

# 三维荧光光谱结合 2DPCA-SSA-GRNN 对柴油占比的检测

陈晓玉<sup>1</sup>, 杜雅欣<sup>1</sup>, 刘亚茹<sup>2</sup>, 孔德明<sup>2\*</sup>

<sup>1</sup>燕山大学信息科学与工程学院, 河北 秦皇岛 066004;

<sup>2</sup>燕山大学电气工程学院, 河北 秦皇岛 066004

**摘要** 柴油和生物柴油混合油液三维荧光光谱中的荧光峰较多, 导致混合油液的主要特征峰不明显, 而且随着柴油占比的减少, 三维荧光光谱的最大荧光强度位置会发生偏移, 荧光强度与柴油占比不满足线性关系。因此, 检测混合油液中的柴油占比较为复杂。本团队采用二维主成分分析(2DPCA)对混合油液的三维荧光光谱进行重构, 重构荧光光谱中的冗余信息减少, 在最佳激发波长 450 nm 下有能代表柴油的发射光谱(465~500 nm)。利用麻雀搜索算法(SSA)对广义回归神经网络(GRNN)进行优化, 构建 2DPCA-SSA-GRNN 预测网络, 该网络输入是训练集中的 9 个样本经 2DPCA 重构后能代表柴油的发射光谱, 网络输出是柴油占比。最后利用建立的网络预测测试集中 4 个样本的柴油占比, 柴油占比分为 85.02%、73.76%、63.80%、53.37%, 平均回收率为 98.39%, 均方根误差为 0.90%, 预测效果较未利用重构发射光谱的网络具有较大提升, 均方误差降低了 0.97 个百分点, 平均回收率提高了 1.24 个百分点。本文为优化神经网络预测物质占比提供了新方法。

**关键词** 光谱学; 三维荧光光谱; 二维主成分分析; 柴油; 广义回归神经网络; 麻雀搜索算法

中图分类号 O443.4

文献标志码 A

DOI: 10.3788/CJL202249.1811002

## 1 引言

近几十年以来, 石油资源过度开发的问题引起了人们的广泛关注, 急需可替代的新型燃料以缓解这一问题。生物柴油具有与柴油相似的理化性质, 而且具有较好的环境效益, 因此人们通常将其与柴油混合后作为柴油的替代产品, 以减少柴油的实际消耗。然而, 柴油在混合油液中的占比对该燃料转化的动力存在较大影响, 若柴油占比过低, 就会出现油雾化不良、油气混合不充分、发动机压力不足以及二氧化氮排放量增加等问题<sup>[1]</sup>。目前, 虽然各国对柴油在混合燃料中的体积比存在不同要求, 但通常要求这一比值要达到 50% 以上<sup>[2]</sup>。因此, 对柴油和生物柴油混合油液中柴油的体积比(以下简称“柴油占比”)进行快速、有效辨别具有重要意义。

目前, 国内外检测柴油占比的方法有色谱法、核磁共振波谱法、红外光谱法、荧光光谱法以及紫外和可见光谱法等<sup>[3]</sup>。其中, 荧光光谱法中的三维荧光光谱法因具有简单、便捷、灵敏度高等优势而被广泛应用于各个领域<sup>[4-7]</sup>。2004 年, 申金媛等<sup>[8]</sup>基于神经网络对非线性荧光光谱进行了分析。王书涛等<sup>[9-10]</sup>在此基础上

通过智能算法优化神经网络提高了定量预测物质占比的精确度, 但是他们在建立神经网络时仅选取了荧光强度较高的发射光谱, 这种方法具有较强的主观性, 当荧光光谱较为复杂时, 主观选择的光谱会导致预测精度下降。Yang 等<sup>[11]</sup>提出的 2DPCA 是基于矩阵的二维主成分分析方法, 本团队利用该方法对三维荧光光谱进行重构, 重构后的荧光光谱既保留了光谱的原始信息, 又选取了能代表光谱特征信息的荧光峰。广义回归神经网络(GRNN)以其对非线性关系的良好处理效果而常被用于物质占比的预测分析。吴希军等<sup>[12]</sup>利用广义回归神经网络实现了掺伪芝麻油的定量分析。但是, 广义回归神经网络参数(光滑因子)对预测结果的影响比较大, Xue 等<sup>[13]</sup>提出的麻雀搜索算法(SSA)能够对广义回归神经网络参数进行优化, 提高其预测效果。麻雀搜索算法的收敛速度快, 易找到全局最优解。

本团队采用 2DPCA 对柴油与生物柴油混合油液的三维荧光光谱进行重构, 该方法可以减少冗余信息, 提高峰值拾取方法的预测能力; 然后利用混合油重构光谱中与柴油光谱拟合较好的发射光谱建立麻雀搜索算法优化后的广义回归神经网络 2DPCA-SSA-

收稿日期: 2021-12-24; 修回日期: 2022-02-02; 录用日期: 2022-02-21

基金项目: 国家自然科学基金(62173289)

通信作者: \*demingkong@ysu.edu.cn

GRNN, 并采用该网络实现了混合油液中柴油占比的准确检测。该方法具有方便快捷、无污染、无损耗的优点, 为物质含量检测提供了新思路。

## 2 基本原理

### 2.1 2DPCA

2DPCA 是一种基于主成分分析的数据降维方法<sup>[11]</sup>, 它不仅可以直接对原始的光谱图像矩阵进行特征提取以及图像重构, 使光谱图像的主要特征更突出, 还可以有效减少冗余信息。2DPCA 的计算过程如下:

