

# 基于 LLE 与 Gold-SA-SVM 的海藻三维荧光光谱门类识别

刘喆, 孟辉, 张永彬, 段玮靓, 陈颖\*

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

**摘要** 针对海藻三维荧光光谱线性降维方法不理想、模型识别准确率低的问题, 提出通过局部线性嵌入(LLE)算法进行降维、通过黄金正弦算法(Gold-SA)对支持向量机(SVM)进行优化来构建分类模型的方法。将 LLE 算法降维后的海藻三维荧光光谱数据作为 SVM 的输入, 并与其他两种降维方法进行对比。结果显示: LLE 算法的降维效果最好, 识别准确率最高。为了进一步提高门类识别能力, 采用 Gold-SA 对 SVM 进行优化并建立 Gold-SA-SVM 模型, 再将其与其他 4 种分类模型进行对比。结果显示, Gold-SA-SVM 模型在分类识别准确率、精准率、召回率和 F1 分数上都有明显提高, 该方法可准确识别抑食金球藻、小球藻、细长聚球藻的门类, 可为褐潮研究工作提供有效的参考。

**关键词** 光谱学; 三维荧光光谱; 褐潮; 局部线性嵌入; 支持向量机; 黄金正弦算法

中图分类号 X834

文献标志码 A

doi: 10.3788/AOS202242.0430001

## Category Recognition of Three-Dimensional Fluorescence Spectra of Algae Based on LLE and Gold-SA-SVM

Liu Zhe, Meng Hui, Zhang Yongbin, Duan Weiliang, Chen Ying\*

Key Laboratory of Test/Measurement Technology and Instrument of Hebei Province, School of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China

**Abstract** Aiming at the problems that the linear dimension reduction method of three-dimensional (3D) fluorescence spectra of algae is not ideal and the model recognition accuracy is low, a classification model is constructed by using local linear embedding (LLE) algorithm to reduce the dimension and using the golden sine algorithm (Gold-SA) to optimize the support vector machine (SVM). The 3D fluorescence spectrum data of algae after dimension reduction by LLE algorithm is used as the input of SVM, and other two dimension reduction methods are compared. The results show that LLE algorithm has the best dimension reduction effect and the highest recognition accuracy. In order to further improve category recognition ability, the Gold-SA is used to optimize SVM and establish a Gold-SA-SVM model, and the other four classification models are compared. The results show that the classification recognition accuracy, precision, recall rate, and F1 score of the Gold-SA-SVM model are significantly improved, and the method can accurately realize the classification of *Aureococcus anophagefferens*, *Chlorella*, and *Synechococcus elongatus*, providing an effective reference for the research of brown tide.

**Key words** spectroscopy; three-dimensional fluorescence spectra; brown tide; locally linear embedding; support vector machine; golden sine algorithm

收稿日期: 2021-08-02; 修回日期: 2021-08-21; 录用日期: 2021-08-31

基金项目: 国家重点研发计划(2018YFC1800903, 2016YFC1400601-3)、河北省重点研发计划(19273901D)、河北省自然科学基金(F2016203188)、中国博士后基金(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

# 1 引言

近年来我国渤海、黄海海域频繁爆发褐潮,其成因是褐潮爆发期间抑食金球藻在数量上占绝对优势,海面呈黄褐色或铁锈色长达几个月,抑食金球藻通过抑制贝类摄食,造成贝类大面积滞长甚至死亡,这会对养殖业和旅游业造成严重影响<sup>[1-4]</sup>。因此,对于抑食金球藻的识别在褐潮监测中尤为重要,三维荧光光谱技术以其快速性、无污染、无损检测等优点,近年来被广泛应用于海藻的门类识别研究中。

由于光谱存在大量冗余信息,一般需要进行降维处理。Bruckman 等<sup>[5]</sup>用线性判别分析法(LDA)对 5 种浮游植物的三维荧光光谱进行了降维和分类。苏荣国等<sup>[6-7]</sup>利用主成分分析法(PCA)对多种海藻的荧光光谱数据进行降维,分别采用非负最小二乘法(NNLS)和 Bayesian 判别法实现了海藻分类。由于 LDA 和 PCA 都是线性降维方法,降维后样本间的非线性相关性有可能会丢失,且 LDA 和 PCA 对异常值和噪声敏感,不适合对非高斯分布样本进行降维。局部线性嵌入(LLE)算法是一种非线性降维方法,能够有效保留光谱数据在高维空间的局部线性关系,对样本数据分布也无特殊要求,且会提高分类效果<sup>[8-9]</sup>。

