第 42 卷 第 12 期/2022 年 6 月/光学学报 研究论文

光学学报

基于三维荧光光谱和 GBDT-LR 的褐潮藻辨识

陈颖[1*]，段玮靓[1]，杨英[1]，刘喆[1]，张永彬[1]，刘俊飞[1]，李少华[2]

[1]燕山大学电气工程学院河北省测试计量技术及仪器重点实验室，河北 秦皇岛 066004；

[2]河北先河环保科技股份有限公司，河北 石家庄 050035

摘要 近年来频繁发生的褐潮污染给沿海地区经济带来巨大损失。准确、高效地识别褐潮藻对预防海洋环境污染意义重大。采用三维荧光光谱、梯度提升决策树(GBDT)和逻辑回归(LR)相结合的方法，实现了对褐潮藻的准确辨识。为解决 LR 模型对非线性数据的特征组合能力较弱的问题，引入 GBDT 算法，充分利用集成学习算法在处理非线性数据上的优势。将 GBDT 的预测结果作为新特征代替原来的特征输入 LR 模型，建立了一种将 GBDT 与 LR 相融合的褐潮藻辨识模型(GBDT-LR)。针对复杂海洋环境中其他门类藻的干扰，实验引入小球藻、细长聚球藻等 5 种不同门类的海藻作为对比，并对处于不同生长周期的褐潮藻辨识情况进行分析。相同条件下通过将所提模型与 LR、支持向量机(SVM)和反向传播(BP)神经网络等模型进行对比。结果表明，GBDT-LR 在分类准确率、召回率和 F1 分数等评价指标上均优于其他模型，处于指数生长期的藻类荧光光谱最为稳定，这一时期的褐潮藻辨识结果最好。

关键词 光谱学；三维荧光光谱；褐潮污染；特征提取；逻辑回归；梯度提升决策树

中图分类号 X834 **文献标志码** A **DOI**：10.3788/AOS202242.1230001

Identification of Brown Tide Algae Based on Three-Dimensional Fluorescence Spectra and GBDT-LR

Chen Ying[1*], Duan Weiliang[1], Yang Ying[1], Liu Zhe[1], Zhang Yongbin[1], Liu Junfei[1], Li Shaohua[2]

[1] *Key Laboratory of Measurement Technology and Instrumentation of Hebei Province, School of Electrical Engineering, Yanshan University, Qinhuangdao* 066004, *Hebei, China*;

[2] *Hebei Sailhero Environmental Protection High-Tech Co., Ltd., Shijiazhuang* 050035, *Hebei, China*

Abstract The frequent occurrence of brown tide pollution in recent years has brought huge losses to the economy of coastal areas. Therefore, the accurate and efficient identification of brown tide algae is of great significance to the prevention of marine environmental pollution. In this paper, a combination method of three-dimensional fluorescence spectroscopy, gradient boosting decision tree (GBDT), and logistic regression (LR) is used to achieve accurate identification of brown tide algae. In order to solve the problem of weak feature combination ability of LR model for nonlinear data, the GBDT algorithm is introduced to make full use of the advantages of the integrated learning algorithm in processing nonlinear data. The prediction result of GBDT model is used as a new feature instead of the original feature which is input into the LR model, and a brown tide algae recognition model (GBDT-LR) that combines GBDT and LR is established. In response to the interference of other types of algae in the complex marine environment, five different types of algae such as *Chlorella* and *Synechococcus elongatus* are

收稿日期：2021-11-29；**修回日期**：2022-01-12；**录用日期**：2022-01-27

基金项目：国家重点研发计划项目(2016YFC1400601-3)、河北省重点研发计划项目(19273901D，20373301D)、河北省自然科学基金(F2020203066)、中国博士后基金项目(2018M630279)、河北省博士后择优资助项目(D2018003028)、河北省高等学校科学技术研究项目(ZD2018243)

通信作者：*chenying@ysu.edu.cn

introduced for comparison in the experiment, and analyzed the identification of brown tide algae in different growth cycles are analyzed. The proposed model is compared with LR, support vector machine (SVM) and back propagation (BP) neural network under the same conditions. The results show that the GBDT-LR model is superior to the other models in terms of classification accuracy, recall rate, and F1-score. The fluorescence spectrum of algae in the exponential growth period is the most stable, and the identification result of the brown tide algae in this period is the best.

Key words spectroscopy; three-dimensional fluorescence spectroscopy; brown tide pollution; feature extraction; logistic regression; gradient boosting decision tree

1 引 言

水体富营养化和藻华污染是全球性的水环境问题，给自然生态和人类生产生活带来了极大的危害[1-2]。中国渤海海域近年来多次发生大规模褐潮污染，给沿海当地经济带来了巨大损失[3-4]。在褐潮爆发期间，其密度最高可达 10^9 cell/L，严重影响鱼、虾等海洋生物的栖息[5]。在藻华污染的早期预防上，及时检测相关海域浮游藻门类信息和浓度信息并进行实时预警非常重要[6]。

目前针对浮游藻的检测方法主要有显微镜计数法[7]、高效液相色谱法[8]、图像识别技术[9]和分子探针方法[10]等，但这些方法操作步骤繁琐且十分考验实验人员的专业性。基于以上问题，光谱学的发展为浮游藻检测提供了更便利的方法。其中，三维(3D)荧光光谱法不仅可以获得激发波长和发射波长，还能够获得变化时的光谱强度信息，获得的荧光信息远多于普通荧光光谱，所以它的检测精度更高[11]。近年来，利用三维荧光光谱数据结合相关分类算法进行浮游藻检测也是常用方法[12-13]。

