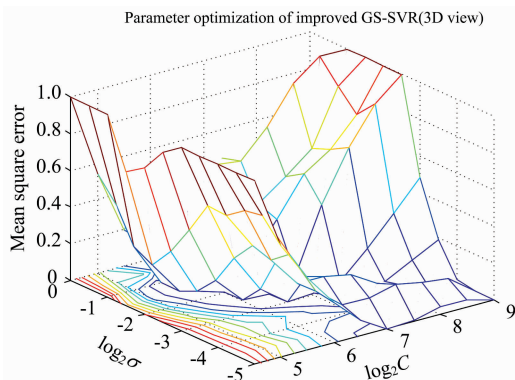


图 7 改进 GS-SVR 第二次参数寻优 3D 视图

Fig. 7 3D view for the third parameter optimization of Improved GS-SVR

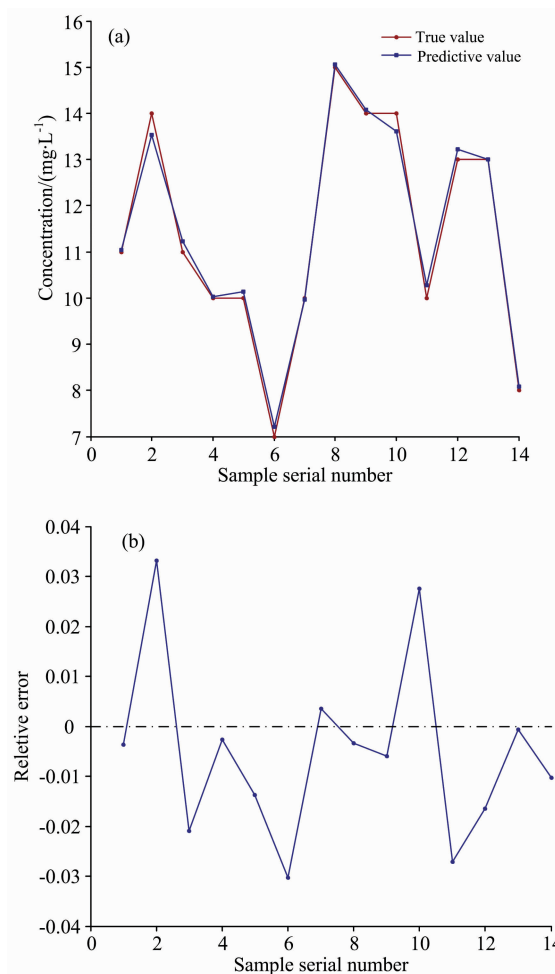


图 8 (a) 改进 GS-SVR 预测值与真实值对比；  
(b) 改进 GS-SVR 预测值相对误差

Fig. 8 (a) Comparison between the predicted value of improved GS-SVR and the real value; (b) Relative error of the predicted value of improved GS-SVR

由图 5—图 7 可知，每一次改变步长后参数搜索空间都

### 3.2 改进 GS-SVR 建模预测

将 3.1 节三次参数寻优的最优参数  $C$  和  $\sigma$  组合 (512, 0.044 2) 应用于支持向量回归模型, 最终获得测试集的预测效果如图 8(a) 所示, 预测结果的相对误差 (relative error, RE) 如图 8(b) 所示。

图 8 和图 9 分析可知, 改进的 GS-SVR 建模的预测效果较好, 其中回归模型的决定系数 (coefficient of determination)  $R^2=0.993 5$ , 预测均方根误差 (root mean square error of prediction, RMSEP)  $RMSEP=0.043 5$ 。14 个样本预测值

的绝对误差 (absolute error, AE) 控制在 0.5 以内, 最大 AE 是 0.465 1, 一半样本的 AE 不超过 0.1。预测值的相对误差最大 3.32%。整个过程平均花费时间为 13 s, 满足快速监测的应用需求。

基于吸光度光谱数据, 同时建立了 BPNN, SVR, GS-SVR, PSO-SVR, GA-SVR 五种预测回归模型, 与文中提出的可变步长 GS-SVR 方法加以对比。每种模型训练 5 次, 在表 1 中统计了硝酸盐浓度预测模型的  $R^2$ , RMSEP, 训练时间。

