

# 基于 SPA 和 多分类 SVM 的 紫外-可见光光谱 饮用水有机污染物判别方法研究

黄平捷, 李宇涵, 俞巧君, 王 柯, 尹 航, 侯迪波\*, 张光新

浙江大学控制科学与工程学院, 工业控制技术国家重点实验室, 浙江 杭州 310027

**摘 要** 快速、有效地识别饮用水中污染物类别对于降低突发饮用水污染事件影响十分重要。目前基于紫外-可见光光谱法的饮用水污染物判别模型大多使用主成分分析(PCA)进行特征提取,然而,对于光谱相似度较高的有机污染物,仅从数据驱动的角度提取其方差最大的方向作为特征进行识别效果往往不佳。针对有机污染物光谱数据多重共线性以及谱峰重叠干扰的问题,开展了基于连续投影算法(SPA)和多分类支持向量机(M-SVM)的紫外-可见光光谱饮用水有机污染物判别方法研究。首先,使用紫外光谱仪测量苯酚、对苯二酚、间苯二酚和间苯二胺的原始光谱数据并进行预处理,在对四种污染物进行了波长与浓度的相关关系对比分析后,发现苯酚和间苯二酚、对苯二酚和间苯二胺的谱峰重叠较为严重;特征提取时,引入 SPA 筛选有机污染物紫外-可见光光谱数据的特征波长组合,并对不同波长个数时的光谱吸光度进行多元线性回归分析,选取对应最小预测标准偏差的参数及波段组合作为最优参数组合;基于最优特征波长组合,构建基于多分类 SVM 的饮用水有机污染物分类识别模型;最后,对比分析了全光谱、PCA 及 SPA 特征提取后的光谱数据在不同分类方法及不同污染物浓度下的分类效果,进一步说明了 SPA 的适用性和稳定性。实验结果表明,SPA 作为一种提取光谱数据原始特征波段的方法,可以有效的对有机污染物的紫外-可见光光谱进行特征提取,提升不同物质之间的差异,在一定程度上消除多重共线性和谱峰重叠干扰,从而提高分类模型的准确率。该方法对于解决饮用水中谱峰重叠的污染物类型判别问题具有参考价值。

**关键词** 紫外-可见光光谱; 有机污染物判别; 连续投影算法; 多分类支持向量机

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)07-2267-06

## 引 言

近年来,饮用水污染事件频发,给居民正常生活带来很大影响<sup>[1]</sup>。为了及时应对此类污染事件,需要在线监测水质并识别污染物,以防止此类事件造成进一步的影响。相较于传统检测手段,基于光谱法的水质检测方法因其具有快速、无需试剂、样本无需预处理、设备简单、检测原理成熟等优点,从而吸引了众多学者针对基于光谱法的水体污染物识别技术展开研究<sup>[2-3]</sup>。在众多光谱波段中,紫外-可见光光谱波段因其检测速度快、检测原理成熟,并且可以间接的反映水中有机物、悬浮物等理化参数,从而受到广泛的关注<sup>[4-5]</sup>。

现有的基于紫外-可见光光谱的水体污染物检测识别方

法大多是基于 PCA 进行特征提取。例如,赵明富<sup>[6]</sup>等利用水体紫外-可见光谱波段,使用主成分分析并结合 Fisher 判别法对水质进行了较好的分类;Hou<sup>[7]</sup>等提出了一种融合主成分分析法和卡方分布的水质异常检测算法,通过紫外光学传感器实现对水质异常状况的判别;郭冰冰<sup>[8]</sup>等提出一种基于基线校正和主元分析的水质异常检测方法,利用主成分分析方法提取正常水质的紫外-可见光光谱的特征并利用统计量识别异常光谱。PCA 对紫外-可见光光谱数据进行了降维并提取了数据的特征,有效的提高了污染物识别的效果,但仍然存在问题,PCA 仅从数据的方差最大的方向对数据进行压缩,没有考虑到紫外-可见光光谱的实际的物理特征,因此,寻找一种更合适的污染物光谱数据特征提取方法来解决饮用水中谱峰重叠的有机污染物分类问题显得十分迫切和

收稿日期: 2019-04-18, 修订日期: 2019-08-09

基金项目: 国家自然科学基金项目(61803333, U1509208), 国家重点研发计划课题(2017YFC1403801), 中央高校基本科研业务费专项资金项目(2019FZA5007)资助