上述研究中所采用的 NNLS 模型是线性分类模型,在处理非线性问题时效果不理想。Bayesian 判别法需要样本数据服从多元正态分布,各组协方差矩阵必须相等。这两种模型在进行分类时都具有局限性。支持向量机(SVM)模型近年来被广泛应用,通过引入核函数解决了非线性问题的分类,且对样本数据分布无特殊要求<sup>[10-11]</sup>。通过选用合适的优化算法对惩罚因子  $c$  和核函数参数  $g$  进行优化,可以提高模型准确率<sup>[12]</sup>。张希萌等<sup>[13-14]</sup>利用遗传算法(GA)优化 SVM,构建了 GA-SVM 分类模型,结合近红外光谱分别对黑木耳多糖含量等级和皮蛋凝胶品质进行了测量。谭念等<sup>[15-16]</sup>利用粒子群(PSO)算法优化 SVM,构建了 PSO-SVM 分类模型,并将其应用于木材近红外光谱树种识别和可见光光谱珍珠分级研究中。GA 算法虽然可以快速接近最优解,要达到真正的最优解却需要花费很长时间。PSO 算法的收敛速度快但容易陷入局部最优。黄金正弦算法(Gold-SA)在种群位置更新时引入黄金正弦系数,使收敛速度和全局最优达到有效的平衡,具有全局寻优能力强、收敛速度快、设置参数少等优点。

因此本文利用 LLE 算法对海藻三维荧光光谱

数据进行降维、利用 Gold-SA 对 SVM 参数进行优化,通过构建 Gold-SA-SVM 分类识别模型,实现抑食金球藻、小球藻、细长聚球藻的门类识别。

## 2 算法与模型

### 2.1 基于 LLE 算法的光谱数据降维

LLE 是针对非线性数据的一种降维方法。该算法假设数据在局部是线性的,即在  $m$  个  $n$  维样本  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$  中,  $\mathbf{x}_i$  可用  $k$  个邻域点  $\mathbf{x}_j$  线性表示为

$$\mathbf{x}_i \approx \sum_{j=1}^k \omega_{ij} \mathbf{x}_j, \quad (1)$$

式中:  $\omega_{ij}$  为权重系数,是使用样本点  $\mathbf{x}_i$  的  $k$  个邻域点  $\mathbf{x}_j$  重构样本点  $\mathbf{x}_i$  时误差  $J(\omega_{ij})$  最小的系数。 $J(\omega_{ij})$  可表示为

$$J(\omega_{ij}) = \sum_{i=1}^m \left\| \mathbf{x}_i - \sum_{j=1}^k \omega_{ij} \mathbf{x}_j \right\|_2^2. \quad (2)$$

LLE 算法假设得到的高维权重系数对应的线性关系在降维后的低维数据里能够保持。即假设  $n$  维样本集  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$  在低维的投影  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m\}$  依然保持线性关系,也就是对应的均方差损失函数最小,最小化损失函数  $J(\mathbf{Y})$  表示为

$$J(\mathbf{Y}) = \sum_{i=1}^m \left\| \mathbf{y}_i - \sum_{j=1}^k \omega_{ij} \mathbf{y}_j \right\| = \text{tr}[\mathbf{Y}^T (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W}) \mathbf{Y}] = \text{tr}(\mathbf{Y}^T \mathbf{M} \mathbf{Y}), \quad (3)$$

式中:  $\mathbf{W}$  为权重系数  $\omega_{ij}$  组成的权重矩阵,  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ , 约束条件为  $\sum_{i=1}^m \mathbf{y}_i = \mathbf{0}$ ;  $\frac{1}{m} \sum_{i=1}^m \mathbf{y}_i \mathbf{y}_i^T = \mathbf{I}$ , 即  $\mathbf{Y}^T \mathbf{Y} = m\mathbf{I}$ 。求出矩阵  $\mathbf{M}$  最小的  $d$  个非 0 特征值所对应的  $d$  个特征向量组成的矩阵  $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_d)$ , 即可得到最小的  $d$  维数据集。

### 2.2 基于黄金正弦算法的改进 SVM 模型

#### 2.2.1 黄金正弦算法

黄金正弦算法是 Tanyildizi 等<sup>[17]</sup>于 2017 年提出的一种基于种群的搜索算法,其灵感来源于正弦函数单位圆内扫描类似于待优化问题解的空间搜索。该算法利用黄金分割率缩小搜索空间,逼近算法的最优解,具有设置参数少、全局寻优能力强、收敛速度快等优点<sup>[18]</sup>。

黄金正弦算法从随机生成种群开始,通过随机生成每个维度的均匀分布更合理地分配搜索空间。种群初始化表示为

$$V_i = n_{\text{rand}} * (x_{\text{ub}} - x_{\text{lb}}) + x_{\text{lb}}, \quad (4)$$

式中:  $n_{\text{rand}}$  为随机数;  $x_{\text{ub}}$  为搜索空间上限值;  $x_{\text{lb}}$  为搜索空间下限值。

Gold-SA 使用黄金分割系数  $x_1, x_2$  对种群位置进行更新, 使搜索和开发达到更合理的分配, 黄金分割系数缩小了搜索空间, 使个体更快趋近最优值。  $x_1, x_2$  表示为

$$\begin{cases} x_1 = a(1-t) + bt \\ x_2 = at + b(1-t) \end{cases}, \quad (5)$$

式中:  $a, b$  为初始值, 一般取  $a = -\pi, b = \pi$ ;  $t$  为黄金分割比率,  $t = (\sqrt{5} - 1)/2$ 。

在迭代过程中, 黄金正弦算法的位置更新公式可表示为

$$V_i^{(t+1)} = V_i^{(t)} |\sin r_1| - r_2 \sin r_1 |x_1 D_i^{(t)} - x_2 V_i^{(t)}|, \quad (6)$$

