

# 嵌入式粒子群-遗传算法的水质 COD 检测特征波长优化算法

漆伟, 冯鹏\*, 魏彪, 郑冬, 于婷婷, 刘鹏勇

重庆大学光电技术及系统教育部重点实验室, 重庆 400044

**摘要** 基于紫外-可见光谱法的水质测量中, 光谱信号易受到系统噪声干扰、悬浮物散射干扰, 且存在信息冗余、多重共线性等特征, 导致水质 COD 测量中特征波长的选取产生较大偏差。因此, 提出了基于嵌入式粒子群-遗传(EPSO\_GA)算法的水质 COD 检测特征波长优化算法, 以提高波长选择精度。为验证检测特征波长优化算法的可行性, 采集了某高校池塘水样、生活污水和排水沟水样的光谱数据, 利用 EPSO\_GA 算法对预处理后的光谱数据选取特征波长。EPSO\_GA 算法采用实数编码方法实现了粒子群(PSO)优化算法和遗传(GA)优化算法的统一编码, 在 PSO 算法中更新粒子时嵌入 GA 算法的选择、交叉、变异等操作, 改善了这两种算法各自在光谱波长特征选取问题上的局限性。将 EPSO\_GA 算法选取的特征波长结合偏最小二乘法(PLS)构建了 EPSO\_GA\_PLS 的水质 COD 预测模型, 并且与传统的 PSO 算法、GA 算法选取特征波长建立的 PSO\_PLS、GA\_PLS 和全光谱构建的 PLS 水质 COD 预测模型做了对比。结果表明: 与 PSO\_PLS、GA\_PLS 和全光谱构建的 PLS 水质 COD 预测模型相比, EPSO\_GA 改善了 PSO 算法和 GA 算法在光谱特征波长选择中早熟和收敛速度慢的问题, 降低了全光谱构建 PLS 水质 COD 预测模型的复杂度, 提高了模型的预测精度。基于 EPSO\_GA 算法建立的 EPSO\_GA\_PLS 水质 COD 预测模型, 均方根误差降到了 0.212 3, 预测精度增加到 0.999 3, 可以快速定量检测水质 COD, 为紫外-可见光谱法测 COD 提供了更好的预测模型。

**关键词** 紫外-可见光谱法; 嵌入式粒子群-遗传算法; 波长特征选择; 偏最小二乘法; 化学需氧量

**中图分类号:** X522 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)01-0194-07

## 引言

化学需氧量(chemical oxygen demand, COD)是衡量水环境污染水平高低的最重要指标之一。利用紫外-可见光谱法构建模型对水质 COD 进行分析, 已广泛应用于在线水质检测领域, 具有实时、在线、原位测量、无二次污染等优点<sup>[1]</sup>。但在实际应用中, 水质检测系统会受到多方面的影响, 其中检测的光谱数据含有大量的噪声对水质 COD 的分析造成严重干扰<sup>[2]</sup>, 因此, 对光谱数据进行去噪是水质 COD 检测的必要前提。在光谱噪声处理方面, 小波变换可以有效去除光谱噪声, 但是小波去噪后依然存在悬浮物散射、光谱信息冗余和微弱噪声等问题。故在小波阈值去噪的基础上, 采用特征波长选取操作, 特征波长选取是利用智能优化算法在原始光谱的整个波段选出最佳的几个特征波长, 用于表征原始光谱信息。

常用的智能优化算法有粒子群算法、遗传算法、蚁群算法等。宾俊等<sup>[3]</sup>在近红外光谱波长选择的研究中, 提出了以蚁群优化(ACO)、遗传优化(GA)、粒子群优化(PSO)、随机青蛙(RF)和模拟退火(SA)等智能优化算法对特征波长进行选择的思想, 找出了总氮和烟碱对应最佳光谱波长组合, 结果表明利用智能优化算法选择特征波长可以排除噪声干扰, 建立的模型可以快速准确的分析出总氮和烟碱的含量。Tang 等<sup>[4]</sup>采用了基于遗传算法和粒子群算法优化的遗传算法-粒子群优化(GA-PSO)算法, 成功实现了在消光光谱测量中选择最佳波长, 该方法具有抗噪声的优点。Ying 等<sup>[5]</sup>将连续投影算法(SPA)与四种群智能优化算法相结合, 包括 GA 算法, PSO 算法, 群搜索优化器(GSO)和萤火虫算法(FA), 用于提取有效波长变量, 可以快速有效地区分苹果汁的掺假。Li 等<sup>[6]</sup>利用差分进化(DE)的波长选择方法, 消除了太赫兹时域光谱的散射和噪声干扰问题。Zhang 等<sup>[7]</sup>提出了使用等距组合多元线性回归(ECMLR)方法进行波长选择, 在时