作者简介: 黄平捷, 1974 年生, 浙江大学控制科学与工程学院副教授 e-mail: huangpingjie@zju.edu.cn

\* 通讯联系人 e-mail: houdb@zju.edu.cn

必要。

针对现有紫外-可见光光谱法对于有机污染物特征提取效果不佳的问题,本文确定了有机污染物之间吸收峰的重叠范围以及光谱的相似度,选用物理意义更强的 SPA 对有机污染物的紫外-可见光光谱进行特征波段提取,并根据特征提取后光谱数据构建分类模型,实现对不同浓度的有机污染物分类,从而确定异常事件的污染物的类别。

## 1 方法研究

### 1.1 紫外光谱法有机污染物分类的总体架构

使用紫外光谱仪得到有机污染物的紫外光谱原始数据后,为了确定污染物类别,还需要以下三个步骤:

(1)光谱数据预处理,通过预处理方法消除紫外-可见光光谱数据中的噪声和基线漂移干扰。

(2)特征提取,由于部分有机物的紫外-可见光光谱的相似度较高,吸收峰所在波长范围部分重叠,因此需要对紫外-可见光光谱数据进行特征提取,增大不同污染物之间的差别。

(3)建模与预测,根据特征光谱数据构建分类模型,基于分类模型对未知样本进行分类。

基于紫外-可见光光谱的有机污染物分类方法的流程如图 1 所示。



图 1 基于紫外-可见光光谱的有机污染物分类的一般流程  
Fig. 1 General procedure for classification of organic contaminants based on the UV-Vis spectra

### 1.2 基于 SPA 的特征提取方法

SPA 是一种使矢量空间共线性最小化的前向变量选择算法,在有效信息获取和降低共线信息的研究中取得了较好的效果<sup>[9-10]</sup>。最初的应用场景是近红外光谱的定量模型中光谱变量的选取,通过投影方式选取线性关系最小的波长组合,从光谱信息中寻找含有最低冗余信息的变量组,使得变量之间的共线性达到最小,同时保留原始数据的绝大部分特征,被选取的特征波长物理意义明确,具有很强的解释能力,因此,可以有效的提高建模的速度以及模型的稳定性,该算法简要介绍如下:

设  $x_{k(0)}$  为选取的原始光谱矩阵的某个波长的吸光度向量,  $N$  为需要选取的特征波长的个数,初始的光谱矩阵为  $m$  列,每次迭代开始前选取的光谱数据第一列不同,算法需要对以下的步骤循环  $m$  次:

(1)迭代开始前,任选光谱矩阵的 1 列,用  $i$  表示,即把原始光谱矩阵的第  $i$  列  $x_i$  赋值给  $x_{k(0)}$ ;

(2)把原始光谱矩阵中未选入的列向量位置的集合记为  $s$ ,记为  $s = \{j, 1 \leq j \leq m; j \notin \{k(0), \dots, k(n-1)\}\}$ ;

(3)计算当前的所选变量  $x_{k(n-1)}$  对剩下原始光谱数据的

列向量  $x_i (i \in s)$  的投影:  $Px_i = x_i (x_i^T x_{k(n-1)}) x_{k(n-1)} (x_{k(n-1)}^T x_{k(n-1)})^{-1}$ ;

(4)提取投影最大值的波长变量序号,  $k(n) = \arg[\max(\|Px_i\|)]$ ;

(5)令  $x_i = Px_i, i \in s$ ;

(6)  $n = n + 1$ ,如果  $n < N$ ,回到步骤(2)循环计算。

最后,提取的特征波长变量为  $\{x_{k(n)}, n = 0, \dots, N-1\}$ 。对应于每一个  $k(0)$  和  $N$ ,每循环一次算法,对提取的特征波长处的光谱吸光度进行多元线性回归分析,计算验证集的预测标准偏差,其中最小预测标准偏差对应的  $k(0)$  和  $N$  就是最优的特征波长组合。

### 1.3 基于 SVM 的有机污染物分类算法

大多数有机物的紫外-可见光光谱具有专属特点,称为指纹特性<sup>[11]</sup>,可以根据有机物的紫外-可见光光谱的指纹特性确定类别。本文选择多分类 SVM 对不同浓度有机污染物的紫外-可见光光谱数据进行分类。