对于测定数据集  $\mathbf{A}$  ( $61 \times 61$ ), 若要求其投影特征向量  $\mathbf{Y}$ , 首先要寻找最佳投影空间  $\mathbf{X}$ 。  $\mathbf{A}$ 、 $\mathbf{Y}$ 、 $\mathbf{X}$  满足

$$\mathbf{Y} = \mathbf{A}\mathbf{X}. \quad (1)$$

寻找最佳投影空间  $\mathbf{X}$  的表达式为

$$\mathbf{J}(\mathbf{X}) = \text{tr}(\mathbf{S}_X), \quad (2)$$

其中,

$$\mathbf{S}_X = \mathbf{E}\{[\mathbf{A} - \mathbf{E}(\mathbf{A})\mathbf{X}][\mathbf{A} - \mathbf{E}(\mathbf{A})\mathbf{X}]^T\}, \quad (3)$$

式中:  $\mathbf{S}_X$  是  $\mathbf{Y}$  的协方差矩阵;  $\text{tr}(\mathbf{S}_X)$  是  $\mathbf{S}_X$  的迹;  $\mathbf{E}(\cdot)$  表示求均值。

令  $\mathbf{G}_i$  表示样本的协方差矩阵,

$$\mathbf{G}_i = \frac{1}{N} \sum_{i=1}^N (\mathbf{A}_i - \bar{\mathbf{A}})^T (\mathbf{A}_i - \bar{\mathbf{A}}), \quad (4)$$

式中:  $N$  表示样本数量;  $\bar{\mathbf{A}}$  表示样本的平均光谱, 即

$$\bar{\mathbf{A}} = \frac{1}{N} \sum_{i=1}^N \mathbf{A}_i. \quad (5)$$

由  $\mathbf{S}_X$  和  $\mathbf{G}_i$  可以得到广义总体散布函数  $\mathbf{J}(\mathbf{X})$  为

$$\mathbf{J}(\mathbf{X}) = \mathbf{X}^T \mathbf{E}\{[\mathbf{X} - \mathbf{E}(\mathbf{X})]^T [\mathbf{X} - \mathbf{E}(\mathbf{X})]\}, \quad (6)$$

$\mathbf{J}(\mathbf{X})$  最大化时对应的投影空间  $\mathbf{X}$  就是最佳投影空间。

选取  $\mathbf{G}_i$  中前  $m$  个最大特征值对应的特征向量  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$ , 并且这些特征向量满足标准正交化, 即满足

$$\begin{cases} \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m\} = \text{argmax}[\mathbf{J}(\mathbf{X})] \\ \mathbf{X}_i^T \mathbf{X}_j = 0, i \neq j, j = 1, \dots, m \end{cases}. \quad (7)$$

将样本光谱  $\mathbf{A}_i$  ( $i = 1, \dots, N$ ) 向最佳向量  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$  投影, 得到的投影特征向量  $\mathbf{Y}_i$  是样本光谱的主要成分向量, 即

$$\mathbf{Y}_i = \mathbf{A}_i [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m]. \quad (8)$$

重构图像  $\mathbf{A}'_i$  为

$$\mathbf{A}'_i = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m] [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m]^T. \quad (9)$$

### 2.2 麻雀搜索算法-广义回归神经网络 (SSA-GRNN)

#### 2.2.1 广义回归神经网络

广义回归神经网络<sup>[12]</sup> 是一种径向基神经网络, 对非线性问题具有良好的逼近能力和预测能力。广义回归神经网络参数 (光滑因子) 的传统寻找方法是遍历法, 这种操作不够便捷, 容易错过最优值。如何寻找最佳光滑因子是提高广义回归神经网络预测的关键。

#### 2.2.2 麻雀搜索算法

麻雀搜索算法<sup>[13]</sup> 主要是利用三种类型的麻雀 (发现者、加入者和侦察者) 分工协作找到全局最优解。发现者负责为加入者提供搜索范围和方向; 加入者寻找食物, 并监视发现者争夺食物; 而侦察者的存在是为了保证整体麻雀的捕食率, 确保雀群在安全范围内进行捕食。

麻雀组成的种群可以表示为

$$\mathbf{X} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^d \\ x_2^1 & x_2^2 & \dots & x_2^d \\ \vdots & \vdots & & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^d \end{bmatrix}, \quad (10)$$

式中:  $d$  表示维数;  $n$  表示麻雀数量。

发现者的位置更新公式为

$$x_{i,j}(t+1) = \begin{cases} x_{i,j}(t) \cdot \exp\left(\frac{-i}{\alpha \cdot T}\right), R_2 < S_T \\ x_{i,j}(t) + Q, R_2 \geq S_T \end{cases}, \quad (11)$$

式中:  $t$  表示当前迭代次数;  $T$  表示最大的迭代次数;  $\alpha$  为  $[0, 1]$  范围内的随机数;  $x_{i,j}(t)$  表示第  $i$  只麻雀在第  $j$  维迭代次数为  $t$  时的位置信息;  $R_2$  表示预警值,  $R_2 \in [0, 1]$ ;  $S_T$  表示安全值,  $S_T \in [0.5, 1]$ ;  $Q$  为服从正态分布的随机数。当  $R_2 \geq S_T$  时, 侦察者发现捕食者, 侦察者向雀群发出信号, 发现者飞往安全区域; 当  $R_2 < S_T$ , 表示周围没有捕食者, 发现者可以继续搜食。

加入者的位置更新公式为

$$x_{i,j}(t+1) = \begin{cases} Q \cdot \exp\left[\frac{x_{\text{worst}}(t) - x_{i,j}(t)}{i^2}\right], i > \frac{n}{2} \\ x_{\text{best},j}(t+1) + |x_{i,j}(t) - x_{\text{best},j}(t+1)| \cdot \mathbf{B}^+ \cdot \mathbf{L}, \text{ others} \end{cases}, \quad (12)$$