传统的物质检测方法有基于特征提取的算法[14-15]和基于机器学习的算法[16]。近年来，集成学习常被用于各类物质的定性判别[17]。有实验结果证明，集成学习方法相对单分类器有更高的分类精度。葛文杰等[18]利用随机森林算法与多元信息相融合的方法，提高了疲劳驾驶检测的精度，与支持向量机(SVM)等算法相比，集成学习算法训练速度更快，准确率更高。周杰英等[19]通过将梯度提升决策树(GBDT)与随机森林相结合，提高了网络入侵检测的准确率，证明了集成学习算法在不平衡数据分类问题上的优势。集成学习在非线性数据处理上表现较好，但是当数据维度过高时模型计算复杂度会大大提高，此时模型不满足海洋环境监测对实时性的要求。

逻辑回归(LR)是一种线性分类器，模型通过 Sigmoid 函数将分类结果转化为概率输出[20]。常钰迪[21]为了提高稀疏矩阵用于分类模型的准确率，提出基于稀疏 LR 的链接神经网络模型，其在手写字和海洋哺乳动物分类数据上的准确率相比神经网络模型有了一定的提升。然而，LR 作为线性分类器，其特征组合能力有限，在非线性数据上表现不佳。

综上所述，本文将 GBDT 与 LR 模型相结合，建立了一种基于三维荧光光谱与 GBDT-LR 的褐潮藻分类辨识模型。首先，将复杂海域褐潮藻分类鉴别转化为二分类问题，然后将集成学习中多个分类器的预测结果转换为稀疏矩阵形式，将其作为先验知识代替原来的特征，并输入 LR 模型中进行训练。不同模型相结合可充分发挥各自优势，从而提高模型的性能，最终实现对褐潮藻的准确鉴别。

2 实验部分

2.1 藻种培养

本实验所用褐潮藻种为抑食金球藻。近海海域常见的藻种，如小球藻、细长聚球藻、圆海链藻、东海原甲藻和杜氏盐藻，因为生长条件与抑食金球藻相似，且相互之间存在光谱重叠的现象，所以被选择作为对比藻种，以此模拟实际海洋环境。所有藻种放置于温度培养箱中进行扩大培养，设置培养箱光照为 1500 lx，温度为 25 ℃，光暗循环比为 12 h∶12 h。

2.2 光谱获取与分析

实验仪器采用 FS920 荧光光谱仪，其检测波长范围为 200～900 nm，设置其激发波长为 400～650 nm，每隔 5 nm 扫描一次。发射波长范围为 630～730 nm，每隔 5 nm 扫描一次。为确保测量精度，比色皿在两组样本的测量间隔期间要使用蒸馏水进行清洁。由于测量环境和仪器在数据采集过程中会受到外界环境噪声、人为因素等众多复杂因素的影响，并且荧光光谱采集过程中会产生瑞利散射，故需要对光谱数据进行预处理。本实验采用 Savitzky-Golay 算法对光谱进行平滑去噪，利用 Delaunay 三角内插值算法消除瑞利散射的影响。

经过去噪平滑和去除散射等处理后，各藻种的

三维荧光光谱图和等高线图如图 2 所示。表 1 列出了各个藻类的主要光合色素组成。藻类的光合色素是藻类荧光发光的物质基础，也是藻类三维荧光光谱体现特异性的根本原因。