表 1 不同预测模型评价参数对比

Table 1 Comparison of evaluation parameters of different prediction models

序号	(C, $\sigma$ )	$R^2$	RMSEP	Time/s	(C, $\sigma$ )	$R^2$	RMSEP	Time/s
BPNN				SVR				
1	—	0.955 5	0.059 3	72	—	0.889 7	0.758 9	<1
2	—	0.969 7	0.056 4	62	—	0.889 7	0.758 9	<1
3	—	0.715 4	0.137 0	53	—	0.889 7	0.758 9	<1
4	—	0.918 7	0.116 9	50	—	0.889 7	0.758 9	<1
5	—	0.981 5	0.041 6	101	—	0.889 7	0.758 9	<1
GS-SVR				PSO-SVR				
1	(512, 0.011)	0.985 8	0.177 5	116	(64, 0.385 8)	0.980 9	0.112 8	348
2	(512, 0.011)	0.985 8	0.177 5	111	(51, 0.133 1)	0.981 2	0.158 5	268
3	(512, 0.011)	0.985 8	0.177 5	133	(100, 0.178 6)	0.986 2	0.078 3	300
4	(512, 0.011)	0.985 8	0.177 5	103	(100, 0.010 0)	0.936 3	0.820 7	271
5	(512, 0.011)	0.985 8	0.177 5	138	(100, 0.852 5)	0.966 5	0.206 7	270
GA-SVR				改进 GS-SVR				
1	(77, 0.128 7)	0.983 6	0.106 1	136	(512, 0.044 2)	0.993 5	0.043 5	13
2	(99, 0.038 1)	0.976 7	0.240 8	164	(512, 0.044 2)	0.993 5	0.043 5	13
3	(99, 0.111 6)	0.985 9	0.092 6	128	(512, 0.044 2)	0.993 5	0.043 5	14
4	(99, 0.036 2)	0.975 9	0.247 4	177	(512, 0.044 2)	0.993 5	0.043 5	13
5	(99, 0.038 1)	0.976 6	0.242 2	124	(512, 0.044 2)	0.993 5	0.043 5	13

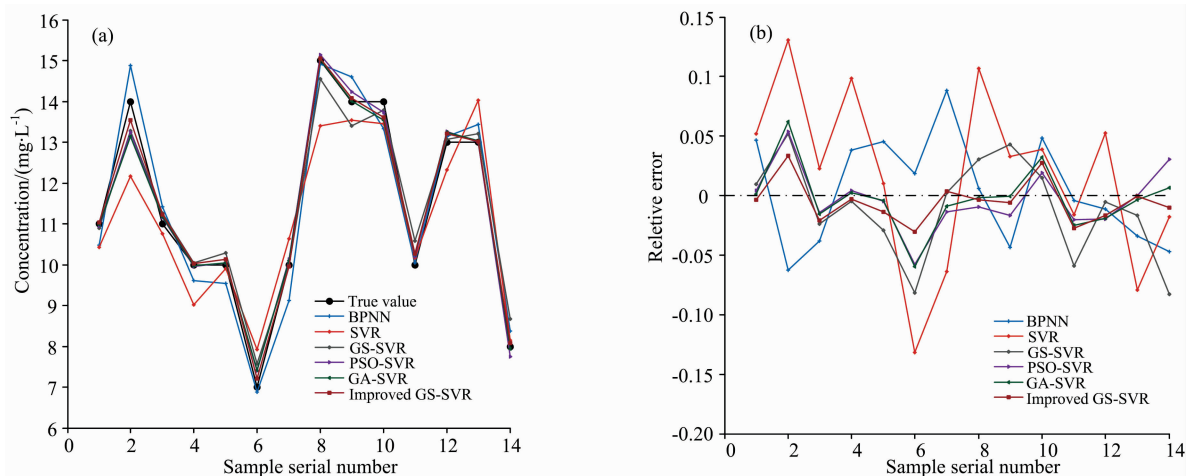


图 9 (a) 不同预测模型的预测值与真实值对比; (b) 不同预测模型的预测值相对误差

Fig. 9 (a) Comparison between predicted value and real value of different prediction models; (b) Relative error of predicted values of different prediction models