式中:  $x_{\text{worst}}(t)$  表示当前全局最差的位置;  $x_{i,j}(t)$  表示发现者占据的最佳位置;  $\mathbf{B}^+ = \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1}$ , 其中  $\mathbf{B}$  表示一个内部元素随机分配为 1 或 -1 的  $1 \times d$  矩阵;  $\mathbf{L}$  表示一个全为 1 的  $1 \times d$  的矩阵;  $x_{\text{best},j}(t+1)$  为更新后的全局最优位置。当  $i > \frac{n}{2}$  时, 表示适应度较差的第  $i$  只加入者处于饥饿状态, 需要到其他区域重新寻

找食物。

侦察者的位置更新公式为

$$x_{i,j}(t+1) = \begin{cases} x_{\text{best},j}(t) + \beta |x_{i,j}(t) - x_{\text{best},j}(t+1)|, f_i > f_g \\ x_{i,j}(t) + K \left[ \frac{|x_{i,j}(t) - x_{\text{worst},j}(t)|}{(f_i - f_w) + e} \right], f_i = f_g \end{cases}, \quad (13)$$

式中： $x_{best,j}(t)$ 表示当前全局最优位置； $f_g$ 和 $f_w$ 表示当前最优值和最劣值； $e$ 是一个常数，是避免分母为0而人为添加的； $\beta$ 是服从均值为0、方差为1的参数； $K$ 的大小表示麻雀的运动步长，正负表示麻雀运动的方向， $K \in [-1, 1]$ ； $f_i$ 表示随机重排采样的第*i*只麻雀的适应度。

麻雀的适应度可以表示为

$$F_x = \begin{bmatrix} f([x_1^1 & x_1^2 & \dots & x_1^d]) \\ f([x_2^1 & x_2^2 & \dots & x_2^d]) \\ \vdots \\ f([x_n^1 & x_n^2 & \dots & x_n^d]) \end{bmatrix} \quad (14)$$

在本团队的研究中，适应度值为折叠度  $Z=4$  时的交叉验证法对广义回归神经网络的预测误差。

### 2.2.3 SSA-GRNN 的建立

麻雀搜索算法模拟麻雀的觅食与反捕食行为，具有较强的搜索能力，能有效避免陷入局部最优。利用麻雀搜索算法寻找最合适的光滑因子，能够有效优化广义回归神经网络，使其预测效果更好。具体步骤如下：

1) 根据 Yang 等<sup>[11]</sup>提出的麻雀搜索算法的参数范围设置麻雀搜索算法的种群数为 20，最大迭代次数为 20，下边界为 0.1，上边界为 2，维度为 1，安全值  $S_T=0.6$ ，发现者在麻雀数量中的占比为 70%，警示者在麻雀数量中占比为 20%，加入者在麻雀数量中的占比为 30%。

2) 初始化种群，计算每只麻雀的适应度，对适应度排序，找出全局最优适应度值及其对应的麻雀的位置。

3) 利用式(11)~(13)逐步更新加入者、发现者和侦察者的位置。当发现者找到更好的食物时，加入者会根据式(12)立即离开前去抢夺。每次迭代随机产生一定比例的侦察者，更新侦察者的位置，保证发现者和加入者一直在安全范围内进行选择。发现者和加入者

的角色可以随意转换，但是所占比例不变。

4) 重新计算每只麻雀的适应度值，对适应度进行排序，寻找最佳适应度值，最佳适应度值(广义回归神经网络的预测误差最小)对应的麻雀位置即为所求光滑因子。

5) 将所求的光滑因子作为参数建立预测网络，对柴油占比进行预测。

### 2.3 品质指数

品质指数(FOM)是用来衡量结果优劣的评价指标。本文引入回收率和均方根误差(RMSE)来衡量结果的准确性，它们的表达式为

$$y_{recovery} = \frac{y'_i}{y_i} \times 100\%, \quad (15)$$

$$E_{RMS} = \frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2, \quad (16)$$

式中： $y_i$ 表示样本的实际占比； $y'_i$ 表示样本的预测占比； $N$ 代表样本个数。

## 3 实验研究

### 3.1 实验仪器设备与参数设置

本文采用 Edinburgh 公司生产的 FS920 稳态荧光光谱仪进行数据采集。仪器参数设置如下：激发波长扫描范围为 300~600 nm，步长 5 nm，发射波长扫描范围为 320~620 nm，步长 5 nm。为避免瑞利散射的影响，设置发射波长始终滞后激发波长 20 nm。

### 3.2 油液样本的配制

生物柴油由科瑞达化工有限公司生产，其主要原料为地沟油；柴油购于当地中国石油加油站。实验样本是将生物柴油和柴油按照不同比例混合而成的 10 mL 混合油液，混合结果如表 1 所示。将其中的 D85、D75、D65、D55 样本作为测试集，其余样本作为训练集。

表 1 样本的配比

Table 1 Proportion of each sample

Sample	Volume of diesel /mL	Volume of biodiesel/mL	Volume fraction of diesel /%	Sample	Volume of diesel /mL	Volume of biodiesel /mL	Volume fraction of diesel /%
D90	9.0	1.0	90	D55	5.5	4.5	55
D85	8.5	1.5	85	D50	5.0	5.0	50
D80	8.0	2.0	80	D40	4.0	6.0	40
D75	7.5	2.5	75	D30	3.0	7.0	30
D70	7.0	3.0	70	D20	2.0	8.0	20
D65	6.5	3.5	65	D10	1.0	9.0	10
D60	6.0	4.0	60				