SVM 是基于统计学习理论的一种模式识别方法<sup>[12]</sup>。SVM 算法最初是针对二分类问题设计的,由于饮用水有机污染物分类是一个多分类问题,因此,需要对 SVM 进行适当变化,从而构造多分类器<sup>[13]</sup>。本文选用一对一方法构造多分类 SVM 模型,该方法与样本分布无关,分类器组合比较方便<sup>[14]</sup>。主要算法流程如下:

(1)对于不同的污染物样本,需要在任意两类样本之间构造一个 SVM,  $k$  类样本就需要设计  $k(k-1)/2$  个分类器;

(2)对一个新样本进行分类时,通过多个分类器投票决定分类结果。

## 2 实验验证与结果分析

### 2.1 设计与数据

为了验证基于紫外-可见光光谱的有机污染物分类的可行性,选取了四种典型的有机污染物,分别是苯酚、对苯二酚、间苯二酚和间苯二胺。这些物质在工业生产中使用较为广泛,被列入 2017 年公布的《世界卫生组织国际癌症研究机构致癌物清单》中,尤其值得关注。配置多浓度梯度的四种污染物水溶液,浓度依次为 30, 40, 50, 100 和 200  $\mu\text{g} \cdot \text{L}^{-1}$ 。训练集中每个浓度的紫外-可见光光谱数据为 10 个。测试集中苯酚 71 个,对苯二酚 58 个,间苯二酚 72 个,间苯二胺 79 个。考虑到水质波动的影响,训练集与测试数据的采集时间间隔一周。

根据朗伯-比尔定律,紫外-可见光光谱的吸光度会随着特征污染物浓度的增大而增大,如果该点的吸光度与特征污染物的浓度阵之间的关系为正相关,说明特征污染物在该波长会对光子产生吸收。将特征污染物的紫外-可见光光谱的每个波长对应的吸光度向量  $x$  与浓度向量  $y$  进行相关性分析,计算公式如下所示

$$R = \left[ \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \right] / \sqrt{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}$$

式中,  $\bar{x} = (\sum_{i=1}^N x_i) / N$ ,  $\bar{y} = (\sum_{i=1}^N y_i) / N$ , 其中,  $N$  为光谱的

数量。

四种特征污染物的浓度与各个波长的吸光度之间的关系如图 2 所示,其中正相关关系所在的波长范围与吸收峰所在位置基本一致。苯酚与间苯二酚、对苯二酚与间苯二胺的相关关系对比图如图 3 所示。

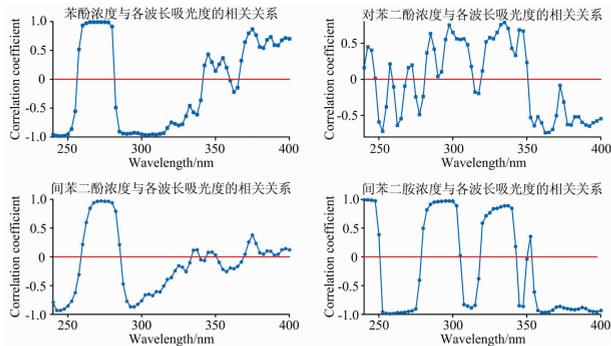


图 2 四种有机污染物浓度与各波长吸光度相关关系图

Fig. 2 Correlation between the concentration of four organic contaminants and the absorbance at each wavelength

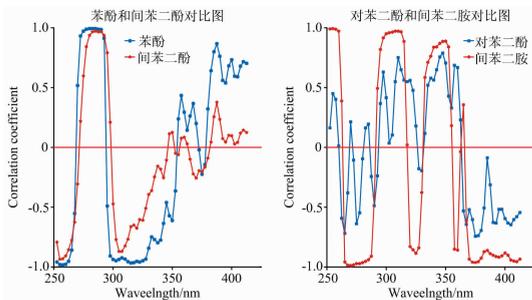


图 3 四种有机污染物的相关关系对比图

Fig. 3 Comparison of the correlation between four organic contaminants

从图 2 和图 3 可以看出,四种特征污染物之间存在部分谱峰重叠,其中苯酚和间苯二酚的谱峰重叠比较严重,原因是两者的分子结构相似,其生色团基本一致,只是助色团的数量和位置不一样,分类时容易相互干扰。对苯二酚和间苯二胺的助色团差别较大,存在部分谱峰重叠,分类决策容易受到浓度的干扰<sup>[15]</sup>。此外,四种特征污染物的紫外-可见光谱存在部分谱峰重叠,相互之间存在一定的干扰,需要选择合适的方法提取光谱数据的特征,增大光谱数据之间的差别。