收稿日期: 2019-12-17, 修订日期: 2020-04-16

基金项目: 国家自然科学基金项目(61401049), 重庆市社会事业与民生保障科技创新专项(cstc2017shmsA00004), 中央高校基本科研业务费医工结合专项(2019CDYGYB019)资助

作者简介: 漆伟, 1995年生, 重庆大学光电技术及系统教育部重点实验室硕士研究生 e-mail: qiwei@cqu.edu.cn

\* 通讯作者 e-mail: coe-fp@cqu.edu.cn

间相关的验证方面, ECMLR 方法的预测效果比使用全光谱构建预测模型的预测效果稍好。总结现有的波长选取方法, 大多只是利用单一的智能优化算法进行波长选择, 但单一的智能算法却存在一定的缺陷, 很少有人将多种智能算法融合成新型的智能算法进行波长选择。

鉴于此, 提出了嵌入式粒子群-遗传(EPSCO\_GA)新型算法用于选取最佳特征波长, 结合偏最小二乘法(PLS)<sup>[8]</sup>对紫外-可见光谱下的 COD 进行了定量分析。

### 1 嵌入式粒子群-遗传(EPSCO\_GA)算法原理

粒子群(PSO)算法是一种进化计算技术, 源于对鸟群捕食行为的研究, 其基本思想是通过群体中粒子之间的协作和信息共享来寻找最优解<sup>[9]</sup>。遗传(GA)算法是模拟达尔文进化论的自然选择和遗传学机理的生物进化过程的计算模型, 是一种模拟自然进化过程搜索最优解的方法<sup>[10]</sup>, 利用选择、交叉和突变等进化因子使得种群的适应度不断增强, 从而达到优胜劣汰的目的。PSO 算法具有自组织和进化性以及记忆功能, 所有粒子都保存优解的相关知识, PSO 算法注重整体, 侧重于整体上的寻优能力, 优化速度较快, 局部搜索能力较差, 精度不高易陷入局部极小值振荡或在寻优过程中全部收敛于全局极值, 而忽略局部最优。GA 算法中变异算子是对群体中的部分个体实施随机变异, 与历史状态和当前状态无关, 在局部信息交流上具有良好的特性, 但是效率比其他优化算法低, 计算量大、运行时间长。两者算法都容易陷入“早熟”的现象, 因此在 PSO 算法中单个粒子更新时嵌入 GA 算法的选择操作、整个种群更新时嵌入交叉操作和变异操作, 形成嵌入式粒子群-遗传(EPSCO\_GA)优化算法。EPSCO\_GA 算法将已经产生的优秀个体与随机产生的微粒个体进行交叉变异操作, 使产生的微粒跳出局部极值点, 并且又保留了优秀个体的特性, 可以加强算法在局部和整体上的寻优能力, 加快搜索的速度, 在解决优化问题中更占优势。EPSCO\_GA 算法的主要步骤如下:

①对 PSO 算法与 GA 算法都采用实数编码, 实现编码上的统一;

②定义  $n$  个粒子组成一个群体对  $D$  维问题空间进行搜索, 根据邻近粒子和自身的经验不断调整其位置  $x$  和速度  $v$ ; 第  $i$  个粒子在  $t$  时刻的位置表示为

$$x_t^i = (x_t^{i1}, x_t^{i2}, \dots, x_t^{iD}), (i = 1, 2, \dots, n) \quad (1)$$

第  $i$  个粒子在  $t$  时刻的速度表示为

$$v_t^i = (v_t^{i1}, v_t^{i2}, \dots, v_t^{iD}), (i = 1, 2, \dots, n) \quad (2)$$

③PSO 算法与 GA 算法都采用同一种目标函数, 实现评价上的统一, 根据求解问题定义的目标函数如式(3)所示

$$f(x) = \min(\text{RMSE}) \quad (3)$$

RMSE 是选取特征波长 COD 真实值与模型解算出测量值之间的均方根误差。

④根据初始化种群参数和个体的位置、速度信息, 更新 PSO 粒子, 同时根据目标函数计算当前个体最优极值和全局最优极值。粒子  $i$  的速度由当前最优粒子的位置 Pbest 和所有粒子中的最优位置 Gbest 来更新迭代, 从而更新粒子  $i$  的