## 4 结果与讨论

### 4.1 光谱特性分析

图 1 为柴油、生物柴油以及柴油与生物柴油混合油液 D10、D70 的三维荧光等高线图和三维荧光光谱。

柴油中含有大量烃类化合物，不同的成分对激发光的吸收效率不同，发射荧光效率也不同，故而构成了较为复杂的荧光光谱。由图 1(a1)、(b1)可知柴油的最佳荧光峰位于 450 nm/480 nm 处，由图 1(a2)、(b2)可知生物柴油的最佳荧光峰位于 370 nm/410 nm。由图 1

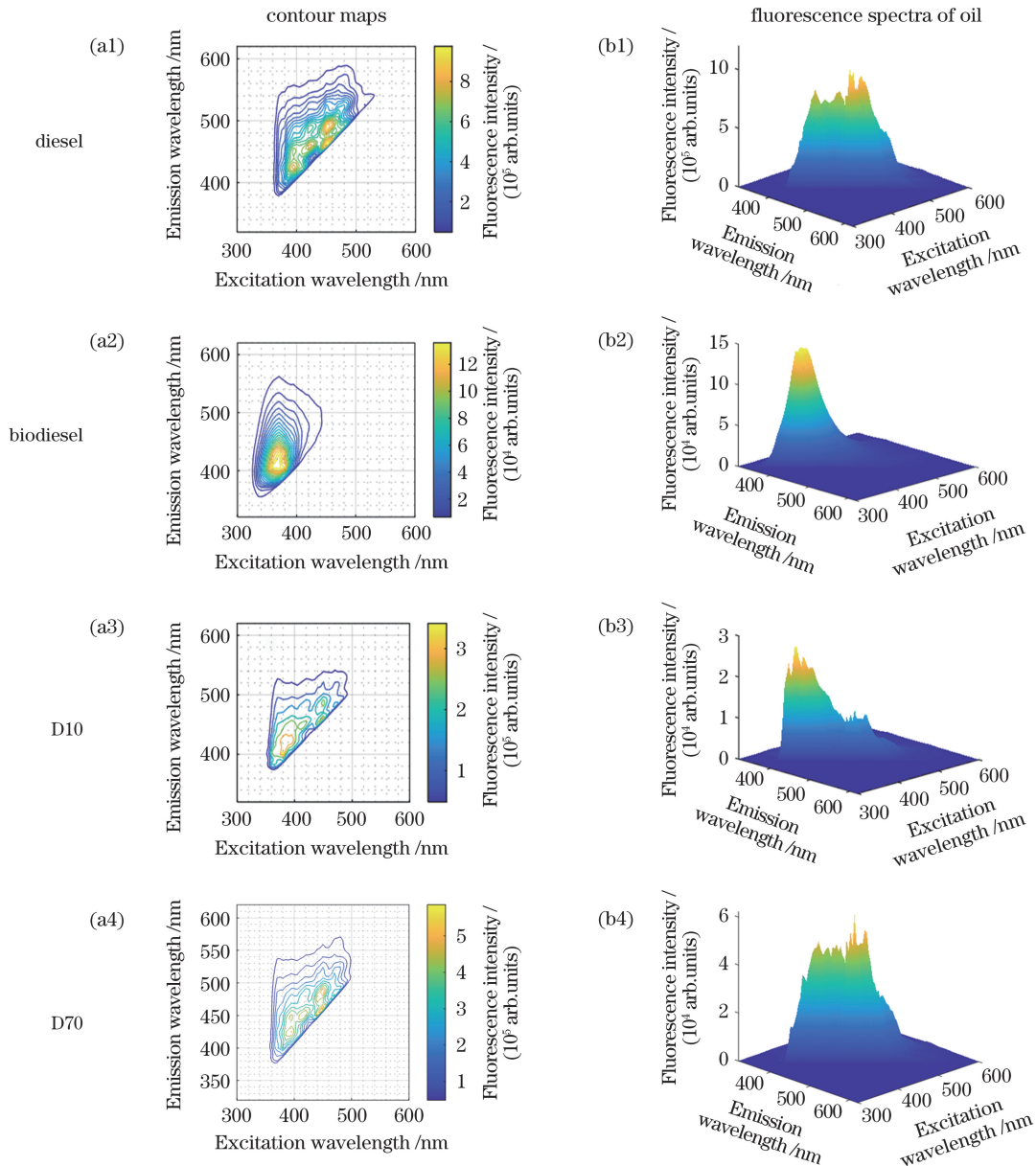


图 1 柴油、生物柴油以及柴油与生物柴油混合油液 D10 和 D70 的三维荧光等高线图和荧光光谱图

Fig. 1 Three-dimensional contour maps and fluorescence spectra of diesel and biodiesel, as well as diesel and biodiesel mixture samples D10 and D70

之间的对比可知柴油与生物柴油混合油液光谱的荧光强度与柴油占比呈非线性关系,随着柴油在混合油中占比的降低,柴油的光谱特征也相应减弱,因此混合油液最佳荧光峰的位置会发生偏移,这种偏移会对柴油占比的分析产生影响。

进一步分析,选取训练集中的混合油液,在柴油的最佳激发波长 450 nm 下,测量混合油液的发射光谱图,测量结果如图 2 所示。混合油液的最大荧光强度随柴油占比的变化如图 3 所示。结合图 2 和图 3 可知,在发射波长范围为 320~620 nm 的发射光谱中,主荧光峰有两个。为避免产生一级瑞利散射而采取发射波长滞后于激发波长 20 nm 的做法导致 450 nm/460 nm 处的荧光峰不完整,无法判断荧光峰是否出现偏移,而且峰的最大荧光强度与柴油占比不呈线性变