## 2.2 光谱数据特征提取

使用 PCA 对四种有机污染物分别提取前三主成分,四种有机污染物的前三主成分如图 4 所示。从图中可以看出,有机污染物的主成分分布比较相似。

采用 SPA 方法对训练集数据进行特征波长选取时,由于 200 和 100  $\mu\text{g} \cdot \text{L}^{-1}$  的数据浓度较高,光谱数据的吸光度与污染物浓度存在一定的非线性,30  $\mu\text{g} \cdot \text{L}^{-1}$  数据中的干扰较大,因此在选取特征波长时,采用的浓度为 50 和 100  $\mu\text{g} \cdot \text{L}^{-1}$ 。对不同特征波长个数下的四种污染物紫外-可见光谱

吸光度进行多元线性回归分析,计算得到的 RMSE 分布曲线如图 5 所示,图中红色方框表示 SPA 算法选择的污染物紫外-可见光谱特征波段的数量。

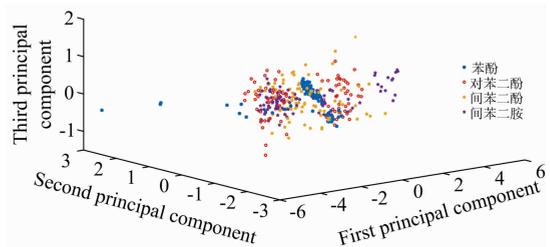


图 4 四种有机污染物的前三主成分的散点图

Fig. 4 Scatter plot of the top three principal components of four organic contaminants

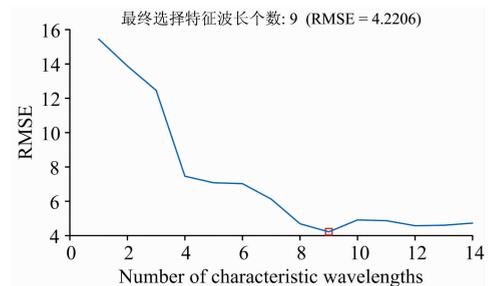


图 5 SPA 不同特征波长个数时 RMSE 分布曲线

Fig. 5 RMSE distribution curve of SPA with different characteristic wavelengths

最后选取的特征波长数为 9 个,根据重要性排序,波长依次为 277.5, 382.5, 295, 322.5, 252.5, 260, 337.5, 310 和 315 nm,这些波长基本都是四种有机污染物的吸收峰所在。四种有机污染物各个波长段的吸光度分布如图 6 所示。从图中可以看出,所选取的波长处的相似有机污染物的吸收光谱的重叠度较低。

## 2.3 种类判别与结果分析

根据特征光谱波段数据构建 KNN 模型和多元 SVM 模型,其中 PCA 方法选取前三主成分,SPA 选取 9 个特征波长,根据特征提取方法对训练集和测试集的数据进行特征提取,构建对应的训练集和测试集。为了对比研究,也使用基于原始光谱的 65 维的数据建立了相应的分类模型。表 1 显示三者对训练集数据以及测试集数据的分类准确度。

通过对比表 1 中不同特征提取方法的分类效果,可以得出以下结论:基于全光谱的分类模型对训练集的分类精度较高,但是对于测试集的分类精度相对较低,说明模型的预测能力较差,原因是全光谱中存在较多的干扰信息和冗余信息;基于 PCA 的特征光谱数据模型对训练集和测试集的分类精度相对较低,原因为 PCA 提取特征的过程中,根据协方差对数据进行降维,引入了干扰信息,相似污染物之间的干扰比较严重,因此分类的准确度不高;基于 SPA 筛选的特征波长模型对训练集和测试集的分类精度都较高,说明模型的稳定性和预测能力都较好。SPA 仅选取了全光谱的 13.8% 的变量,变量之间的共线性更小,有助于构建一个相