位置, 更新公式如式(4)和式(5)

$$v_{t+1}^i = \omega \times v_t^i + c_1 \times r_1 \times (P_{\text{best}_t}^i - x_t^i) + c_2 \times r_2 \times (G_{\text{best}_t}^i - x_t^i) \quad (4)$$

$$x_{t+1}^i = x_t^i + v_{t+1}^i \quad (5)$$

⑤嵌入 GA 算法的选择算子和交叉算子, 根据计算出的个体极值选择出优良的个体, 再根据交叉概率交换粒子之间的基因, 将有益的基因组合, 提高个体的目标函数值, 此时产生的新个体的位置表示为

$$x_{t+1}^i = \alpha x_t^i + (1 - \alpha)x_t^j, (i, j = 1, 2, \dots, n, i \neq j)$$

$$x_{t+1}^j = \alpha x_t^j + (1 - \alpha)x_t^i, (i, j = 1, 2, \dots, n, i \neq j) \quad (6)$$

⑥根据自然变异的思想, 嵌入 GA 算法的变异算子, 对粒子群按照一定的变异概率进行变异操作, 保证粒子群的多样性, 产生子代粒子种群;

⑦将粒子群父代精英个体染色体与子代个体染色体合并, 孕育出适应度更强的粒子种群;

⑧更新当前目标函数值, 更新最优粒子的位置 Pbest 和所有粒子中的最优位置 Gbest;

⑨重复步骤④—⑧, 直到满足结束条件, 结束算法, 并记录当前粒子的目标函数值和位置。

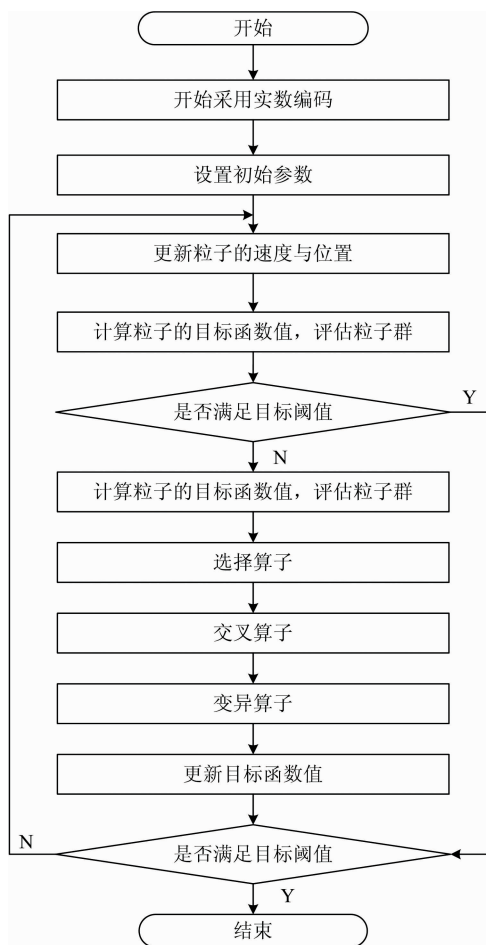


图 1 嵌入式粒子群-遗传优化算法流程图  
Fig. 1 Flow chart of embedded particle swarm optimization-genetic algorithm

在式(4)中  $\omega$  是惯性权重, 决定先前速度对当前速度的影响, 取值一般在  $[0, 1]$  范围。  $c_1$  和  $c_2$  为学习因子, 经验设置  $c_1 = c_2 = 2$ ,  $r_1$  和  $r_2$  为随机因子, 在  $[0, 1]$  区间随机取值,  $P_{\text{best}}$  是单个粒子在搜索过程中的历史最优值,  $G_{\text{best}}$  是粒子群在搜索过程中的历史最优值。在式(6)中,  $\alpha$  为交叉概率, 一般设置  $\alpha = 0.7$ 。EPSO\_GA 算法的流程图如图 1 所示。

## 2 实验部分

### 2.1 光谱数据采集

利用化学方法, 使用重铬酸钾溶液对某高校的池塘水样、生活污水和排水沟水样分别进行滴定, 温度为  $15\text{ }^\circ\text{C}$ , 溶液  $20\text{ mL}$ , 测得池塘水样 COD 浓度为  $10.4\text{ mg}\cdot\text{L}^{-1}$ 、生活污水水样 COD 浓度为  $40\text{ mg}\cdot\text{L}^{-1}$ 、排水沟水样 COD 浓度为  $12.6\text{ mg}\cdot\text{L}^{-1}$ , 接着对每种水样进行稀释, 每次加蒸馏水  $5\text{ mL}$ , 共得到 46 组 COD 标准溶液。在水体环境中, 有机物对紫外-可见光的吸收遵循 Lambert Beer 定律<sup>[11]</sup>, 因此, 根据紫外-可见光的吸收度确定特定条件下的水质 COD 浓度值。紫外-可见吸收光谱法是基于测量水体中有机物对紫外-可见光的吸收强度来确定 COD 值的方法<sup>[12]</sup>, 使用海洋光学生产的 USB2000+工业级光谱仪和长寿命、无需预热的脉冲氙灯光源对 46 组 COD 标准溶液进行光谱采集, 可得到溶液在紫外-可见光波段的吸收光谱数据。部分数据如表 1 所示。从中随机选择 32 组数据用作模型训练数据集, 剩余 14 组数据用作测试集。