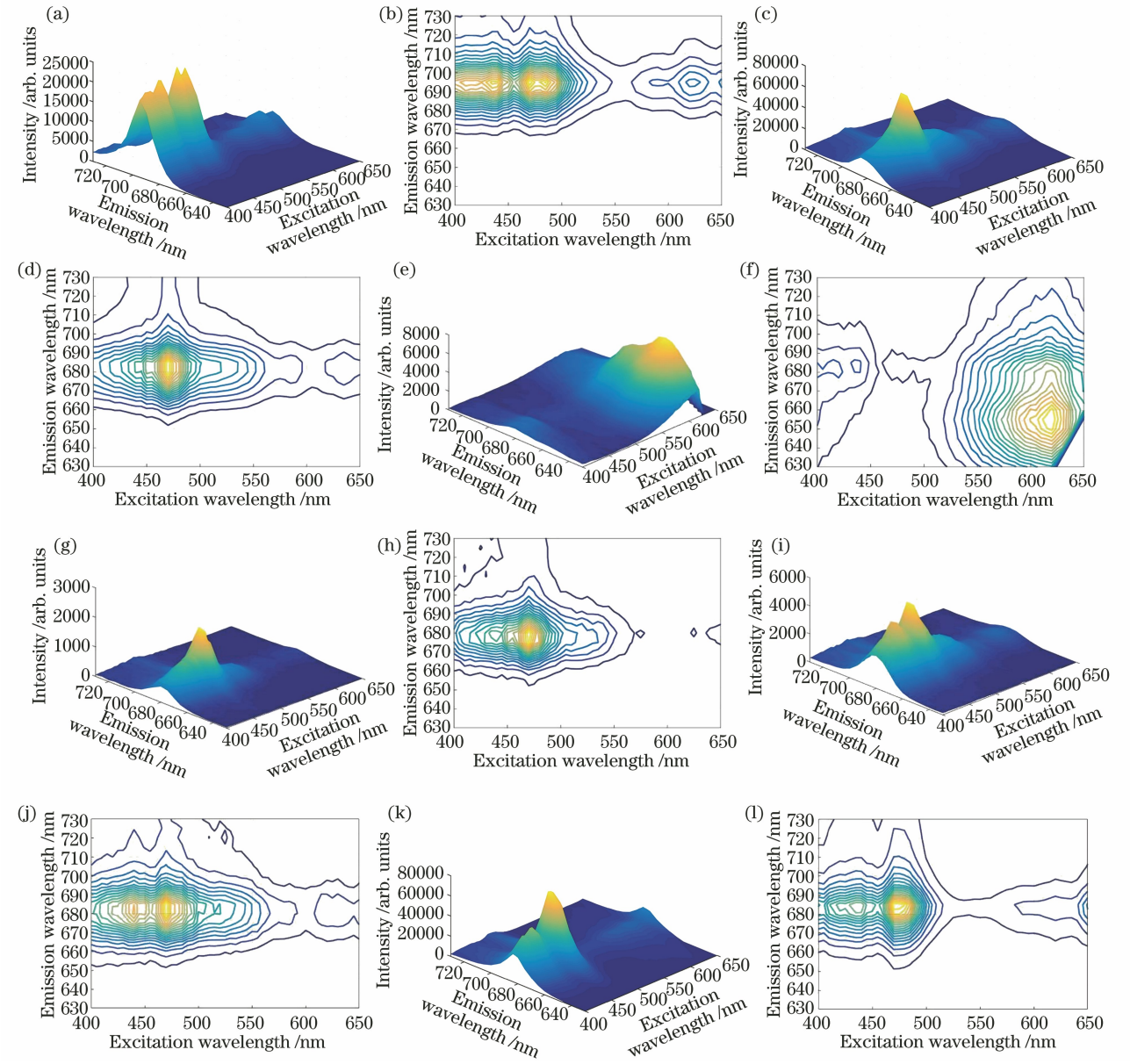

图 1 海藻三维荧光光谱图和等高线图。(a)小球藻的三维荧光光谱图；(b)小球藻的等高线图；(c)抑食金球藻的三维荧光光谱图；(d)抑食金球藻的等高线图；(e)细长聚球藻的三维荧光光谱图；(f)细长聚球藻的等高线图；(g)东海原甲藻的三维荧光光谱图；(h)东海原甲藻的等高线图；(i)圆海链藻的三维荧光光谱图；(j)圆海链藻的等高线图；(k)杜氏盐藻的三维荧光光谱图；(l)杜氏盐藻的等高线图

Fig. 1 3D fluorescence spectra and contour maps of algae. (a) 3D fluorescence spectrum of *Chlorella*; (b) contour map of *Chlorella*; (c) 3D fluorescence spectrum of *Aureococcus anophagefferens*; (d) contour map of *Aureococcus anophagefferens*; (e) 3D fluorescence spectrum of *Synechococcus elongatus*; (f) contour map of *Synechococcus elongatus*; (g) 3D fluorescence spectrum of *Prorocentrum donghaiense*; (h) contour map of *Prorocentrum donghaiense*; (i) 3D fluorescence spectrum of *Thalassiosira rotula*; (j) contour map of *Thalassiosira rotula*; (k) 3D fluorescence spectrum of *Dunaliella salina*; (l) contour map of *Dunaliella salina*

结合图 1 和表 1 可以看出，不同藻类之间可能包含相同的光合色素，且不同色素荧光吸收峰的位置可能相近，如叶绿素 b 的吸收峰在 453 nm 处，岩藻黄素的吸收峰在 450～470 nm 范围内，这也导致荧光峰出现重叠现象[22]。当利用三维荧光光谱结合化学计量学分析算法进行藻类区分时，将光谱数

据直接用于模型识别的运算复杂度较高，且光谱数据包含很多冗余信息，这给分类识别算法带来一定的困难。为降低模型运算复杂度、有效提取重叠峰位置的主要光谱信息和提高模型辨识准确率，需要对光谱数据进行特征提取。

表 1 主要色素与藻类的对应关系

Table 1 Corresponding relationship between main pigments and algae

Algae	Category	Main pigment
Aureococcus anophagefferens	Ochromonadaceae	Chlorophyll b, fucoxanthin
Chlorella	Chlorophyta	Chlorophyll a, chlorophyll b
Synechococcus elongatus	Cyanophyta	Phyeoeyanin
Prorocentrum donghaiense	Pyrrophyta	Chlorophyll b
Dunaliella salina	Chlorophyta	Chlorophyll a, chlorophyll b
Thalassiosira rotula	Bacillariophyta	Chlorophyll a, chlorophyll b

3 原理分析

3.1 基于 LR 的褐潮辨识模型

LR 是一种广义线性模型，主要用于解决分类问题[23]。在褐潮藻辨识过程中，LR 有两个输出：1 代表褐潮藻样本；0 代表其余 5 种干扰藻类样本。LR 的表达式为

$$\hat{y}=w_0+w_1x_1+\cdots+w_nx_n, \tag{1}$$

式中：x_i 为样本特征光谱值，$i=1,2,\cdots,n$，n 为光谱特征维度；w_i 为权重；$\hat{y}$ 为模型预测值。LR 通过 Sigmoid 函数将输出结果映射到[0,1]范围内，并以概率的形式输出。Sigmoid 函数的表达式为

$$g(z)=\frac{1}{1+e^{-\hat{y}}}。 \tag{2}$$

Sigmoid 函数的输出不再是分类结果，而是一个样本被预测为正例的概率 $p(x_i)$，预测为负例的概率为 $1-p(x_i)$。对于模型参数中的权重 w_i，可以利用最小化负对数似然函数进行求解，似然函数可以表示为

$$L(w_i)=\prod[p(x_i)]^{y_i}[1-p(x_i)]^{1-y_i}, \tag{3}$$

式中：y_i 为第 i 个样本的真实值。

对式(3)两边同时取负对数，将似然函数表示成对数似然函数，通过随机梯度下降法求得权重 w_i，将其代入函数表达式中得到对应的预测概率 $p(x_i)$。

以抑食金球藻为例，在激发波长处于 470 nm 的前提下，分别绘制每组样本的荧光发射光谱，如图 2 所示。不同浓度样本对应的荧光强度差异较大，将其直接用于 LR 模型训练容易造成欠拟合。因此，在模型训练前，首先应对光谱信息进行特征提取，再输入模型进行训练。