式中:  $V_i^{(t+1)}$  表示第  $i$  个个体第  $t+1$  次迭代时的位

置;  $V_i^{(t)}$  表示第  $i$  个个体第  $t$  次迭代时的位置;  $D_i^{(t)}$  表示第  $i$  个个体第  $t$  次迭代的最优位置;  $r_1$  是  $[0, 2\pi]$  范围内的随机数;  $r_2$  是  $[0, \pi]$  范围内的随机数。

### 2.2.2 Gold-SA-SVM 分类模型

在 SVM 分类模型中, 惩罚因子  $c$  和核函数参数  $g$  对分类结果和模型复杂度有重要影响。为了进一步提高分类准确率, 利用黄金正弦算法对  $c, g$  参数进行优化。首先设置 Gold-SA 初始化参数, 针对 SVM 模型的参数调节寻优个数与范围, 设置 Gold-SA 的寻优个数为 2,  $c$  和  $g$  参数的范围均设置为  $[0.01, 100]$ , 设置种群数量为 20, 迭代次数为 50, 以分类准确率作为适应度函数, 结合 5 折交叉验证方法对模型进行训练, 当准确率最高时模型达到最优。Gold-SA-SVM 分类模型的具体步骤如图 1 所示。

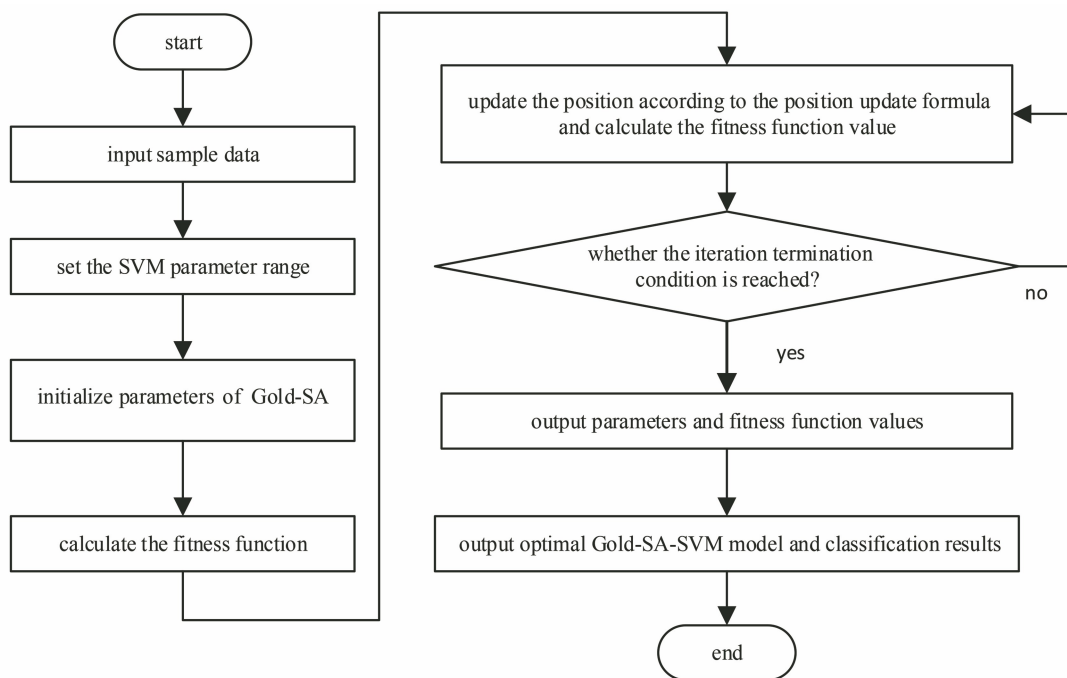


图 1 Gold-SA-SVM 分类模型流程图

Fig. 1 Flowchart of Gold-SA-SVM classification model

本文降维方法和分类识别模型构建均使用 MATLAB 仿真软件实现。

## 3 光谱数据采集

藻种选用褐潮成因藻抑食金球藻, 秦皇岛海域常见的绿藻门的小球藻、蓝藻门的细长聚球藻在形态和大小上和抑食金球藻十分接近, 利用显微镜很难将其区分, 且它们生存环境相似, 因此小球藻、细长聚球藻用来作为对比藻种。

使用英国 Edinburgh Instruments 公司生产的

FS920 稳态荧光光谱仪采集三种藻的荧光光谱。设置激发和发射狭缝宽度为 0.44 mm, 激发波长的扫描范围为 400~650 nm, 步长为 5 nm, 发射波长范围为 630~730 nm, 步长为 5 nm。为了避免一级瑞利散射的干扰, 设置发射波长滞后激发波长 10 nm。二级瑞利散射一般发生在发射波长是激发波长的两倍处, 本实验通过设置采样范围避开二级瑞利散射的干扰。对三种海藻每天采集一次荧光光谱, 得到每种藻在整个生长周期不同阶段的 90 个荧光光谱数据, 以采集到的 270 个样本数据作为研究对象。