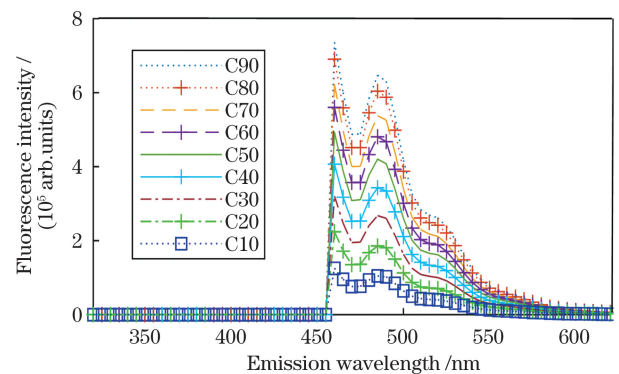


图 2 混合油液的发射光谱

Fig. 2 Emission spectra of oil mixtures

化关系;而在相同的激发波长下,475~500 nm 发射波长范围的荧光峰波形完整,最大荧光强度与 450 nm/

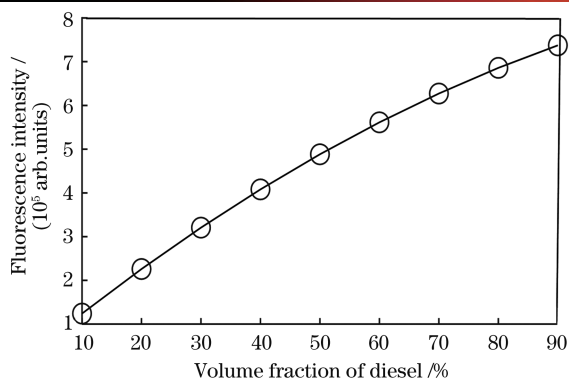


图 3 混合油液的最大荧光强度

Fig. 3 Maximum fluorescence intensity of oil mixtures

460 nm 处荧光峰的最大荧光强度相差较小,且在柴油占比较小时荧光峰的位置向左偏移。根据王书涛预测物质占比的方法<sup>[9-10]</sup>,本实验中选取的荧光强度较强的两荧光峰均包含较多荧光信息,但利用二者的荧光信息建立预测网络会造成过拟合现象,从而降低物质占比的预测精度。因此,需要选取一种方法拾取荧光信息较多的荧光峰。

#### 4.2 样本预处理

利用 2DPCA 对荧光光谱进行重构,拾取能代表样本性质的荧光峰建立预测网络。根据 2DPCA 提取的光谱的特征信息,可以得到对应的主成分贡献率。为保证重构的混合油荧光光谱能代表柴油占比,由图 4 所示的混合油液重构后的发射光谱和柴油发射光谱对比结果可以看出贡献率在 90% 以上即可重构出所需光谱信息。根据表 2 所示主成分的贡献率,选取前三个主成分 PC1、PC2、PC3 对荧光光谱进行重构。

表 2 主成分贡献率

Table 2 Contribution rate of principal components

Principal component	Contribution rate / %	Cumulative contribution rate / %
PC1	70.78	70.78
PC2	17.11	87.88
PC3	5.06	92.94
PC4	2.37	95.31
PC5	1.13	96.45
PC6	0.62	97.07
PC7	0.52	97.60
PC8	0.42	98.01
PC9	0.30	98.32
PC10	0.25	98.57

经过重构后的混合油液 D10 的三维等高线图和荧光光谱如图 5(a1)、(b1) 所示, D70 的三维等高线图和荧光光谱如图 5(a2)、(b2) 所示。2DPCA 利用由柴油、混合油液(D10~D90)、生物柴油组成的多个二维矩阵同时投影并选取主成分对样本进行重构,重构光

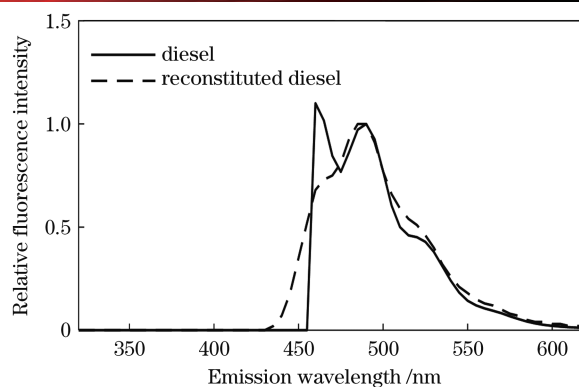


图 4 混合油液光谱经 2DPCA 重构后的发射光谱与柴油发射光谱的对比

Fig. 4 Comparison between reconstructed emission spectrum of oil mixture by 2DPCA and diesel emission spectrum

谱中包含了能代表柴油和生物柴油占比信息的荧光峰。由图 4 可知经过 2DPCA 重构的荧光光谱在保留混合油本身荧光特性的前提下,可使荧光的主要特征峰得以突出,荧光光谱的组成成分得以简化。

训练集中混合油液样本的发射光谱在柴油最佳激发波长 450 nm 处的变化情况如图 6 所示。为避免柴油占比对光谱的影响,对发射光谱利用 450 nm/490 nm 的荧光强度进行归一化。柴油发射光谱以及混合油液光谱经 2DPCA 重构后的发射光谱的对比如图 4 所示。由图 4 和图 6 可知重构后的混合油发射光谱得到了精简,拾取出的荧光峰在发射波长 475~500 nm 范围内与柴油本身荧光峰的拟合情况较好,位置和波形基本未发生改变,能够代表柴油的荧光信息,进而能够用来对柴油的占比进行预测。

#### 4.3 混合油液中柴油的定量分析

利用 2DPCA 重构训练集,将重构后的训练集中各混合油液样本在激发波长 450 nm、发射波长 475~500 nm 下的相对荧光强度组成的 9×6 矩阵作为输入,柴油在混合油液中的占比作为输出,构建预测网络,对柴油占比进行预测。