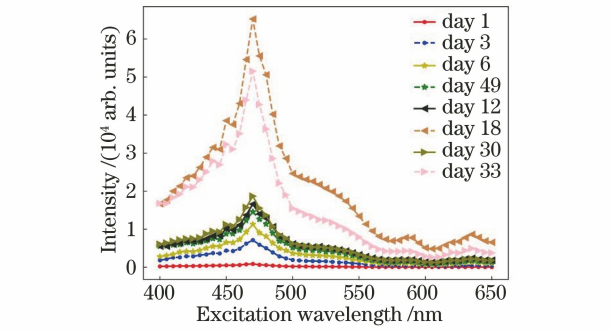

图 2 不同生长周期抑食金球藻的荧光激发光谱

Fig. 2 Fluorescence excitation spectra of *Aureococcus anophagefferens* at different growth periods

GBDT 中每一棵树的形成过程都会首先考虑区分度高的特征，然后再考虑区分度较低的特征，以挖掘有区分度的特征。因此，在利用 LR 进行褐潮藻辨识之前，先利用 GBDT 进行原始光谱数据的特征提取。

3.2 基于 GBDT 的光谱特征提取

GBDT 通过迭代多棵树进行共同决策，将所用树的结果进行求和得到最终结果，这是一种将弱学习算法提升为强学习算法的统计学习算法[24]。GBDT 模型结构示意图如图 3 所示，其中 f_1 为第一棵树，f_n 为第 n 棵树。

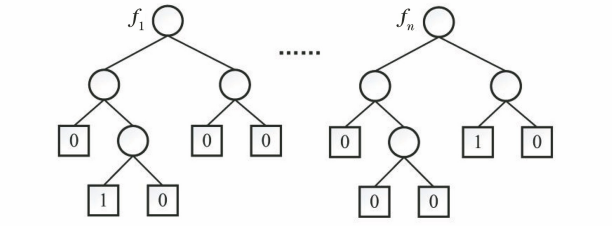

图 3 GBDT 模型结构示意图

Fig. 3 Structural diagram of GBDT model

GBDT 中每棵树学习的是前一棵树结果的残差，将残差与预测值相加后得到真实值的估计，最终

将每个弱分类器的分类结果进行加权求和，得到最终的藻类分类结果。新决策树会向前一棵决策树残差降低的方向形成。具体算法步骤如下：

1）对于给定的藻类光谱数据 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$（$N$ 为样本数量）和损失函数 $L[y_{i'},f(x_{i'})]$，初始化第一棵树的公式为

$$f_0(x_{i'})=\underset{c}{\operatorname{argmin}}\sum_{i'=1}^{N}L(y_{i'},c);\tag{4}$$

2）对于 M 棵决策树，重复以下运算。当 $i'=1,2,\cdots,N$ 时，计算损失函数的负梯度在当前模型的值，并将其作为残差估计，残差计算公式为

$$\begin{cases}r_{m,i'}=-\dfrac{\partial L[y,f_m(x_{i'})]}{\partial f_m(x_{i'})}\\ f(x_{i'})=f_{m-1}(x_{i'})\end{cases},\tag{5}$$

式中：m 表示第 m 棵树。以 $r_{m,i'}$ 作为新的数据集，拟合一颗新的决策树，得到第 m 棵树的叶节点区域 $R_{m,j}$（$j=1,2,\cdots,J$，J 为叶节点数）。求 $R_{m,j}$ 区域内使得损失函数达到最小的值 $c_{m,j}$。对于 $j=1,2,\cdots,J$，计算每一个节点区域的输出值

$$c_{m,j}=\underset{c}{\operatorname{argmin}}\sum_{x_{i'}\in R_{m,j}}L[y_{i'},f_{m-1}(x_{i'})+c],\tag{6}$$

决策树的更新策略可以表示为

$$f_m(x_{i'})=f_{m-1}(x_{i'})+v\sum_{j=1}^{J}c_{m,j}I,x_{i'}\in R_{m,j},\tag{7}$$

式中：v 是正则化项，用来防止模型出现过拟合，v 的取值范围为 $0<v<1$。v 越小表示需要更多的树来迭代，v 越大表示需要的树的数量越少。当 $x_{i'}\in R_{m,j}$ 时，$I=1$。当 $x_{i'}\notin R_{m,j}$ 时，$I=0$；

3）得到最终决策树，计算公式为

$$\hat{f}(x_{i'})=f_M(x_{i'})=\sum_{m=1}^{M}\sum_{j=1}^{J}c_{m,j}I,x_{i'}\in R_{m,j}。\tag{8}$$

在利用 GBDT 进行光谱特征提取前，首先通过构造学习曲线确定 GBDT 的最优参数。将原始藻类光谱数据通过 GBDT 进行特征提取，构造新特征，再将新特征通过 LR 模型进行训练，得到模型分类准确率与决策树个数 n_{est} 的关系，如图 4(a)所示。可以看出，当决策树个数为 36 时，模型的分类准确率最高。固定决策树的个数为 36，通过网格搜索确定最大叶节点个数，模型的分类准确率与最大叶节点个数 n_{max} 的关系如图 4(b)所示。当最大叶节点为 6 时，预测准确率达到最高。综上所述，设置 GBDT 模型的决策树个数为 36，最大叶节点个数为 6。