表 1 紫外-可见吸收光谱数据集

Table 1 UV-Visible absorption spectrum data set

样本	COD/ ( $\text{mg}\cdot\text{L}^{-1}$ )	198.98 (nm)	199.78 (nm)	200.58 (nm)	...	993.06 (nm)	993.78 (nm)
1	10.40	0.032	0.049	0.032	...	0.176	0.000
2	9.45	-0.005	0.016	0.013	...	0.000	0.477
3	8.67	0.017	0.009	0.013	...	0.000	0.117
...	...	...	...	...	...	...	...
43	10.67	0.165	0.178	0.157	...	0.000	0.000
44	10.00	0.165	0.167	0.165	...	-0.301	0.000
45	9.41	0.154	0.156	0.153	...	0.000	0.000
46	8.89	0.163	0.152	0.153	...	-0.477	-0.602

### 2.2 光谱数据预处理

由光谱仪采集到的水质吸收光谱存在大量噪声, 既有低频的基线干扰, 又有高频噪声的影响<sup>[13]</sup>。因此在使用吸收光谱数据构建预测模型之前需对光谱数据进行预处理。针对噪声干扰, 预处理可以滤除部分噪声, 消除相关因素对模型预测结果的影响。采取小波阈值去噪方法对现有的光谱数据进去降噪, 选取“sym4”小波, 将光谱信号经“sym4”小波分解得到相应的小波系数, 运用极值阈值估计方法设定阈值, 定义大于阈值的小波系数为有用信号产生的, 将其保留, 小于阈值的小波系数认定为噪声信号, 将其滤除。最后将分解之后保留的小波进行重构, 即可得到降噪之后的光谱数据。原始光谱图和预处理之后的光谱图分别如图 2 所示。在处理之后的光谱图中可以明显观察到吸收峰主要分布在  $230\sim 550$  和

$620\sim 960\text{ nm}$  这两个波段。但是实际上, 去噪后的吸收光谱仍存在水中悬浮物的散射干扰<sup>[14]</sup>, 因此, 需要对吸收光谱进行散射校正处理, 便于后续建模的精度提升。早期研究多采用  $254\text{ nm}$  波长的吸光度为模型输入量, 但在实时在线测量时, 由于水中悬浮物散射对  $254\text{ nm}$  吸光度产生干扰, 导致测量结果偏差较大。为了补偿水中悬浮物对吸光度造成的散射影响, 本工作采用多波长的吸光度进行建模, 即利用智能优化算法选择  $N$  个特征波长作为模型输入量。