麻雀搜索算法的迭代次数随误差的变化曲线如图 7 所示。可见,麻雀搜索算法在第 5 次迭代时就找到了能使广义回归神经网络误差最小的光滑因子,收敛速度较快且收敛到最小值后保持不变。

利用最佳的光滑因子建立广义回归神经网络,对测试集的样本进行预测,得到柴油占比的预测值。表 3 是 2DPCA-SSA-GRNN 模型和 SSA-GRNN 模型预测结果的对比,其中 SSA-GRNN 模型选取训练集中未经 2DPCA 重构的样本中激发波长 450 nm 下荧光强度较高的 460~500 nm 范围内的发射光谱建立同样的广义回归神经网络进行预测。因为广义回归神经网络在预测时的权重具有随机性,因此预测时会有误差。本文预测值是麻雀搜索算法寻找到最佳光滑因子后进行 5 次预测并取平均值得到的。由表 3 可以看出,2DPCA-SSA-GRNN 模型对柴油占比的预测效果

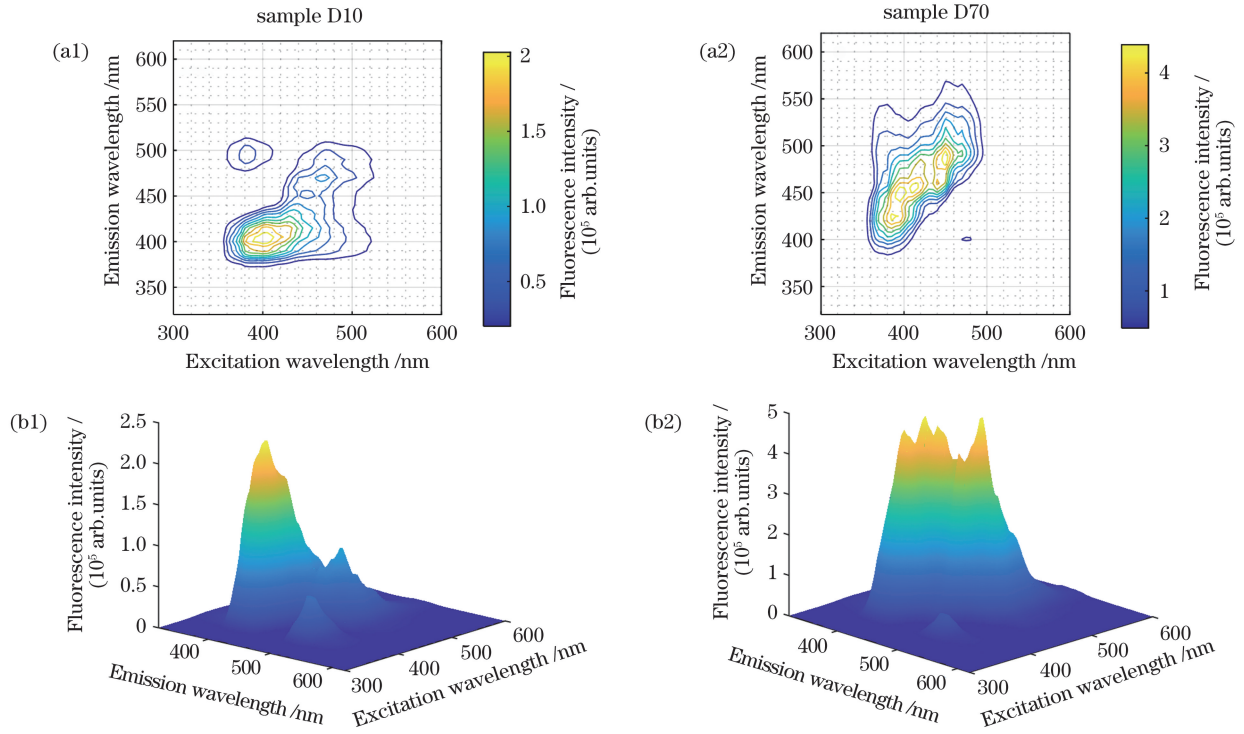


图 5 2DPCA 重构之后混合油液 D10、D70 的三维等高线图和荧光光谱。(a1)(a2)等高线图;(b1)(b2)三维荧光光谱  
Fig. 5 Three-dimensional contour maps and fluorescence spectra of oil mixture samples D10 and D70 after 2DPCA reconstruction. (a1)(a2) Three-dimensional contour maps; (b1)(b2) fluorescence spectra

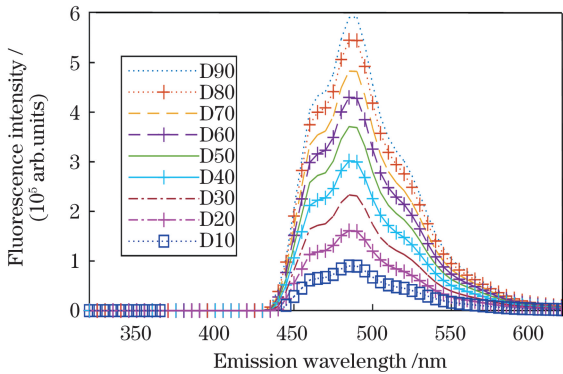


图 6 2DPCA 重构后混合油液的发射光谱  
Fig. 6 Emission spectra of oil mixture samples reconstructed by 2DPCA