## 4 光谱降维结果及对比分析

对采集到的光谱进行本底扣除处理和 S-G 平滑去噪处理,预处理后细长聚球藻、小球藻、抑食金球藻的三维荧光光谱图及等高线图如图 2 所示,分析其荧光光谱特性。

图 2(a)、图 2(b)为细长聚球藻的三维荧光光谱图和等高线图,激发波长为 625 nm、发射波长为 650 nm 处有一个荧光特征峰,恰为藻蓝蛋白对应

的激发、发射波长,可以看出细长聚球藻作为蓝藻时,细胞内藻蓝蛋白占绝对优势。图 2(c)、图 2(d)为小球藻的三维荧光光谱图和等高线图,在激发波长为 440 nm 和 470 nm 时,685 nm 发射波长处有两个荧光特征峰,分别对应叶绿素 a 和叶绿素 b。图 2(e)、图 2(f)为抑食金球藻的三维荧光光谱图和等高线图,荧光峰位于激发波长为 470 nm、发射波长为 685 nm 处,叶绿素 b 含量明显多于其他色素。

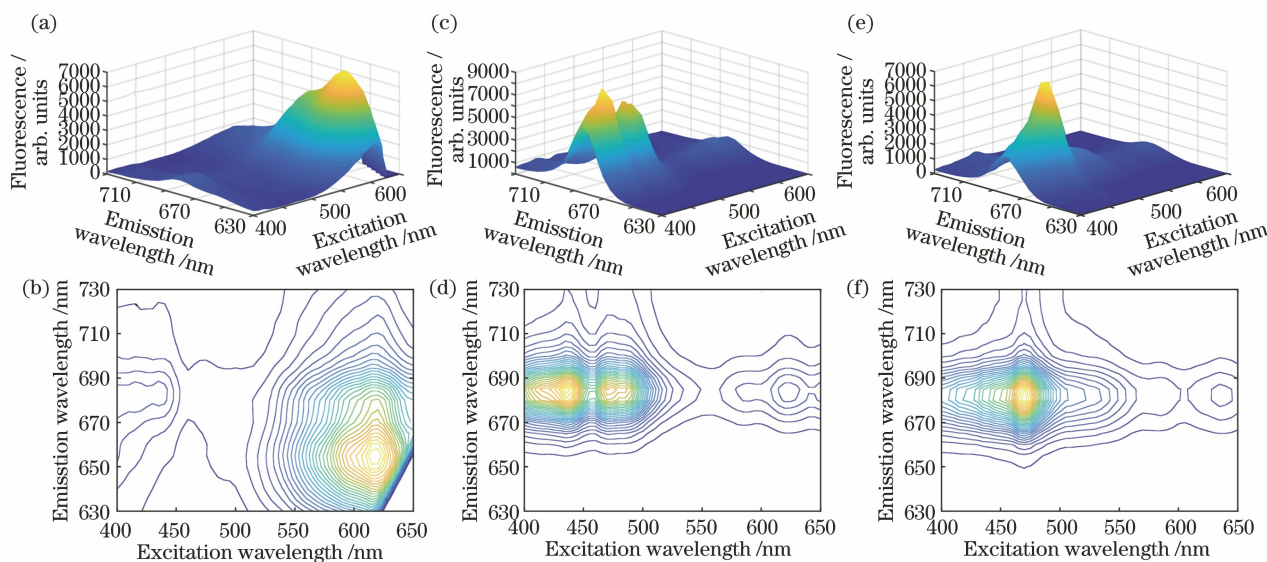


图 2 海藻三维荧光光谱图及等高线图。(a)细长聚球藻的三维荧光光谱图;(b)细长聚球藻的等高线图;

(c)小球藻的三维荧光光谱图;(d)小球藻的等高线图;(e)抑食金球藻的三维荧光光谱图;(f)抑食金球藻的等高线图

Fig. 2 3D fluorescence spectra and contour maps of algae. (a) 3D fluorescence spectrum of *Synechococcus elongatus*;

(b) contour map of *Synechococcus elongatus*; (c) 3D fluorescence spectrum of *Chlorella*; (d) contour map of *Chlorella*;

(e) 3D fluorescence spectrum of *Aureococcus anophagefferens*; (f) contour map of *Aureococcus anophagefferens*

在培养过程中,海藻荧光强度会随着海藻丰度的增加而增强,光谱形状也会随着内部色素的变化而发生变化。以小球藻为例,其生长周期不同阶段的三维荧光光谱图如图 3 所示。

由图 3 可知,小球藻在生长过程中的荧光强度不断增强,内部色素比例也在发生变化。由图 3(a)~(d)可知,在小球藻培养前期和中期,440 nm 激发波长处荧光峰强度明显大于 470 nm 激发波长处的荧光峰强度,叶绿素 a 的含量大于叶绿素 b 的含量。由图 3(e)、图 3(f)可知,在小球藻培养后期,440 nm 激发波长处荧光峰强度小于 470 nm 激发波长处荧光峰的强度,叶绿素 a 的含量低于叶绿素 b 的含量。