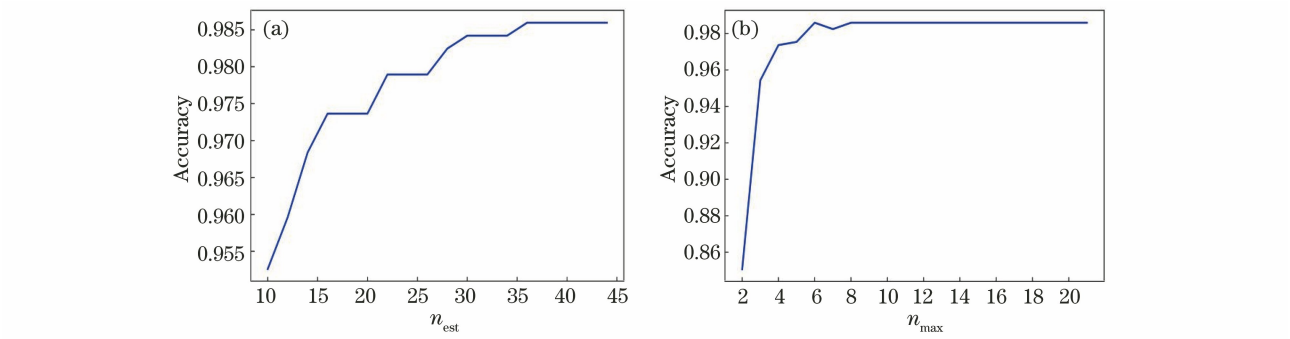

图 4　模型准确率与决策树参数的关系曲线图。(a)预测准确率与决策树个数的关系图；(b)预测准确率与最大叶节点个数的关系图

Fig. 4　Relationship between model accuracy and decision tree parameters. (a) Relationship between prediction accuracy and number of decision trees; (b) relationship between prediction accuracy and maximum number of leaf nodes

3.3　GBDT-LR 融合模型

GBDT-LR 模型的训练流程如图 5 所示，具体步骤如下：

1）构造学习曲线，确定 GBDT 最佳参数，由原始藻类荧光数据训练 GBDT 模型，以最小化残差为目标不断拟合新的决策树，最终组成一个强分类器；

2）GBDT 模型不以分类结果作为输出，而是以模型中每棵树预测值所在叶节点位置为特征，提取样本预测后在所有树中叶节点的位置信息，形成新的数据集；

3）对叶节点位置信息进行 One-hot 编码，将样本输出所属叶节点位置记为 1，其他叶节点位置记为 0，得到每个样本的位置信息向量。所有样本输出结果组成含有叶节点位置信息的稀疏矩阵。假设 GBDT 模型共生成 N' 棵树，每棵树有 m' 个叶节点，对于原始数据来说，每一个样本都会被转换为 $N'\times m'$ 维稀疏向量，其中有 N' 个元素的值为 1，其余元素的值为 0；

4）将该稀疏矩阵作为训练数据输入到 LR 模型中进行训练。

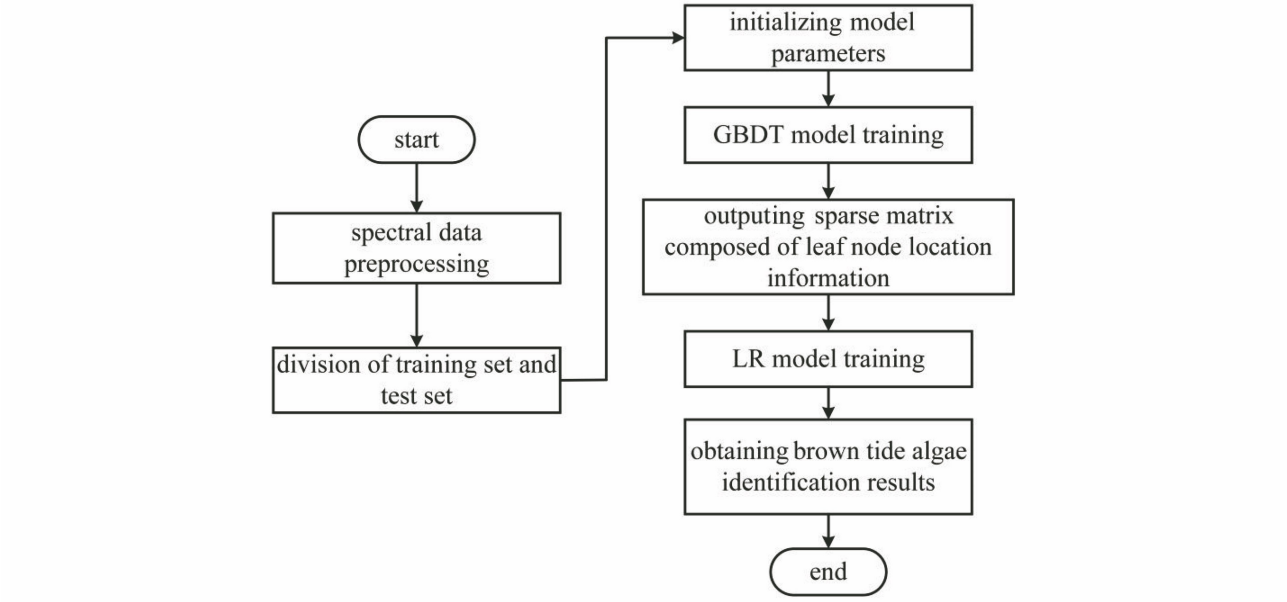

图 5 GBDT-LR 模型流程图

Fig. 5 Flowchart of GBDT-LR model

4 结果与讨论

4.1 辨识模型评价指标

本次实验主要采用以下几种指标评价模型好坏：

1）准确率(A)，表示分类正确的样本占总体样本的比例，表达式为

$$A=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{FP}+N_{FN}+N_{TN}},\tag{9}$$