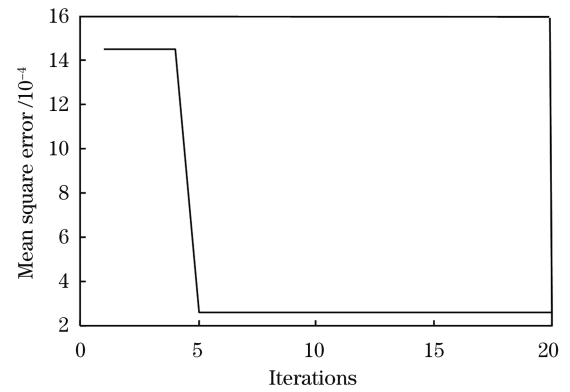


图 7 麻雀搜索算法的收敛曲线  
Fig. 7 Convergence curve of sparrow search algorithm

表 3 预测结果对比  
Table 3 Comparison of prediction results

Model	Predictive percentage /%	Recovery /%	Spread	RMSE /%	Average recovery /%
2DPCA-SSA-GRNN	85.02 (D85)	100.02 (D85)	0.44	0.90	98.39
	73.76 (D75)	98.34 (D75)			
	63.80 (D65)	98.16 (D65)			
	53.37 (D55)	97.03 (D55)			
SSA-GRNN	85.50 (D85)	98.24 (D85)	0.48	1.87	97.15
	75.96 (D75)	101.28 (D75)			
	61.69 (D65)	94.90 (D65)			
	51.79 (D55)	94.17 (D55)			

良好, D85、D75、D65、D55 油液样本的占比分别为 85.02%、73.76%、63.80%、53.37%, 平均回收率为 98.39% (较 SSA-GRNN 模型预测的平均回收率提高了 1.24 个百分点), 均方根误差 RMSE 为 0.90% (约

为 SSA-GRNN 模型预测 RMSE 的 1/2)。由上述两种模型的对比可以看出,采用 2DPCA 对荧光光谱进行重构有助于提高荧光峰发生偏移时的预测准确率。

## 5 结 论

利用峰值拾取结合智能算法优化神经网络,可以对物质的占比进行定量分析,但当荧光光谱比较复杂时,主观的峰值拾取会导致准确率降低。本团队利用 2DPCA 对混合油的荧光光谱进行重构,重构后的荧光光谱既保留了光谱原始的荧光特性,又简化光谱中的冗余信息,提高了混合油液中柴油的定量分析准确率,为物质的定量分析提供了新思路。下一步拟将 2DPCA 与其他光谱分析方法相结合,进一步提高光谱重叠严重情况下荧光信息的提取准确度,实现复杂物质的定量与定性分析。

## 参 考 文 献

- [1] 张志清. 燃生物柴油船用发动机高效燃烧与低污染排放研究[D]. 长沙: 湖南大学, 2019.  
Zhang Z Q. Research on high efficiency combustion and low emission of marine engine fueled with biodiesel[D]. Changsha: Hunan University, 2019.
- [2] 张雁玲, 孟凡飞, 王家兴, 等. 国内外生物柴油发展现状[J]. 现代化工, 2019, 39(10): 9-14.  
Zhang Y L, Meng F F, Wang J X, et al. Global and China's development status of biodiesel industry[J]. Modern Chemical Industry, 2019, 39(10): 9-14.
- [3] Mazivila S J. Trends of non-destructive analytical methods for identification of biodiesel feedstock in diesel-biodiesel blend according to European Commission Directive 2012/0288/EC and detecting diesel-biodiesel blend adulteration: a brief review[J]. Talanta, 2018, 180: 239-247.
- [4] 赵兴涛, 车先阁, 王书涛, 等. 荧光光谱法结合 BLLS/RBL 用于快速检测水中喹诺酮类抗生素[J]. 光学学报, 2020, 40(9): 0930001.  
Zhao X T, Che X G, Wang S T, et al. Rapid detection of quinolones in water based on fluorescence spectrometry and BLLS/RBL[J]. Acta Optica Sinica, 2020, 40(9): 0930001.
- [5] 黄尧, 赵南京, 孟德硕, 等. 非平滑非负矩阵分解解析土壤多环芳烃三维荧光光谱[J]. 中国激光, 2020, 47(10): 1011002.  
Huang Y, Zhao N J, Meng D S, et al. Analysis on three-

- dimensional fluorescence spectra of PAHs in soil using non-smooth non-negative matrix factorization[J]. Chinese Journal of Lasers, 2020, 47(10): 1011002.
- [6] 王书涛, 展书杰, 刘诗瑜, 等. 三维荧光光谱结合 ICSO-SVM 对性激素的分类鉴别[J]. 光学学报, 2021, 41(10): 1030004.  
Wang S T, Zhan S J, Liu S Y, et al. Classification and identification of sex hormones by three-dimensional fluorescence spectroscopy combined with ICSO-SVM[J]. Acta Optica Sinica, 2021, 41(10): 1030004.
- [7] 程钊, 赵南京, 殷高方, 等. 基于 SWATLTD 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 SWATLTD [J]. Acta Optica Sinica, 2021, 41(14): 1430001.
- [8] 申金媛, 韩应哲, 常胜江, 等. 非线性荧光光谱的神经网络分析及其应用[J]. 光学学报, 2004, 24(7): 1000-1003.  
Shen J Y, Han Y Z, Chang S J, et al. Neural network analysis and application of nonlinear fluorescence spectra[J]. Acta Optica Sinica, 2004, 24(7): 1000-1003.
- [9] 王书涛, 陈东营, 魏蒙, 等. 荧光光谱法和 PSO-BP 神经网络在山梨酸钾浓度检测中的应用[J]. 中国激光, 2015, 42(5): 0515004.  
Wang S T, Chen D Y, Wei M, et al. Application of fluorescence spectroscopy and PSO-BP neural network in the detection of potassium sorbate concentration[J]. Chinese Journal of Lasers, 2015, 42(5): 0515004.
- [10] 王书涛, 陈东营, 王兴龙, 等. 荧光分析法和 ABC-BP 神经网络相结合的多环芳烃的检测[J]. 中国激光, 2015, 42(11): 1115001.  
Wang S T, Chen D Y, Wang X L, et al. Detection of polycyclic aromatic hydrocarbons combining fluorescence analysis with ABC-BP neural network[J]. Chinese Journal of Lasers, 2015, 42(11): 1115001.
- [11] Yang J, Zhang D, Frangi A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(1): 131-137.
- [12] 吴希军, 崔耀耀, 潘钊, 等. 三维荧光光谱结合 Zernike 图像矩快速鉴别掺伪芝麻油[J]. 光谱学与光谱分析, 2018, 38(8): 2456-2461.  
Wu X J, Cui Y Y, Pan Z, et al. 3D fluorescence spectra combined with Zernike image moments for rapid identification of doping sesame oil [J]. Spectroscopy and Spectral Analysis, 2018, 38(8): 2456-2461.
- [13] Xue J K, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm [J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.