由图 2 和图 3 可以发现,不同海藻的荧光光谱图有较明显的差异,但其光谱数据庞大,分类识别运算量较大,且海藻生长周期不同阶段的色素变化也

会对分类识别难度产生影响。为了降低分类识别运算量,同时使海藻特征更加明显,需对光谱数据进行降维处理。

预处理后的光谱数据格式为  $21 \times 51$  的矩阵,将每个样本矩阵按列首尾相连,转换为  $1 \times 1071$  的数据形式,270 个海藻荧光光谱数据构成  $270 \times 1071$  的矩阵形式。采用 LLE 算法在此基础上对海藻荧光光谱数据进行降维,近邻点  $K$  和本征维度  $d$  的选择会对降维效果产生影响,取  $K$  的范围为  $5 \sim 20$ ,取  $d$  的范围为  $2 \sim 10$ ,以 SVM 为分类模型,分析 LLE 算法取不同近邻点和本征维度时的三种海藻的分类效果,标记细长聚球藻为第一类,小球藻为第二类,抑食金球藻为第三类。结合 5 折交叉验证法,将 270 个样本数据平均分成 5 份,轮流将其中 4 份作为训练集,1 份作为测试集,分别进行 5 组实验,以保证每一个样本都被作为训练数据和测试数据。

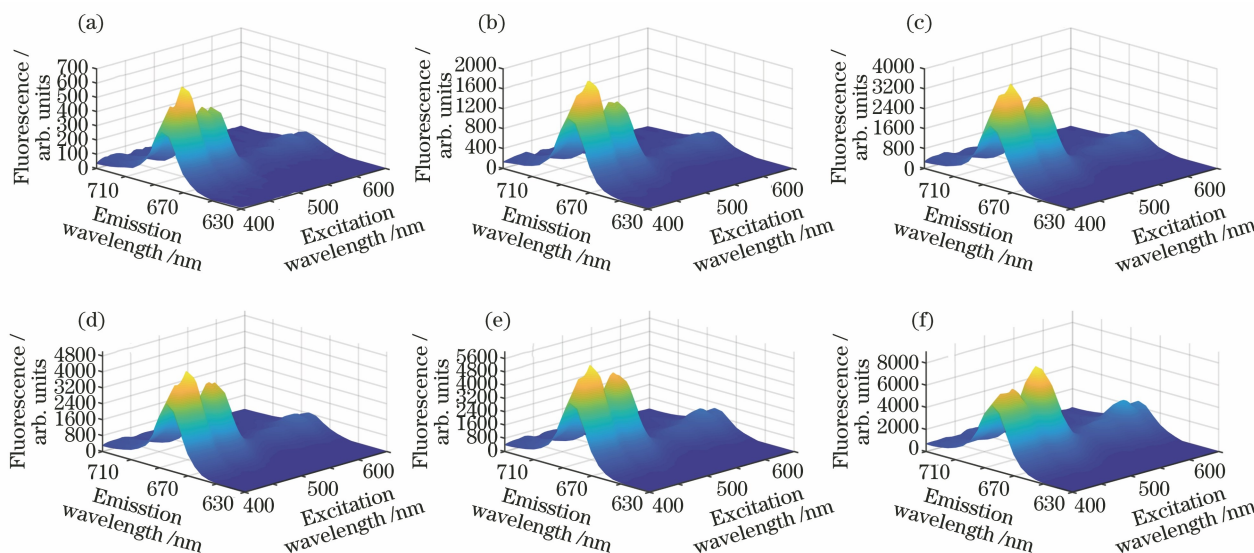


图 3 小球藻生长周期不同阶段的三维荧光光谱图。(a)第 1 天;(b)第 3 天;(c)第 6 天;(d)第 9 天;(e)第 12 天;(f)第 15 天  
Fig. 3 3D fluorescence spectra of *Chlorella* at different stages of growth cycle. (a) 1st day; (b) 3rd day; (c) 6th day; (d) 9th day; (e) 12th day; (f) 15th day

图 4 为 LLE-SVM 取不同近邻点  $K$  和本征维度  $d$  时准确率的波动情况。

由图 4 可知,局部线性嵌入算法取近邻点  $K = 8$ 、本征维数  $d = 5$  时降维效果最好,准确率为 96.3%,光谱数据经 LLE 算法降维后的部分样本特征值 PC1~PC5 如表 1 所示。PC1~PC5 是通过计算权重系数矩阵得到的特征值,是原始数据降维后的表现形式,其正负和数值大小代表了降维过程中近邻点之间的权重关系。

降维后抑食金球藻、小球藻、细长聚球藻在生长周期不同阶段的荧光光谱特征可视化结果如图 5 所示。

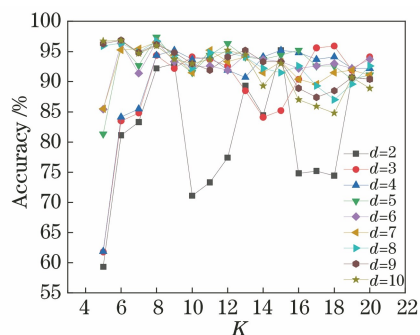


图 4 不同  $K, d$  下 LLE-SVM 的准确率变化情况  
Fig. 4 Change of accuracy of LLE-SVM for different  $K$  and  $d$