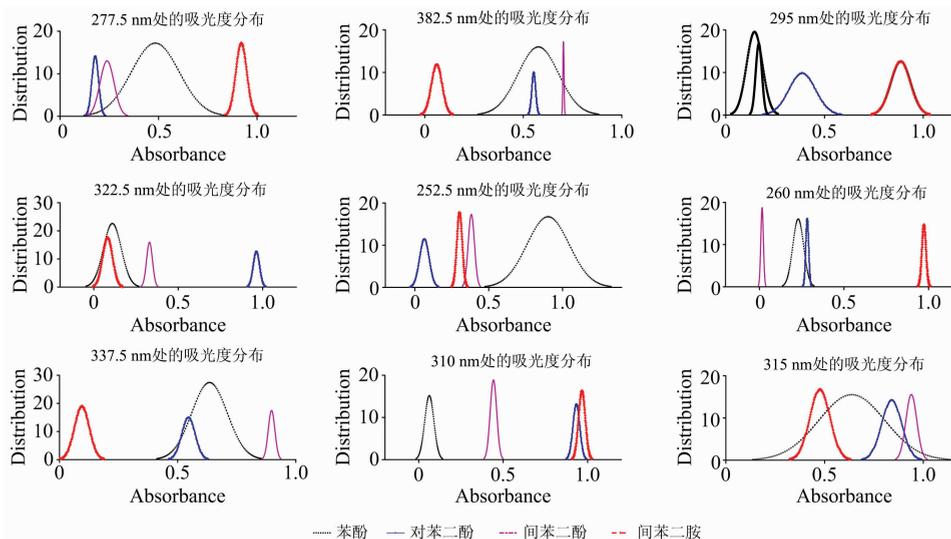


图 6 SPA 选取的特征波长处的有机污染物的吸光度分布

Fig. 6 Absorbance distribution of organic contaminants at characteristic wavelengths selected by SPA

表 1 有机污染物在不同特征提取方法下的分类效果

Table 1 Classification effect of organic contaminants under different feature extraction methods

建模方法	变量数	训练集/%	测试集/%
全光谱+KNN	65	94.51	85.36
全光谱+多分类 SVM	65	100	87.14
PCA+KNN	3	89.20	60.70
PCA+多分类 SVM	3	97.44	61.07
SPA+KNN	9	91.48	88.21
SPA+多分类 SVM	9	99.15	95.36

对简单、稳定的分类模型，并且 SPA 在两种分类方法下的分类准确率都很高，表明该方法的适用性较强。

整体分类准确率虽然可以反映多分类 SVM 模型整体的分类效果，但是无法反映有机污染物之间的相互干扰，因此，需要通过混淆矩阵分析多分类 SVM 模型对有机污染物的分类结果。基于全光谱的四种有机污染物分类效果如图 7 所示；基于 PCA 特征提取的分类效果如图 8 所示；基于 SPA