# Detection of Diesel Proportion Using Three-Dimensional Fluorescence Spectrum and 2DPCA-SSA-GRNN

Chen Xiaoyu<sup>1</sup>, Du Yaxin<sup>1</sup>, Liu Yaru<sup>2</sup>, Kong Deming<sup>2\*</sup>

<sup>1</sup>School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China;

<sup>2</sup>School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China

## Abstract

**Objective** Recently, overexploitation of oil resources has become a major problem, attracting extensive attention. Thus, alternative new fuels are needed to alleviate the problem. Biodiesel is usually mixed with diesel as an alternative to diesel because of its similar properties and excellent environmental benefits, thereby reducing the actual consumption of diesel. However, the proportion of diesel in mixed oil containing diesel and biodiesel greatly affects the power transformed. A low diesel proportion will cause problems, such as poor oil atomization, insufficient oil-gas mixing,

insufficient engine pressure, and increased nitrogen dioxide emission. Although several countries have different specifications for the proportion of diesel in mixed fuel, the proportion of diesel should be more than 50%. Therefore, the rapid and effective identification of the proportion of diesel in the mixed oil of diesel and biodiesel is greatly significant for the qualitative detection of imported diesel.

**Methods** In this paper, the peak pick-up method was used to detect the proportion of diesel in the mixed oil. Since there were many hydrocarbons in the mixed oil, the fluorescence spectrum of the sample was complex. Thus, selecting the appropriate fluorescence peak was difficult. The sample was projected directly on the two-dimensional plane using two-dimensional principal component analysis (2DPCA), and the first three principal components were selected to reconstruct the fluorescence spectrum. Observe that the reconstructed fluorescence spectrum retained the fluorescence peak with more principal component information. Using this peak intensity to predict the proportion of the diesel oil will greatly improve the prediction accuracy. However, no linear relationship between the retained fluorescence peak intensity and the proportion of diesel oil exists, thus, introducing a neural network for further analysis is necessary. Furthermore, we selected a generalized regression neural network (GRNN) to predict the proportion of diesel oil because of its advantages in dealing with nonlinear relationships. Additionally, three sparrows of the small slope approximation (SSA) method were used to search the parameters of the GRNN-spread, and the GRNN was established using the best spread to accurately detect the proportion of diesel in the mixed oil.

**Results and Discussions** For the oil mixture sample of diesel and biodiesel (Table 1), we divided the sample into test and training sets, respectively. It can be found from Fig. 1 that the fluorescence spectrum of the sample was complex, with more fluorescence peaks and less difference in fluorescence intensity; thus, accurately identifying fluorescence peaks containing more principal components was difficult. In the Fig. 1, the best fluorescence peak of diesel oil was at 450/480 nm and that of biodiesel was at 370/410 nm. With the decrease in diesel proportion, the fluorescence intensity at the best fluorescence peak position of diesel failed to satisfy the linear relationship with the proportion of diesel, as shown in Fig. 3. Furthermore, the fluorescence spectrum of biodiesel greatly affected that of diesel. It was impossible to simply extract the strongest fluorescence intensity and establish a neural network for prediction. The three-dimensional fluorescence spectrum of the sample was reconstructed using 2DPCA, as shown in Fig. 5. Since the remaining fluorescence peaks of the reconstructed fluorescence spectrum contain more principal components, selecting the spectral information with strong fluorescence intensity for prediction reduced the error. According to Fig. 4, the reconstructed emission spectrum of the sample matched the original diesel sample. Therefore, the strong part of the reconstructed fluorescence spectrum ( $9 \times 6$ ) was selected as the training set. The GRNN optimized using SSA was established to predict the test set and the prediction results are shown in Table 3. The results revealed that the prediction effect of the 2DPCA-SSA-GRNN model is good, with an average recovery of 98.39% and a mean square error of 0.90%, which is about 1/2 the mean square error predicted by SSA-GRNN alone.

**Conclusions** The intelligent algorithm was used to optimize the neural network, which realized the quantitative analysis of substances. However, its accuracy was low considering the complex situation of the fluorescence spectrum. In this paper, we proposed 2DPCA to reconstruct the fluorescence spectrum of the sample. Our method retains the original fluorescence characteristics of the spectrum, simplifies the redundant information in the spectrum, improves the quantitative analysis results of diesel in mixed oil, and provides a new idea for the quantitative analysis of substances. With the progress in science and technology, more additives can be used to optimize the performance of substances, even with more complex spectra. 2DPCA can screen out the main fluorescence information in the spectrum and improve the peak picking method to analyze substances. In future work, 2DPCA can be combined with other spectral analysis methods to further improve the accuracy of fluorescence information extraction for serious spectral overlap and realize the quantitative and qualitative analysis of complex substances.

**Key words** spectroscopy; three-dimensional fluorescence spectroscopy; two-dimensional principal component analysis; diesel; generalized regression neural network; sparrow search algorithm