表 1 采用 LLE 算法得到的特征值

Table 1 Feature values extracted from LLE algorithm

Feature data	PC1	PC2	PC3	PC4	PC5
Sample 1	-1.00587	-2.54722	1.43672	0.32511	0.04593
Sample 2	-0.95251	-2.06625	0.91802	0.05146	-0.03223
Sample 3	-1.00584	-2.54735	1.44139	0.37201	0.08087
⋮	⋮	⋮	⋮	⋮	⋮
Sample 270	2.82157	-0.71266	0.50778	-2.63531	0.37356

由图 5(a)可知,在培养周期不同阶段的抑食金球藻的 PC2、PC3、PC4、PC5 的特征值变化明显,其中 PC2、PC3、PC4 的特征值随着培养天数的增加呈下降趋势,PC5 的特征值随着培养天数的增加呈上升趋势。由图 5(b)可知,在培养周期不同阶段的小球藻的 PC2、PC3、PC4、PC5 的特征值变化明显,其中 PC2、PC5 的特征值随着培养天数的增加呈上升

趋势,PC3、PC4 的特征值随着培养天数的增加呈下降趋势。由图 5(c)可知,在培养周期不同阶段的细长聚球藻的 PC1、PC3、PC4 的特征值变化明显,其中 PC1 的特征值随着培养天数的增加呈上升趋势,PC3 的特征值随着培养天数的增加呈下降趋势,但是在培养早期 PC1 的特征值变化不明显。PC4 的特征值在培养早期呈下降趋势,但是随着生长过程

中色素比例的变化,培养后期 PC1 的特征值变化规律不明显。由图 5 可以看出,海藻不同时期的特征

值有明显变化,且不同藻类的特征值变化规律明显不同,这有利于对海藻进行分类识别。

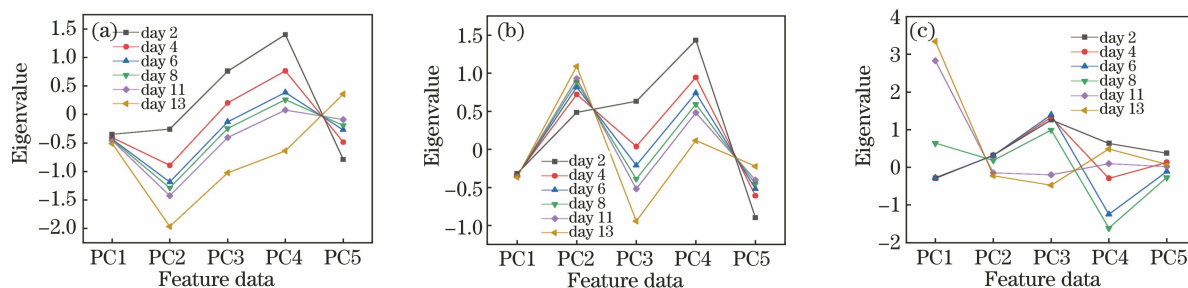


图 5 抑食金球藻、小球藻和细长聚球藻的荧光光谱特征。(a)抑食金球藻;(b)小球藻;(c)细长聚球藻  
Fig. 5 Fluorescence spectrum characteristics of *Aureococcus anophagefferens*, *Chlorella*, and *Synechococcus elongatus*. (a) *Aureococcus anophagefferens*; (b) *Chlorella*; (c) *Synechococcus elongatus*

仅以 SVM 作为门类识别模型,分别以原始光谱数据、文献[5]中采用的 LDA 降维的光谱数据、文献[6-7]中采用的 PCA 降维的光谱数据和 LLE 降维后的光谱数据作为输入,对比门类识别的准确率,结果如表 2 所示。可以发现,LLE 算法降维得到的光谱数据用于门类识别较其他两种数据更有效。与此同时,惩罚因子  $c$  和核函数参数  $g$  会影响 SVM 模型性能,SVM 作为门类识别模型,所得到的识别准确率仍有待提高。下面将引入 Gold-SA 对 SVM 的参数进行优化,并采用 Gold-SA-SVM 分类模型进行海藻门类识别。

表 2 不同降维方法的分类准确率

Table 2 Classification accuracy of different dimension reduction methods %

Dimension reduction method	LDA	PCA	LLE
Accuracy	95.1	95.6	96.3

## 5 海藻门类识别结果及对比分析

在对海藻门类识别方法的评价中,准确率  $A$  代表分类正确的样本数占样本总数的比例;精确率  $P$  代表在预测为正类的样本中真正类所占的比例;召回率  $R$  代表所有正类中被预测为正类的比例; $F_1$  分数  $F_1$  为精确率和召回率的调和平均,用于衡量模型的性能。以上各变量的计算公式可分别表示为

$$A = \left( \frac{x_{TP} + x_{TN}}{x_{TP} + x_{FP} + x_{TN} + x_{FN}} \right) \times 100\%, \quad (7)$$