分别从预测精度, 预测误差, 训练效率以及稳定性四个方面来对比六种预测模型的性能。由表 1 分析可知, 相比于

BPNN, SVR, GS-SVR, PSO-SVR, GA-SVR 五种预测回归模型 (取 5 次预测精度最高的对比), 改进 GS-SVR 预测精度

最高,  $R^2=0.993\ 5$ , 分别提高了 1.22%, 11.66%, 0.78%, 0.74% 和 0.77%。RMSEP=0.043 5, 分别降低了 -4.36%, 1 644%, 308%, 80% 和 112%。改进 GS-SVR 平均训练时间为 13 s, 相较 BPNN 的 67 s, GS-SVR 的 121 s, PSO-SVR 的 291 s, GA-SVR 的 146 s, 效率分别提升了 4.15 倍, 8.30 倍, 21.38 倍, 10.23 倍, 尽管 SVR 模型训练时间最短, 但其精度不高, 且误差很大。GS-SVR, PSO-SVR, GA-SVR 模型存在一个明显的缺陷就是计算量大, 训练效率较低, 需要等待很长时间才能完成一次参数的解算, 参数寻优过程耗时过长, 尤其后两者预测结果不稳定, 容易陷入局部极优。BPNN 的最大缺点就是预测效果极不稳定, 效果较好时  $R^2$  高达 0.981 5, 训练效果较差时,  $R^2$  只有 0.715 4。究其原因, BPNN 是局部搜索的优化算法, 在计算非线性函数的全局极值时极大概率陷入局部极值, 导致训练精度降低。同时, 网络结构的选取尚无成熟的理论指导, 只能凭经验进行选择, 因此如何采用合适的网络结构也是 BPNN 预测模型训练的关键问题。比较可知, 改进的 GS-SVR 模型综合预测精度, 预测误差, 训练效率以及稳定性方面皆优于其他模型, 在提升预测精度的同时提高了算法的优化效率。

六种硝酸盐浓度回归模型的预测结果与真实值如图 9(a)所示, 相对误差如图 9(b)所示。改进 GS-SVR 最大绝对误差控制在 0.5 以内,  $0.465\ 1$  (改进 GS-SVR)  $< 0.729\ 7$  (GS-SVR)  $< 0.755\ 1$  (PSO-SVR)  $< 0.865\ 9$  (GA-SVR)  $<$

$0.874\ 1$  (BPNN)  $< 1.830\ 3$  (SVR), 为六种模型中最优。相对误差为 3.32% (改进 GS-SVR)  $< 5.77\%$  (PSO-SVR)  $< 6.18\%$  (GA-SVR)  $< 8.30\%$  (GS-SVR)  $< 8.84\%$  (BPNN)  $< 14.12\%$  (SVR), 无论是绝对误差还是相对误差, 改进 GS-SVR 模型都保持在最低水平, 预测效果明显优于其他五种模型。

## 4 结 论

基于精细全光谱结合支持向量回归技术提出了一种基于可变步长的 GS 法优化 SVR, 将其用于混合水体硝酸盐浓度的建模预测。通过可变步长的 GS 寻优算法获得最优参数  $C$  和  $\sigma$  为 (512, 0.044 2), 建立 SVR 硝酸盐浓度预测模型, 模型  $R^2=0.993\ 5$ , RMSEP=0.043 5, 平均训练时间为 13 s。并同目前流行的 BPNN, SVR, GS-SVR, PSO-SVR, GA-SVR 算法对比,  $R^2$  分别提高了 1.22%, 11.66%, 0.78%, 0.74% 和 0.77%, 训练效率分别提升 4.15 倍 (BPNN), 8.30 倍 (GS-SVR), 21.38 倍 (PSO-SVR), 10.23 倍 (GA-SVR)。结果显示, 提出的 GS-SVR 方法不仅提高了模型的预测精度, 而且在寻优效率方面取得了很大的提升, 可为复杂基质水体硝酸盐浓度的快速在线监测提供一种新的有效方法, 具备潜在的应用价值。