式中：N_{TP} 表示真阳性；N_{TN} 表示真阴性；N_{FP} 表示假阳性；N_{FN} 表示假阴性；

2）召回率(R)，表示分类正确的正样本占总的正样本的比例，表达式为

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}};\tag{10}$$

3）F1 分数(F)，表示模型精确度和召回率的调和平均值，表达式为

$$F=2\frac{PR}{P+R},\tag{11}$$

式中：P 为精确度，表示预测为正类的样本中真正为正类的样本所占的比例；

4）接受者操作特征(ROC)曲线，输出概率分布的二分类器分类能力的一种图形化展示，概率大于阈值则预测为正，否则为负；

5）ROC 曲线下的面积(A_{UC})，A_{UC} 越大，说明分类器的效果更好。

4.2 辨识结果分析

实验数据来自不同生长周期的 6 种藻类，其中抑食金球藻为褐潮成因藻，其余均为近海海域的常见藻种。分别取各个样本荧光光谱数据作为 GBDT-LR 模型的输入，样本标签值作为输出，将抑食金球藻标签值设为 1，其余藻种标签值设为 0。

测试集选择不同生长周期的藻类样本共 57 组，其中生长天数在 1～7 d 内的有 10 组，生长天数在 8～14 d 的有 12 组，生长天数在 15～21 d 的有 18 组，生长天数在 22～33 d 的有 17 组。使用的性能指标为准确率、召回率、F1 分数和 A_{UC}。为进一步评价 GBDT-LR 模型性能，GBDT、SVM、LR 和 BP 模型作为对比也被用于褐潮藻辨识。各模型 ROC 曲线变化情况如图 6 所示，各模型应用在不同生长周期藻类上的分类结果如图 7 所示，各模型的评价指标如表 2 所示。

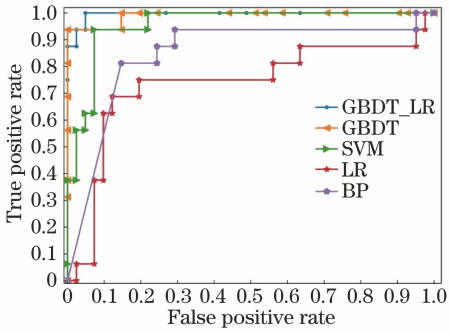

图 6 不同分类模型的 ROC 曲线对比

Fig. 6 Comparison of ROC curves of different classification models

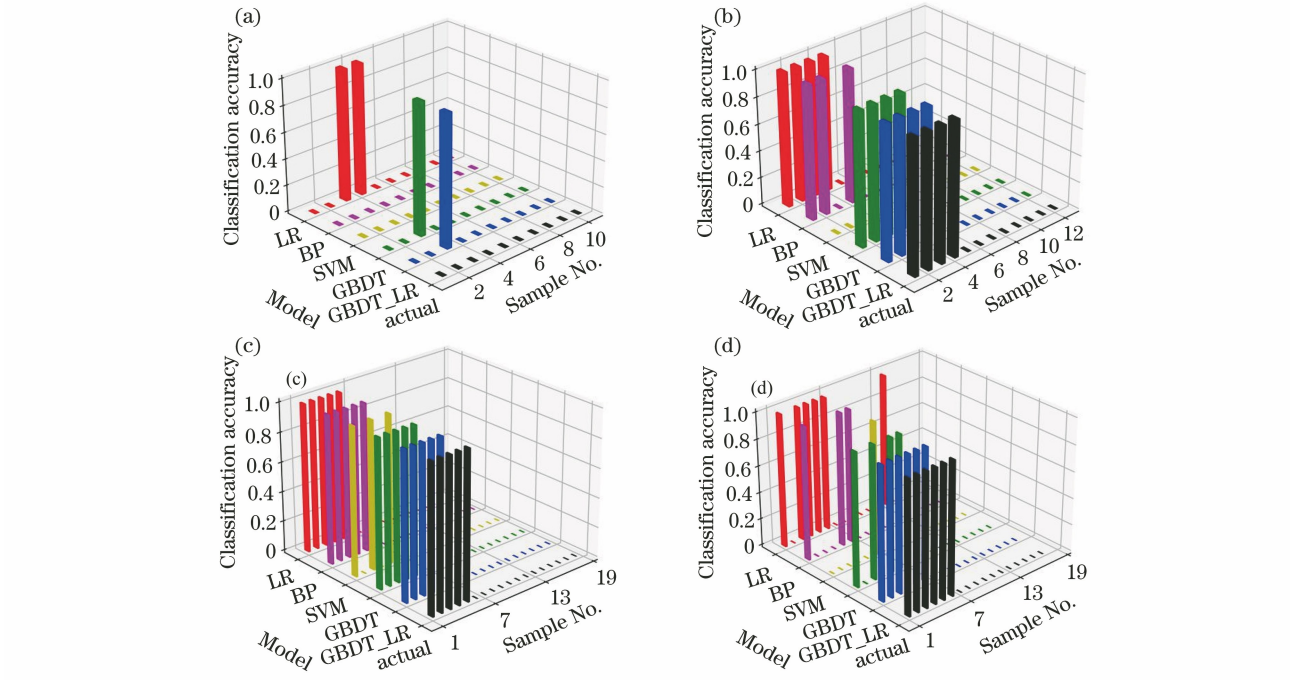

图 7　不同生长周期下各模型分类结果比较。(a)第一周藻类样本;(b)第二周藻类样本;(c)第三周藻类样本;(d)第三周之后的藻类样本

Fig. 7　Comparison of classification results of various models under different growth periods. (a) Algae samples of first week; (b) algae samples of second week; (c) algae samples of third week; (d) algae samples after third week