$$P = \frac{x_{TP}}{x_{TP} + x_{FP}} \times 100\%, \quad (8)$$

$$R = \frac{x_{TP}}{x_{TP} + x_{FN}} \times 100\%, \quad (9)$$

$$F_1 = \frac{2PR}{P + R} \times 100\%, \quad (10)$$

式中: $x_{TP}$  代表真阳性; $x_{TN}$  代表真阴性; $x_{FP}$  代表假阳性; $x_{FN}$  代表假阴性。

选用准确率、精确率、召回率和  $F_1$  分数作为评价指标,将 LLE 算法采用  $K=8, d=5$  时降维的数据作为输入,利用 Gold-SA-SVM 模型进行分类,并与文献[6]中的 NNLS 模型、文献[7]中的 Bayesian 模型、文献[13-14]中的 GA-SVM 模型、文献[15-16]中的 PSO-SVM 模型进行对比。各模型分类结果的混淆矩阵如表 3 所示,评价指标对比如表 4 所示。

由表 3 和表 4 可以发现,NNLS 模型将 6 个小球藻识别为细长聚球藻,将 11 个小球藻识别为抑食金球藻,将 6 个抑食金球藻识别为细长聚球藻,模型准确率仅为 91.5%,该模型对细长聚球藻的识别精确率为 88.2%,对小球藻的召回率只有 81.1%,对抑食金球藻的识别精确率为 88.4%。三类藻的  $F_1$  分数分别为 93.8%, 89.6%, 90.8%,模型分类效果较差。Bayesian 判别模型将 1 个细长聚球藻识别为小球藻,将 1 个细长聚球藻识别为抑食金球藻,将 4 个小球藻识别为细长聚球藻,将 6 个小球藻识别为抑食金球藻,将 1 个抑食金球藻识别为细长聚球藻,模型准确率为 95.2%。Bayesian 判别模型对小球藻的识别效果不好,召回率只有 88.9%,其他评价指标都在 90% 以上。GA-SVM、PSO-SVM 分别将 7 个和 5 个样本错误识别,准确率分别为 97.4%、98.1%,二者对三类海藻的识别精确率、召回率和  $F_1$  分数都在 90% 以上,分类效果相对较好,说明采用 SVM 作为海藻门类识别模型是可行的。本文提出的 Gold-SA-SVM 模型对三类藻的识别都正确无误,准确率达到 100%,识别效果最好。



表 3 不同分类模型的混淆矩阵

Table 3 Confusion matrices of different classification models

Model	Actual	Predict		
		<i>Synechococcus elongatus</i>	<i>Chlorella</i>	<i>Aureococcus anophagefferens</i>
NNLS in Ref. [6]	<i>Synechococcus elongatus</i>	90	0	0
	<i>Chlorella</i>	6	73	11
	<i>Aureococcus anophagefferens</i>	6	0	84
Bayesian in Ref. [7]	<i>Synechococcus elongatus</i>	88	1	1
	<i>Chlorella</i>	4	80	6
	<i>Aureococcus anophagefferens</i>	1	0	89
GA-SVM in Refs. [13-14]	<i>Synechococcus elongatus</i>	90	0	0
	<i>Chlorella</i>	3	84	3
	<i>Aureococcus anophagefferens</i>	1	0	89
PSO-SVM in Refs. [15-16]	<i>Synechococcus elongatus</i>	90	0	0
	<i>Chlorella</i>	1	87	2
	<i>Aureococcus anophagefferens</i>	1	1	88
Gold-SA-SVM in this paper	<i>Synechococcus elongatus</i>	90	0	0
	<i>Chlorella</i>	0	90	0
	<i>Aureococcus anophagefferens</i>	0	0	90

表 4 不同模型分类结果评价指标对比

Table 4 Comparison of evaluation indexes of different classification models

%

Model	<i>Synechococcus elongatus</i>			<i>Chlorella</i>			<i>Aureococcus anophagefferens</i>			A
	P	R	F <sub>1</sub>	P	R	F <sub>1</sub>	P	R	F <sub>1</sub>	
Bayesian in Ref. [7]	94.6	98.8	92.7	97.8	88.9	98.9	96.2	93.6	95.7	95.2
GA-SVM in Refs. [13-14]	95.7	100.0	97.8	100.0	93.3	96.6	96.7	98.9	97.8	97.4
PSO-SVM in Refs. [15-16]	97.8	100.0	98.9	98.9	96.7	97.8	97.8	87.8	97.8	98.1
Gold-SA-SVM in this paper	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

## 6 结 论

通过采集抑食金球藻、小球藻、细长聚球藻的三维荧光光谱,结合 LLE 与 Gold-SA-SVM 模型,实现了对不同时期三种藻的门类鉴别。LLE 算法降维取得预期效果,将 LLE 算法降维后的数据作为支持向量机的输入,近邻点  $K=8$ 、本征维数  $d=5$  时模型的降维效果最好,准确率为 96.3%,优于 LDA 算法和 PCA 算法降维时的准确率。为了进一步提高分类效果,提出 Gold-SA-SVM 模型,将 NNLS 模型、Bayesian 判别模型、GA-SVM 模型、PSO-SVM 模型与 Gold-SA-SVM 模型进行对比,对比模型的准确率分别为 91.5%,95.2%,97.4%,98.1%,100%,Gold-SA-SVM 模型的识别准确率有明显提高。这表明本文提出的海藻三维荧光光谱门类识别方法可行,可为褐潮研究提供有效的参考。