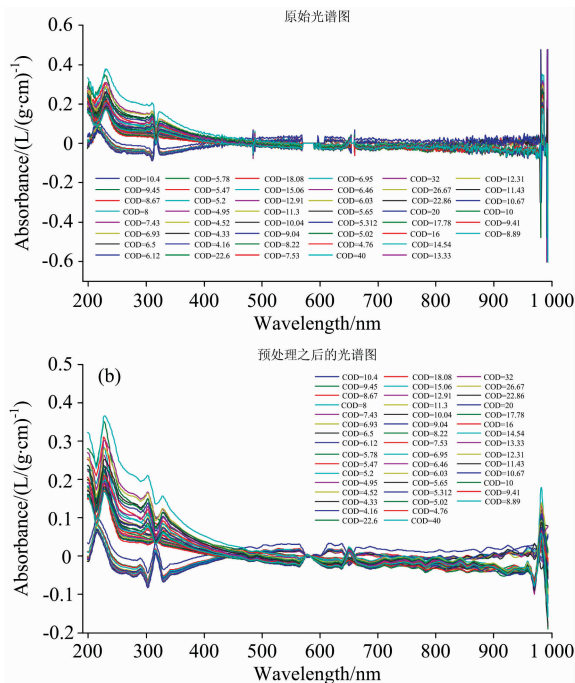


图 2 原始光谱与去噪之后的对比图

Fig. 2 Comparison between the original and denoising spectra

### 2.3 特征波长选择

观察去噪之后的光谱, 发现吸收光谱中存在多个吸收峰。受益于主成分分析(principal component analysis, PCA) 方法的思想, 对所有数据吸收光谱的吸收峰进行筛选, 结果表明每条数据的 6 个主吸收峰可以贡献所有吸收峰 98% 的信息, 故使用智能优化算法选择 6 个特征波长作为模型输入。利用 PSO 算法选择特征波长, 选择均方根误差(root mean square error, RMSE)作为目标函数, 迭代 200 次, 不断更新粒子的位置与速度, 选取的特征波长为:  $289.76, 386.67, 445.23, 542.68, 777.37$  和  $787.80\text{ nm}$ , 此时的目标函数值为  $0.1653$ 。利用 GA 算法选择特征波长时, 同样以均方根误差(root mean square error, RMSE)作为目标函数, 设置迭代次数为 200, 选取的特征波长为:  $289.76, 372.56, 486.36, 507.24, 539.61$  和  $883.87\text{ nm}$ , 此时的目标函数值为  $0.1547$ 。PSO 算法和 GA 算法虽都具有深刻的智能背景, 但在某些方面依然存在一定缺陷。在 GA 算法中, 染色体之间相互共享信息, 整个种群是比较均匀地向最优区域移动<sup>[15]</sup>。而 PSO 算法的性能依赖于全局搜索和局部搜索之间(exploration and exploitation)的平衡能力, 即搜索空间的全

局搜索能力和快速收敛于有希望的区域的能力, 分别对应  $G_{best}$  变量和  $P_{best}$  变量<sup>[16]</sup>, 虽然搜索速度较快, 但是收敛性较差。PSO 算法具有记忆, 能将历史优解都保存, 但 GA 算法没有记忆, 以前的知识随着种群的变化容易被破坏。为了弥补 PSO 算法与 GA 算法的缺陷, 采用 EPSO\_GA 算法选择特征波长, EPSO\_GA 算法将 GA 算法的选择算子、交叉算子、变异算子嵌入到 PSO 算法中, 不仅具备了 PSO 算法的搜索速度和全局寻优能力, 还有着 GA 算法的局部寻优能力。通过不断更新以均方根误差 (root mean square error, RMSE) 作为目标函数的函数值, 实现粒子群的整体优化。实验设置迭代次数为 200 次, 寻得粒子群中的最优个体, 得到最佳的 6 个吸收波长, 6 个波长分别为: 285.80, 372.56, 496.42, 544.99, 638.19 和 945.96 nm, 此时的目标函数值为 0.106 9。对比 EPSO\_GA 算法、PSO 算法、GA 算法选择特征波长的迭代流程和实验结果, 迭代流程比较图如图 3 所示, 实验测试结果如表 2 所示, 可以明显发现 EPSO\_GA 既有 PSO 算法的收敛速度, 又有 GA 算法的精度, EPSO\_GA 算法在迭代 81 次之后就接近收敛, 而且精度得到了大大的提升, 达到了 0.106 9。分析选取的特征波长, 基于 PSO 算

法和基于 GA 算法筛选出的波长组合表征的光谱的信息不完善, 不是最优组合, 基于 EPSO\_GA 算法选取的特征波长目标函数值最小, 在吸收光谱上分配的更加均匀, 更具合理性, 故认为基于 EPSO\_GA 算法筛选出的波长组合对光谱散射问题更敏感, 几乎可以表征完整的光谱信息, 跳出了局部最优解, 找到了近似于全局最优解的波长组合。

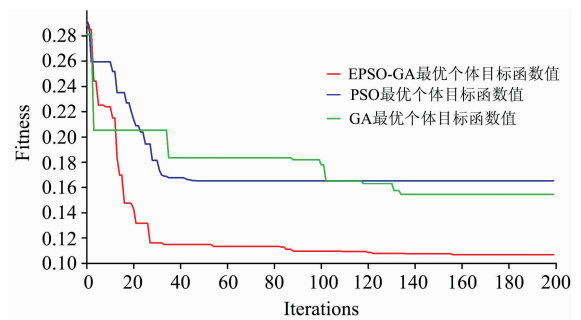


图 3 EPSO\_GA, PSO, GA 算法迭代流程比较  
Fig. 3 Comparison of iterative processes of EPSO\_GA, PSO and GA algorithms

表 2 EPSO\_GA, PSO, GA 算法实验结果

Table 2 Experimental results of EPSO\_GA, PSO and GA algorithms

算法	特征波长 1 /nm	特征波长 2 /nm	特征波长 3 /nm	特征波长 4 /nm	特征波长 5 /nm	特征波长 6 /nm	迭代次数 /次	RMSE
PSO	289.76	386.67	445.23	542.68	777.37	787.80	48	0.165 3
GA	289.76	372.56	486.36	507.24	539.61	883.87	131	0.154 7
EPSO_GA	285.80	372.56	496.42	544.99	638.19	945.96	81	0.106 9