表 2　不同模型的评价指标对比

Table 2　Comparison of evaluation indexes of different models

Model	*A*	*R*	*F*	Accuracy
GBDT-LR	0.982	1.000	0.973	0.995
LR	0.929	0.944	0.895	0.736
GBDT	0.947	0.889	0.914	0.991
SVM	0.807	0.389	0.560	0.956
BP	0.929	0.778	0.875	0.848

由图 6 可以看出,GBDT-LR 模型在测试集上的表现优于其他模型,经过计算,GBDT-LR 模型的准确率为 0.995,LR 模型的准确率为 0.736,GBDT 模型的准确率为 0.991,SVM 模型的准确率为 0.956,BP 模型的准确率为 0.848。从图 7(d)可以看出,各模型在藻类生长末期(第 3 周之后)的分类准确率较低,经过计算,处于生长末期的藻类的平均分类准确率为 0.859,其中 LR 模型有 2 次错误分类,GBDT 模型有 2 次错误分类,SVM 模型有 5 次错误分类,BP 模型有 3 次错误分类,GBDT-LR 没有出现错误分类。分析其原因可能是:处于生长末期的藻类细胞浓度较高,从而发生荧光猝灭现象,这会导致荧光强度骤降,影响判别结果[25]。在第一、二、三周,模型的平均分类准确率分别为 0.920、0.917、0.978,这一时期藻类生长情况良好,荧光数据较为稳定。其中,第三周处于藻类指数生长期,这一时期仅 SVM 模型出现 2 次错误分类,其他模型未出现错误分类。从表 2 可以看出,GBDT-LR 模型的准确率、召回率、F1 分数和准确率分别为 0.982、1.000、0.973、0.995,各项评价指标均优于其他几种模型。综上所述,GBDT-LR 模型在褐潮藻辨识中的效果最好,符合海洋环境监测系统的设计要求。另一方面,良好的分类效果表明三维荧光光谱与 GBDT-LR 结合能够准确识别具有重叠光谱的物质。

5　结　　论

建立了一种基于三维荧光光谱和 GBDT-LR 的褐潮藻辨识方法。为提高 LR 模型在非线性数据上的特征组合能力,引入 GBDT 对原始光谱数据进行特征提取。首先通过构造学习曲线确定 GBDT 模型最佳参数,将决策树叶节点位置信息组成的稀疏矩阵作为 LR 模型的输入,最终的分类准确率为 0.982,召回率为 1.000,这表明 GBDT-LR 模型可以对褐潮藻进行准确识别。通过对不同生长周期的藻类荧光光谱数据进行对比,发现处于指数生长期的藻类的荧光光谱稳定性最好。为验证模型的可行性,将 GBDT-LR 与 GBDT、SVM 和 BP 等几种算法进行横向对比。结果表明,GBDT-LR 模型的各

项评价指标均优于其他方法，该方法为褐潮污染辨识工作提供了一种有效的技术参考。

参 考 文 献

[1] Domangue R J, Mortazavi B. Nitrate reduction pathways in the presence of excess nitrogen in a shallow eutrophic estuary [J]. Environmental Pollution, 2018, 238: 599-606.

[2] Qing S, Runa A, Shun B R, et al. Distinguishing and mapping of aquatic vegetations and yellow algae bloom with Landsat satellite data in a complex shallow Lake, China during 1986 — 2018 [J]. Ecological Indicators, 2020, 112: 106073.

[3] 张建乐，王全颖，张永丰，等. 秦皇岛海域褐潮生消过程中营养盐特征[J]. 应用生态学报，2020，31(1)：282-292.
Zhang J L, Wang Q Y, Zhang Y F, et al. Characteristics of seawater nutrients during the occurrence of brown tide in the coastal area of Qinhuangdao, China[J]. Chinese Journal of Applied Ecology, 2020, 31(1): 282-292.

[4] 俞志明，陈楠生. 国内外赤潮的发展趋势与研究热点[J]. 海洋与湖沼，2019，50(3)：474-486.
Yu Z M, Chen N S. Emerging trends in red tide and major research progresses [J]. Oceanologia et Limnologia Sinica, 2019, 50(3): 474-486.

[5] 乔玲，甄毓，米铁柱. 抑食金球藻(*Aureococcus anophagefferens*)褐潮研究概述[J]. 海洋环境科学，2016，35(3)：473-480.
Qiao L, Zhen Y, Mi T Z. Review of the brown tides caused by *Aureococcus anophagefferens* [J]. Marine Environmental Science, 2016, 35(3): 473-480.

[6] 于湖洋，崔磊，潘霖，等. 秦皇岛海域浮游植物的群落结构特征[J]. 海洋科学，2016，40(5)：66-75.
Yu H Y, Cui L, Pan L, et al. Characteristics of the phytoplankton community structure in the Qinhuangdao coastal area [J]. Marine Sciences, 2016, 40(5): 66-75.

[7] 张婷婷，刘晶，张莉. 直接显微镜菌落观察法诊断念珠菌[J]. 泰山医学院学报，2014，35(12)：1272-1273.
Zhang T T, Liu J, Zhang L. Diagnosis of *Candida* by direct microscopic colony observation[J]. Journal of Taishan Medical College, 2014, 35(12): 1272-1273.

[8] Osterrothová K, Culka A, Němečková K, et al. Analyzing carotenoids of snow algae by Raman microspectroscopy and high-performance liquid chromatography [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 212: 262-271.

[9] 杨寿勇，张海阳，李成，等. 基于卷积神经网络模型的微藻种类识别[J]. 环境科学与技术，2020，43(S2)：158-164.
Yang S Y, Zhang H Y, Li C, et al. Recognition of microalgae species based on convolutional neural network model [J]. Environmental Science & Technology, 2020, 43(S2): 158-164.