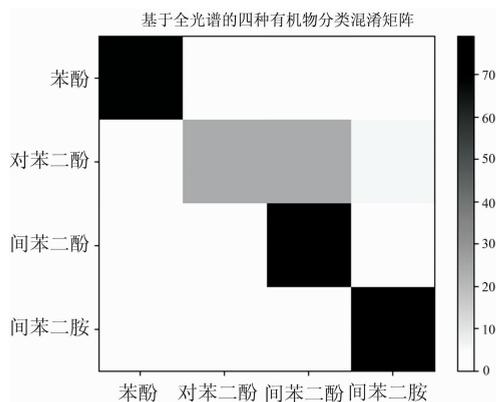


图 7 基于全光谱的有机污染物分类效果

Fig. 7 Classification effect of organic contaminants based on full spectrum

特征提取的分类结果如图 9 所示。混淆矩阵图中，颜色越深代表对应的两种污染物相似度越高，所有的正确分类都集中

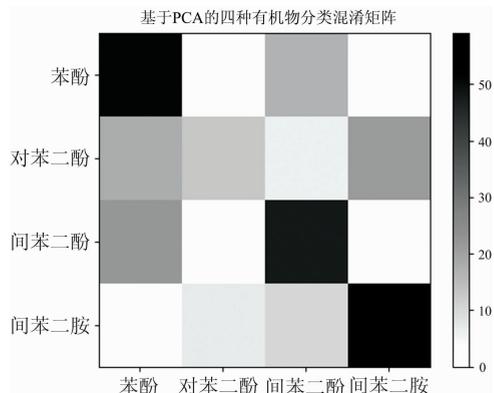


图 8 基于 PCA 特征提取的有机污染物分类效果

Fig. 8 Classification effect of organic contaminants based on PCA

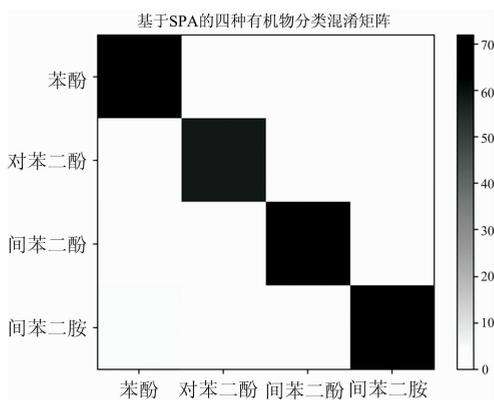


图 9 基于 SPA 特征提取的有机污染物分类效果

Fig. 9 Classification effect of organic contaminants based on SPA

在矩阵的对角线上。如果一张混淆矩阵的图呈现出对角线相似度很高的特征,则认为该模型分类效果较好<sup>[16]</sup>。

通过对比可以看出,基于 SPA 的四种污染物分类混淆矩阵对角线相似度最高,说明该模型下四种污染物之间的相互干扰最小,分类性能比较稳定;而基于全光谱和基于 PCA 的分类模型中,混淆矩阵相似度分布相对更加平均,表明污染物之间相互干扰严重,性能不够稳定。

为了分析浓度的干扰,需要分析不同浓度下有机污染物的分类准确率,从而评估其分类性能。基于全光谱的不同浓度有机污染物分类准确率如表 2 所示;基于 PCA 特征提取的不同浓度有机污染物分类准确率如表 3 所示;基于 SPA 特征提取的不同浓度有机污染物分类准确率如表 4 所示。

表 2 基于全光谱的不同浓度有机污染物分类准确率 (%)

Table 2 Classification accuracy of organic contaminants at different concentrations based on full spectrum (%)

类别	浓度/( $\mu\text{g} \cdot \text{L}^{-1}$ )					总准确率
	30	40	50	100	200	
苯酚	92.86	100	92.86	100	100	97.18
对苯二酚	0	0	100	83.33	16.67	41.38
间苯二酚	100	100	100	100	100	100
间苯二胺	100	100	100	100	100	100

表 3 基于 PCA 特征提取的不同浓度有机污染物分类准确率 (%)

Table 3 Classification accuracy of organic contaminants at different concentrations based on PCA (%)

类别	浓度/( $\mu\text{g} \cdot \text{L}^{-1}$ )					总准确率
	30	40	50	100	200	
苯酚	85.71	73.33	57.14	64.29	78.57	71.83
对苯二酚	36.36	18.18	8.33	16.67	33.33	22.41
间苯二酚	40.00	64.29	73.33	71.43	85.71	66.67
间苯二胺	86.67	100	100	0	93.75	74.68

表 4 基于 SPA 特征提取的不同浓度有机污染物分类准确率 (%)

Table 4 Classification accuracy of organic contaminants at different concentrations based on SPA (%)

类别	浓度/( $\mu\text{g} \cdot \text{L}^{-1}$ )					总准确率
	30	40	50	100	200	
苯酚	100	93.33	100	98.71	100	95.77
对苯二酚	100	100	100	100	100	100
间苯二酚	100	100	93.33	85.71	100	95.83
间苯二胺	100	100	100	58.82	100	91.44

上述不同浓度分类准确率的结果表明,全光谱+多分类 SVM 模型对低浓度和高浓度的对苯二酚分类效果较差,对其他三种有机污染物的分类效果较好;PCA+多分类 SVM 模型对各浓度梯度的有机污染物分类精度较差;SPA+多分类 SVM 对各浓度梯度的有机污染物分类效果较好,对四种有机污染物的分类精度可以达到 90% 以上,稳定性较好。

### 3 结 论

针对有机污染物紫外-可见光光谱存在谱峰重叠导致分类效果不佳的问题,开展了有机污染物特征提取和判别方法的研究。

通过实验获取苯酚、对苯二酚、间苯二酚、间苯二胺这四种谱峰相似有机污染物在不同浓度下的紫外-可见光光谱数据,并比较了不同特征提取和分类方法对不同浓度有机污染物的分类效果。实验结果表明,与全光谱和 PCA 特征提取方法相比,SPA 能有效地提取有机污染物紫外吸收光谱的特征波长组合并表征各类有机污染物的特征,不仅能够消除光谱数据的多重共线性干扰以及相似污染物之间的干扰,同时增强了有机污染物之间的差别,具有很好的物理意义和解释性,结合多分类 SVM 可实现对不同浓度有机污染物的有效分类,为实现饮用水中谱峰重叠污染物类型的判别提供有效方法。