### 3 结果与讨论

将全部特征波长和根据 PSO 算法、GA 算法、以及 EPSO\_GA 算法选出的特征波长采用 PLS 方法根据训练集的 32

组数据分别建立 PLS, PSO\_PLS, GA\_PLS, EPSO\_GA\_PLS 的水质 COD 预测模型, 再利用构建的模型对 14 组测试数据进行预测, 得到各个模型下水质 COD 预测值, 表 3 为各个模型下的 COD 预测值以及与 COD 真实值的误差数据。图 4 是各个模型下水质 COD 预测值与真实值比较的曲线图。

表 3 各个模型下 COD 预测值与 COD 真实值的误差

Table 3 Errors between COD predicted values and true COD values for each model

COD 真实值	COD 预测值 (PLS)	COD 预测值 (PSO_PLS)	COD 预测值 (GA_PLS)	COD 预测值 (EPSO_GA_PLS)	误差 (PLS)	误差 (PSO_PLS)	误差 (GA_PLS)	误差 (EPSO_GA_PLS)
14.54	13.06	14.32	14.59	14.54	-1.48	-0.22	0.05	0.00
10.67	11.16	10.91	10.62	10.77	0.49	0.24	-0.05	0.10
6.12	5.79	5.76	6.54	6.19	-0.33	-0.36	0.42	0.07
16.00	15.75	16.34	16.06	16.03	-0.25	0.34	0.06	0.03
32.00	31.59	31.82	31.59	31.64	-0.41	-0.18	-0.41	-0.36
7.53	6.85	7.81	7.58	7.38	-0.68	0.28	0.05	-0.15
22.6	24.91	22.86	22.98	22.92	2.31	0.26	0.38	0.32
4.76	5.53	4.67	4.41	4.63	0.77	-0.09	-0.35	-0.13
15.06	13.89	15.16	15.51	15.12	-1.17	0.10	0.45	0.06
5.02	4.89	4.6	5.25	5.05	-0.13	-0.42	0.23	0.03
4.16	2.95	4.48	4.05	4.11	-1.21	0.32	-0.11	-0.05
5.312	6.07	5.31	5.32	4.79	0.758	-0.002	0.008	-0.522
10.04	8.43	10.52	10.03	10.17	-1.61	0.48	-0.01	0.13
5.47	6.56	5.73	5.82	5.69	1.09	0.26	0.35	0.22

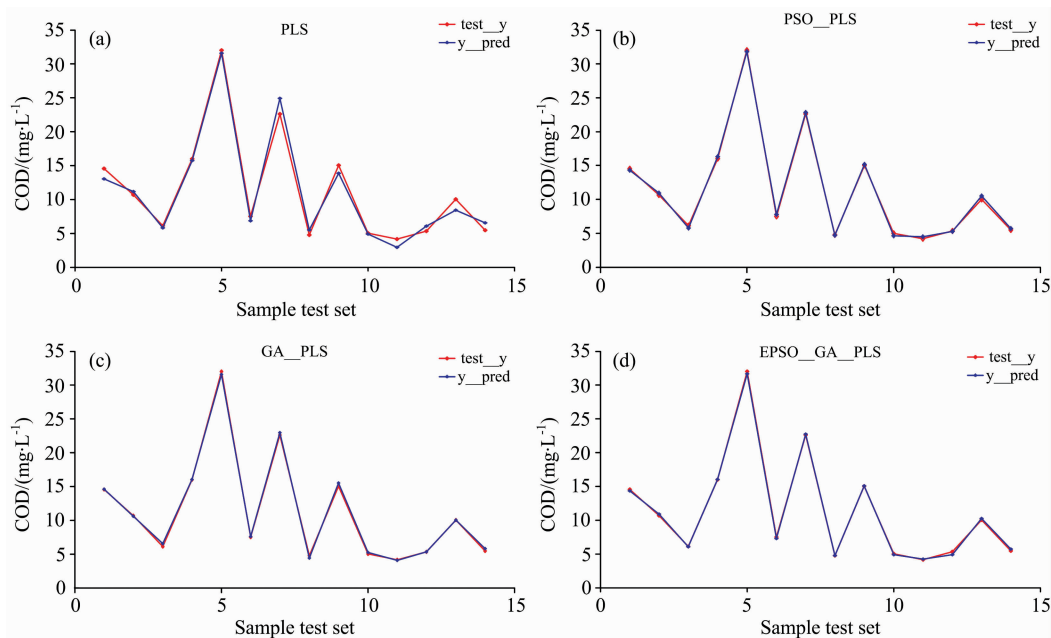


图 4 模型效果展示图

(a): 全光谱 PLS; (b): PSO\_PLS; (c): GA\_PLS; (d): EPSO\_GA\_PLS

Fig. 4 Performance of each model for COD prediction

(a): Full-spectrum PLS; (b): PSO\_PLS; (c): GA\_PLS; (d): EPSO\_GA\_PLS

表 4 各个模型的指标性能

Table 4 Indicator performance of each model

模型	RMSE	MAE	MedAE	$R^2$
PLS	1.081 9	0.906 5	0.761 2	0.980 7
PSO_PLS	0.284 1	0.254 7	0.261 5	0.998 6
GA_PLS	0.269 4	0.170 5	0.209 4	0.998 8
EPSO_GA_PLS	0.212 3	0.155 2	0.112 7	0.999 3