[10] 甄毓，于志刚，蔡青松，等. 运用双特异分子探针技术对胶州湾三种硅藻的检测[J]. 海洋与湖沼，2010，41(1)：24-32.
Zhen Y, Yu Z G, Cai Q S, et al. Detection of three diatom species in Jiaozhou Bay using sandwich hybridization integrated with nuclease protection assay[J]. Oceanologia et Limnologia Sinica, 2010, 41(1): 24-32.

[11] Zhao N J, Zhang X L, Yin G F, et al. On-line analysis of algae in water by discrete three-dimensional fluorescence spectroscopy [J]. Optics Express, 2018, 26(6): A251-A259.

[12] 程钊，赵南京，殷高方，等. 基于SWTATLD算法的藻类群落离散三维荧光光谱识别方法[J]. 光学学报，2021，41(14)：1430001.
Cheng Z, Zhao N J, Yin G F, et al. Identification of algae community discrete three-dimensional fluorescence spectrum based on SWTATLD[J]. Acta Optica Sinica, 2021, 41(14): 1430001.

[13] 李潇凡，王胜强，翁轩，等. 基于UNet深度学习算法的东海大型漂浮藻类遥感监测[J]. 光学学报，2021，41(2)：0201002.
Li X F, Wang S Q, Weng X, et al. Remote sensing of floating macroalgae blooms in the East China Sea based on UNet deep learning model[J]. Acta Optica Sinica, 2021, 41(2): 0201002.

[14] Bruckman L S, Richardson T L, Swanstrom J A, et al. Linear discriminant analysis of single-cell fluorescence excitation spectra of five phytoplankton species[J]. Applied Spectroscopy, 2012, 66(1): 60-65.

[15] 苏荣国，梁生康，胡序朋，等. 我国东海常见6种有毒赤潮藻的三维荧光光谱识别技术[J]. 海洋环境科学，2008，27(3)：265-268.
Su R G, Liang S K, Hu X P, et al. Discrimination of 6 toxic red tide algae occurred in East China Sea by 3D fluorescence spectra [J]. Marine Environmental Science, 2008, 27(3): 265-268.

[16] 齐晓丽，吴珍珍，张传松，等. 基于支持向量机回归的3种常见有毒赤潮藻荧光识别技术[J]. 中国海洋大学学报(自然科学版)，2016，46(12)：73-80.
Qi X L, Wu Z Z, Zhang C S, et al. A fluorescence technology for discriminating toxic algae by support

sector machine regression[J]. Periodical of Ocean University of China, 2016, 46(12): 73-80.

[17] 王娟, 赵吉祥, 单春芝, 等. 基于集成学习的海岸带变化检测方法研究[J]. 海洋开发与管理, 2021, 38(7): 48-54.
Wang J, Zhao J X, Shan C Z, et al. Research on coastal zone change detection method based on ensemble learning[J]. Ocean Development and Management, 2021, 38(7): 48-54.

[18] 葛文杰, 陈龙. 基于随机森林与多源信息融合的疲劳驾驶检测方法[J]. 软件导刊, 2021, 20(10): 73-77.
Ge W J, Chen L. Fatigue driving detection method based on random forest and multi-source information fusion[J]. Software Guide, 2021, 20(10): 73-77.

[19] 周杰英, 贺鹏飞, 邱荣发, 等. 融合随机森林和梯度提升树的入侵检测研究[J]. 软件学报, 2021, 32(10): 3254-3265.
Zhou J Y, He P F, Qiu R F, et al. Research on intrusion detection based on random forest and gradient boosting tree[J]. Journal of Software, 2021, 32(10): 3254-3265.

[20] 王坤, 蒋宁, 李敏, 等. 基于 SMOTE 算法和逻辑回归模型算法的江苏短时强降水潜势预报[J]. 科学技术与工程, 2020, 20(28): 11447-11454.
Wang K, Jiang N, Li M, et al. The potential forecast for short-term heavy precipitation in Jiangsu Province based on SMOTE and logistic regression combination algorithm[J]. Science Technology and Engineering, 2020, 20(28): 11447-11454.

[21] 常鈺迪. 基于稀疏逻辑回归的链接模型在分类问题的应用[J]. 软件工程, 2021, 24(6): 2-5.
Chang Y D. Application of link model based on sparse logistic regression in classification problem[J]. Software Engineering, 2021, 24(6): 2-5.

[22] 徐胜. 基于三波长荧光光谱的浮游藻测量方法研究[D]. 杭州: 浙江大学, 2019.
Xu S. Research on measurement method of planktonic algae based on three-wavelength fluorescence spectrometry[D]. Hangzhou: Zhejiang University, 2019.

[23] Li G, Wang H, Liu H K, et al. Classification of grounding system defects in cross-bonded HV cables based on logistic regression[J]. High Voltage Engineering, 2021, 47(10): 3674-3683.
李根, 王航, 刘海康, 等. 基于逻辑回归的高压电缆交叉互联接地系统缺陷分类识别方法[J]. 高电压技术, 2021, 47(10): 3674-3683.

[24] 李新春, 赵忠婷, 于洪仕. 基于局部线性嵌入和梯度提升决策树的信道状态信息室内指纹定位算法研究[J]. 激光与光电子学进展, 2022, 59(2): 0215008.
Li X C, Zhao Z T, Yu H S. Channel state information indoor fingerprint localization algorithm based on locally linear embedding and gradient boosting decision tree[J]. Laser & Optoelectronics Progress, 2022, 59(2): 0215008.

[25] Rousso B Z, Bertone E, Stewart R A, et al. Light-induced fluorescence quenching leads to errors in sensor measurements of phytoplankton chlorophyll and phycocyanin[J]. Water Research, 2021, 198: 117133.