## References

- [ 1 ] Liu F, Zheng P, Huang B, et al. International Conference on Computer and Computing Technologies in Agriculture. Springer, Cham, 2015, 619.
- [ 2 ] Causse J, Thomas O, Jung A V, et al. Water Research, 2016, 108: 312.
- [ 3 ] Pons M N, Assaad A, Oucacha C, et al. Ecohydrology & Hydrobiology, 2017, 17(1): 46.
- [ 4 ] MI Yun-ping, WANG Xiao-ping, JIN Xin(宓云骈, 王晓萍, 金 鑫). Journal of Zhejiang University • Engineering Science(浙江大学学报 • 工学版), 2008, 42(5): 790.
- [ 5 ] WANG Li-li, LIU Xian-hua, MI Ma, et al(王莉丽, 刘宪华, 米 玛, 等). Transducer and Microsystem Technologies(传感器与微系统), 2014, 33(1): 75.
- [ 6 ] YANG Qiong, ZHU Qian-hua, REN Peng, et al(杨 琼, 朱乾华, 任 鹏, 等). Journal of Analytical Science(分析科学学报), 2016, 32(4): 485.
- [ 7 ] CHEN Ying, HE Lei, CUI Xing-ning, et al(陈 颖, 何 磊, 崔行宁, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(5): 1489.
- [ 8 ] He Z, Wen X, Liu H, et al. Journal of Hydrology, 2014, 509: 379.
- [ 9 ] Bao Y, Hu Z, Xiong T. Neurocomputing, 2013, 117: 98.
- [ 10 ] Sajjan K S, Kumar V, Tyagi B. International Journal of Electrical Power & Energy Systems, 2015, 73: 200.
- [ 11 ] Gao X, Hou J. Neurocomputing, 2016, 174: 906.

# Research on the Quantitative Analysis Method of Nitrate in Complex Water by Full Scale Spectrum With GS-SVR

LEI Hui-ping<sup>1, 2</sup>, HU Bing-liang<sup>1</sup>, YU Tao<sup>1\*</sup>, LIU Jia-cheng<sup>1</sup>, LI Wei<sup>1, 2</sup>, WANG Xue-ji<sup>1</sup>, ZOU Yan<sup>3</sup>, SHI Qian<sup>3</sup>

1. Xi'an Institute of Optics and Precision Mechanics, Chinese Academy Sciences, Xi'an 710119, China

2. Photoelectrical Engineering Institute, University of Chinese Academy Sciences, Beijing 100049, China

3. Institute of Oceanographic Instrumentation, Shandong Academy of Sciences, Qingdao 266000, China

**Abstract** Nitrate is an important index of water quality monitoring. The high concentration of nitrate in water results in the decrease of biodiversity and the degradation of the ecosystem. Meanwhile, it will cause irreversible harm to human health. Water quality monitoring technology based on the spectrum is the trend of modern water environment monitoring. Compared with the traditional method, nitrate field sampling and laboratory analysis, it has the advantages of simple operation, no pretreatment, fast detection, good repeatability and no pollution. Due to the complexity and diversity of water components, there is a high degree of nonlinearity between water parameters and absorbance. Traditional linear regression prediction models are not applicable, such as single wavelength method, dual wavelength method and partial least square method. Therefore, this paper proposes a new method for the determination of nitrate in water by fine full spectrum combined with the improved variable step grid search algorithm optimized support vector regression (GS-SVR). In cooperation with the college of chemistry and chemical engineering of Shaanxi University of Science and Technology, 94 groups of solution samples with different concentrations were prepared according to different concentration gradients and the experimental requirements by using standard nitrate solution, platinum cobalt standard solution and formazine standard suspension. Firstly, the transmittance spectrum was converted to absorbance, and 94 solution samples were divided into 80 training sets and 14 test sets by Kennard stone algorithm. Secondly, the improved GS algorithm combined with 5-fold cross validation is used to optimize the parameters of SVR by reducing the search range and changing the search step for many times, and the optimal penalty parameters and kernel function width are used to build model, which is used to predict the test set. Meanwhile, the prediction results are compared with those of BPNN, SVR, GS-SVR, PSO-SVR and GA-SVR. The results show that the coefficient of determination  $R^2 = 0.9935$ , root means square of prediction RMSEP = 0.0435. The optimal parameters are (512, 0.0442), and the average training time is 13 s. Compared with the above five prediction models,  $R^2$  increased by 1.22%, 11.66%, 0.78%, 0.74%, 0.77%, training efficiency increased by 4.15 times (BPNN), 8.30 times (GS-SVR), 21.38 times (PSO-SVR), 10.23 times (GA-SVR). The prediction accuracy and training efficiency of the model has been greatly improved, which provides a novel approach basis for rapid and real-time online monitoring of nitrate concentration in the complex water body. This method is also suitable for the establishment prediction models of other water quality parameters.

**Keywords** Fine full spectrum; Nitrate; Improved grid search; Support vector regression

(Received Jan. 6, 2020; accepted Apr. 20, 2020)

\* Corresponding author