### References

- [1] Miao X, Tang Y, Wong C W Y, et al. Environmental Pollution, 2015, 196: 473.
- [2] Assaad A, Pontvianne S, Marie-Noëlle Pons. Environmental Monitoring and Assessment, 2017, 189(5): 229.
- [3] Zhang H, Tian J, Huang J, et al. RSC Adv., 2016, 6: 110356.
- [4] Huang P, Wang K, Hou D, et al. Applied Optics, 2017, 56(22): 6317.
- [5] Yu Q, Yin H, Wang K, et al. Water, 2018, 10(11): 1566.
- [6] ZHAO Ming-fu, TANG Ping, TANG Bin, et al(赵明富,唐平,汤斌,等). Journal of Atmospheric and Environmental Optics(大气与环境光学学报), 2018, 13(6): 436.
- [7] Hou D, Zhang J, Yang Z, et al. Optics Express, 2015, 23(13): 17487.
- [8] GUO Bing-bing, HOU Di-bo, JIN Yu, et al(郭冰冰,侯迪波,金玉,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(5): 1460.
- [9] Zhu H, Chu B, Zhang C, et al. Scientific Reports, 2017, 7(1): 4125.
- [10] Gomes A A, Schenone A V, Goicoechea H C, et al. Analytical & Bioanalytical Chemistry, 2015, 407(19): 5649.
- [11] Terrile, Amélia E, Marcheafave, et al. Journal of the Brazilian Chemical Society, 2016, 27(7): 1254.
- [12] Vapnik V. The Nature of Statistical Learning Theory. Springer, 1995.
- [13] Huang P, Jin Y, Hou D, et al. Sensors, 2017, 17(3): 581.

- [14] Daengduang S, Vateekul P. IEEE 2017 9th International Conference on Knowledge and Smart Technology (KST)-Chonburi, Thailand (2017. 2. 1—2017. 2. 4). 2017 9th International Conference on Knowledge and Smart Technology (KST)-Applying One-Versus-One SVMs to Classify Multi-Label Data with Large Labels Using Spark, 2017. 72.
- [15] CHU Xiao-li(褚小立). Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011.
- [16] Ting K M. Confusion Matrix. Encyclopedia of Machine Learning and Data Mining. Springer US, 2017.

## Classification of Organic Contaminants in Water Distribution Systems Developed by SPA and Multi-Classification SVM Using UV-Vis Spectroscopy

HUANG Ping-jie, LI Yu-han, YU Qiao-jun, WANG Ke, YIN Hang, HOU Di-bo\*, ZHANG Guang-xin  
State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

**Abstract** Quickly and effectively identifying the water contaminants is vital for reducing the impact of sudden drinking water pollution incidents. PCA is mostly used to extract the feature of different contaminants in drinking water with UV-Vis spectra. However, for the organic contaminants with high similarity in UV-Vis spectra, the identification result is ineffective when only extracting the feature of the largest variance direction from the data-driven point of view. This paper studies the classification of organic contaminants in water distribution systems developed by SPA and multi-classification SVM using UV-Vis spectroscopy. Firstly, the original spectral data of phenol, hydroquinone, resorcinol and m-phenylenediamine are measured by UV spectrometer and pretreated. The correlation between wavelength and concentration of four contaminants was compared. The peaks between phenol and resorcinol, hydroquinone and m-phenylenediamine are overlapped seriously, the classification results can interfere easily. In feature extraction, the SPA is introduced to select the organic contaminants' characteristic wavelengths of UV-Vis spectra. Then, multiple linear regression analysis is carried out to choose the optimal parameter combination, which corresponds to the minimum prediction standard deviation. Based on this, the multi-classification support vector machine is used to form an identification model for drinking water organic contaminants. Finally, the classification results of spectral data based on full spectrum, PCA and SPA under different classification methods and different concentrations are compared and analyzed, and the applicability and stability of SPA are further explained. Experimental results demonstrate that SPA-based feature extraction method eliminates the interference of multi-collinearity and amplifies the difference among the UV-Vis spectra of different organic contaminants, thereby improving the accuracy of the classification model. This method has certain reference value for solving the problem of identifying the types of pollutants with overlapped peaks in the drinking water.

**Keywords** UV-Vis spectroscopy; Identification of organic contaminants; Successive projections algorithm; Multi-classification support vector machine

(Received Apr. 18, 2019; accepted Aug. 9, 2019)

\* Corresponding author