根据所得数据再计算出各个模型的均方根误差 (RMSE)、平均绝对误差 (MAE)、中值绝对误差 (MedAE)、确定系数 ( $R^2$ )、作为模型的评估标准, RMSE 指标是用来衡量预测值同真实值之间的偏差, MAE 指标用于评估预测结果和真实数据的接近程度, MedAE 反映模型的对异类预测的情况, 值小说明拟合效果越好,  $R^2$  用来度量测试样本是否可能通过模型被很好地预测, 其值越大说明模型的效果越强, 值最大为 1, 表 4 反映了各个模型指标性能。观测各个模型下水质 COD 预测值与真实值相比较的曲线图, 对比各个模型的性能指标, 可直观发现在测试数据集中由全光谱建立的 PLS 模型效果不理想, 均方根误差太大, 若将模型应用到实际场景中将会存在一定的问题。而利用特征波长选择方法构建的 PSO\_PLS, GA\_PLS, EPSO\_GA\_PLS 模型较全光谱建立的模型而言, 选用的特征波长大幅度减少, 降低了光谱数据的冗余度, 弱化了光谱散射影响, 使建模更简单, 稳定性更强, 效果更优。对比利用特征波长选择方法构建的 3 个模型, 则以 EPSO\_GA\_PLS 模型效果最优, 它的均方根误差达到了 0.212 3, 水质 COD 的预测准确率突破到了 99.93%,

同时异类预测的情况更少。

## 4 结论

将 GA 算法嵌入到 PSO 算法, 得到新型 EPSO\_GA 算法, 将 EPSO\_GA 应用到水质 COD 检测特征波长优化算法研究中, EPSO\_GA 算法在选取特征波长的过程中具有高收敛性、高精度解的优点。结合目前光谱法测水质 COD 最常用的 PLS 定量分析模型, 建立了 EPSO\_GA\_PLS 的水质 COD 预测模型, 比较了其与全光谱, PSO 算法选择特征波长和 GA 算法选择特征波长构建的 PLS 水质 COD 预测模型的不同, 分析了 EPSO\_GA 算法选择的特征波长组合优于 PSO 算法和 GA 算法选择的波长组合的原因。结果表明, 利用 EPSO\_GA 算法选择的特征波长组合表征原始光谱的信息更加完善, 比 PSO 算法和 GA 算法选择的特征波长组合的适应性更强, 是紫外-可见光谱法水质 COD 检测特征波长的最佳选择, 其选取的波长组合建立的 PLS 水质 COD 预测模型的复杂度较全光谱构建的模型而言大大减小, 减少了特征变量的个数, 解决了光谱噪声干扰和浊度散射问题, 提高了模型的预测精度。综上所述: 基于 EPSO\_GA 算法建立的 EPSO\_GA\_PLS 的水质 COD 预测模型, 优化了单一智能算法的早熟和收敛速度慢的问题, 弱化了噪声干扰、光谱散射干扰, 数据冗余的问题, 实现了紫外-可见光谱法下 COD 的精准检测。故本文提出的用 EPSO\_GA 算法选择特征波长建立的 PLS 水质 COD 预测模型可以为实际场景快速精准定量检测水质 COD 提供有效方案, 为水环境保护事业做出贡献。