## 参 考 文 献

- [1] 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.  
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.
- [2] Cira E K, Wetz M S. Spatial-temporal distribution of *Aureococcus lagunensis* ("brown tide") in Baffin Bay, Texas[J]. Harmful Algae, 2019, 89: 101669.
- [3] He F, Xie L, Xu X J, et al. Effect of *Aureococcus anophagefferens* (brown tide) on growth and feeding of the mussels *Perna Viridis* [J]. Marine Environmental Science, 2019, 38(4): 489-493, 502.  
贺凡, 谢磊, 徐晓娇, 等. 抑食金球藻对翡翠贻贝生长及摄食的影响研究[J]. 海洋环境科学, 2019, 38(4): 489-493, 502.
- [4] Jiang B, Song G J. Marine biological disasters in

- Bohai Sea and countermeasures for prevention and control[J]. Hebei Fisheries, 2019(6): 51-53.  
姜冰, 宋广军. 渤海海洋生物灾害及防控管理对策[J]. 河北渔业, 2019(6): 51-53.
- [5] 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.
- [6] 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.  
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.
- [7] Liu N. Assessment of marine phytoplankton using *in vivo* excitation-emission matrix fluorescence spectra[D]. Qingdao: Ocean University of China, 2008.  
刘娜. 基于荧光特征的海洋浮游植物快速分类方法研究[D]. 青岛: 中国海洋大学, 2008.
- [8] Zhang J J, Zhou X Y, Liu Q. Improved dimensionality reduction algorithm of large-scale hyperspectral scenes using manifold[J]. Acta Optica Sinica, 2013, 33(11): 1128001.  
张晶晶, 周晓勇, 刘奇. 一种改进的大尺度高光谱流形降维算法[J]. 光学学报, 2013, 33(11): 1128001.
- [9] Zhuo L, Cheng B, Zhang J. A comparative study of dimensionality reduction methods for large-scale image retrieval[J]. Neurocomputing, 2014, 141: 202-210.
- [10] Sha W, Li J T, Lu C P. Quantitative analysis of Mn in soil based on laser-induced breakdown spectroscopy optimization[J]. Chinese Journal of Lasers, 2020, 47(5): 0511001.  
沙文, 李江涛, 鲁翠萍. 基于激光诱导击穿光谱技术寻优定量分析土壤中 Mn 元素[J]. 中国激光, 2020, 47(5): 0511001.
- [11] Chen X L, Huang M Z, Fu Y C. Recognition method of dorsal hand vein with liveness detection function[J]. Acta Optica Sinica, 2021, 41(6): 0610002.  
陈秀莲, 黄梅珍, 富雨超. 具有活体检测功能的手背静脉身份识别方法研究[J]. 光学学报, 2021, 41(6): 0610002.
- [12] Zheng K, Chen Y B, Jiang Y, et al. A SVM based ship collision risk assessment algorithm[J]. Ocean Engineering, 2020, 202: 107062.
- [13] Zhang X M. Research on quality classification of auricularia auricula polysaccharide based on support vector machine optimized by genetic algorithm[D]. Harbin: Northeast Forestry University, 2019: 30-37.  
张希萌. 基于 GA-SVM 的黑木耳多糖品质分类研究[D]. 哈尔滨: 东北林业大学, 2019: 30-37.
- [14] Wang Q H, Mei L, Ma M H, et al. Nondestructive testing and grading of preserved duck eggs based on machine vision and near-infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(24): 314-321.  
王巧华, 梅璐, 马美湖, 等. 利用机器视觉与近红外光谱技术的皮蛋无损检测与分级[J]. 农业工程学报, 2019, 35(24): 314-321.
- [15] Tan N, Sun Y D, Wang X S, et al. Research on near infrared spectrum with principal component analysis and support vector machine for timber identification[J]. Spectroscopy and Spectral Analysis, 2017, 37(11): 3370-3374.  
谭念, 孙一丹, 王学顺, 等. 基于主成分分析和支持向量机的木材近红外光谱树种识别研究[J]. 光谱学与光谱分析, 2017, 37(11): 3370-3374.
- [16] Peng J, Gong X F, Luo R S, et al. Pearl rapid grading method based on visible spectrum[J]. Computer Applications and Software, 2021, 38(6): 143-148.  
彭杰, 龚晓峰, 雒瑞森, 等. 基于可见光光谱的珍珠快速分级方法研究[J]. 计算机应用与软件, 2021, 38(6): 143-148.
- [17] Tanyildizi E, Demir G. Golden sine algorithm: a novel math-inspired algorithm[J]. Advances in Electrical and Computer Engineering, 2017, 17(2): 71-78.
- [18] Zhou Y R, Li N, Zhou F H. Application of golden sine algorithm in hydrogeological parameter optimization[J]. Pearl River, 2020, 41(6): 117-120, 128.  
周有荣, 李娜, 周发辉. 黄金正弦算法在水文地质参数优化中的应用[J]. 人民珠江, 2020, 41(6): 117-120, 128.