## References

- [ 1 ] ZHANG Zheng, WEI Biao, TANG Ge, et al(张 峥, 魏 彪, 汤 戈, 等). Laser Journal(激光杂志), 2016, 37(4): 21.
- [ 2 ] TANG Bin, WEI Biao, MAO Ben-jiang, et al(汤 斌, 魏 彪, 毛本将, 等). Progress in Laser and Optoelectronics(激光与光电子学进展), 2014, 51(4): 201.
- [ 3 ] BIN Jun, FAN Wei, ZHOU Ji-heng, et al(宾 俊, 范 伟, 周冀衡, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(1): 95.
- [ 4 ] Tang Ge, Wei Biao, Wu Decao, et al. Journal of Applied Spectroscopy, 2018, 85(1): 109.
- [ 5 ] Li Ying, Guo Yajing, Liu Chang, et al. Food Analytical Methods, 2017, 10(6): 1965.
- [ 6 ] Li Zhi, Chen Weidong, Lian Feiyu, et al. Applied Spectroscopy, 2017, 71(12): 2653.
- [ 7 ] Zhang Linna, Sun Meixiu, Wang Zhennan, et al. Analytical Methods, 2016, 8(2): 381.
- [ 8 ] Moreno-Martin Gustavo, León-González María Eugenia, Madrid Yolanda. Talanta, 2018, 188: 393.
- [ 9 ] HU Xiao-xi, ZHOU Chen, ZHAO Jun, et al(胡小希, 周 晨, 赵 军, 等). Chinese Journal of Radio Science(电波科学学报), 2018, 33(6): 708.
- [10] LIU Hao-ran, ZHAO Cui-xiang, LI Xuan, et al(刘浩然, 赵翠香, 李 轩, 等). Journal of Instrumentation(仪器仪表学报), 2016, 37(7): 1573.
- [11] Maentele Werner, Deniz Erhan. Spectrochimica Acta, Part A. Molecular and Biomolecular Spectroscopy, 2017, 173: 965.
- [12] Wu Shaofeng, Guan Li, Li Dongbo, et al. Optik: Zeitschrift Fur Licht-und Elektronenoptik; Journal for Light-and Electronoptic, 2018, 174: 591.
- [13] ZHAO Jing-xiao, WEI Biao, TANG Bin, et al(赵敬晓, 魏 彪, 汤 斌, 等). Environmental Science & Technology(环境科学与技术), 2015, 38(6): 83.
- [14] WU De-cao, WEI Biao, TANG Ge, et al(吴德操, 魏 彪, 汤 戈, 等). Acta Optica Sinica(光学学报), 2017, 37(2): 356.
- [15] LI Xin-ran, JIN Yan-xia(李欣然, 靳雁霞). Computer Systems & Applications(计算机系统应用), 2012, 21(8): 127.
- [16] Armin Azad, Hojat Karami, Saeed Farzin, et al. Water Science and Engineering, 2019, 12(1): 45.

# Feature Wavelength Optimization Algorithm for Water Quality COD Detection Based on Embedded Particle Swarm Optimization-Genetic Algorithm

QI Wei, FENG Peng\*, WEI Biao, ZHENG Dong, YU Ting-ting, LIU Peng-yong

Key Laboratory of Optoelectronic Technology and Systems, Ministry of Education, Chongqing University, Chongqing 400044, China

**Abstract** In the water quality measurement based on UV-Visible spectroscopy, the spectral signal is easily disturbed by system noise and the scattering of suspended solids, and there are information redundancy, multicollinearity and other characteristics, resulting in a large deviation in the selection of characteristic wavelength in the COD measurement of water quality. Therefore, this paper proposes an optimization algorithm of water COD detection characteristic wavelength based on embedded particle swarm genetic (EPSO\_GA) algorithm to improve the accuracy of wavelength selection. In order to verify the feasibility of the optimization algorithm for detecting characteristic wavelength, spectral data of water samples from a university pond, domestic sewage and drainage ditch were collected, and EPSO\_GA algorithm was used to select characteristic wavelength from the pre-processed spectral data. EPSO\_GA algorithm adopts the real coding method to realize unified coding of particle swarm optimization (PSO) algorithm and genetic optimization (GA) algorithm. The operations of selection, crossover, and mutation of the GA algorithm are embedded when the particles are updated in the PSO algorithm, which improves the limitations of these two algorithms in spectral wavelength feature selection. The characteristic wavelength selected by EPSO\_GA algorithm was combined with the partial least square method (PLS) to construct the water COD prediction model of EPSO\_GA\_PLS, and compared with the traditional PSO algorithm and the GA algorithm, the PSO\_PLS, GA\_PLS established by the characteristic wavelength and the PLS water quality COD prediction model constructed by the full spectrum are compared. The results showed that compared with PSO\_PLS, GA\_PLS and the PLS water quality COD prediction model constructed by full-spectrum, EPSO\_GA improves the precocious and slow convergence speed of PSO and GA in spectral characteristic wavelength selection,

and reduces the complexity of constructing PLS water quality COD prediction model in the whole spectrum, the prediction accuracy of the model is improved. In the EPSO\_GA\_PLS water quality COD prediction model established based on EPSO\_GA algorithm, the root-mean-square error decreased to 0.212 3, and the prediction accuracy increased to 0.999 3, which can quickly and quantitatively detect water quality COD, provides a better prediction model for COD measurement by UV-Visible spectroscopy.

**Keywords** Ultraviolet-Visible spectroscopy; Embedded particle swarm optimization-genetic algorithm; Wavelength feature selection; Partial least squares method; Chemical oxygen demand

(Received Dec. 17, 2019; accepted Apr. 16, 2020)

\* Corresponding author