

基于连续投影算法的食用油激光诱导 荧光光谱特征波长筛选

周孟然,余道洋,胡锋,卞凯,来文豪,孙磊,闫鹏程

(安徽理工大学 电气与信息工程学院,安徽 淮南 232000)

摘要:针对目前质量监督领域中难以快速准确地识别食用油种类的问题,提出了一种激光诱导荧光技术结合连续投影算法的食用油光谱识别方法.实验搭建激光诱导荧光系统采集了 5 种食用植物油共计 500 组荧光光谱数据.首先,通过实验对比选择 Savitzky-Golay 卷积平滑算法对荧光光谱进行预处理;然后使用连续投影算法筛选特征波长;最后,将筛选出的特征波长送入建立的神经网络模型中进行训练和测试,并通过不同时间采集的油样组成独立验证集进行再次验证.结果表明:通过特征波长筛选,从全光谱 2 048 个波长中优选出 11 个,减少了冗余信息,波长数缩减为原来的 0.54%.并且在神经网络模型中的训练样本准确率和测试样本准确率分别达到 100%和 95%,效果好于径向基函数神经网络和 BP(Back Propagation)神经网络.在独立验证集的预测准确率也达到了 91%.因此,将连续投影算法用于食用油激光诱导荧光光谱特征波长筛选并结合神经网络模型可以实现食用油的快速准确分类且具备通用性,并为进一步设计专用的在线式食用油种类识别仪提供了理论依据.

关键词:激光诱导荧光;特征波长;连续投影算法;神经网络;食用油

中图分类号:O657.3

文献标志码:A

食用油在人们的日常饮食中不可或缺,其品质的优劣直接关系到人们的身体健康,不同种类食用油之间的成分、营养价值、市场价格都不尽相同,但外观上很难辨识,因此对食用油真实品类进行迅速精准的识别具有很大实际意义.常规的食用油检测主要是气相色谱法^[1],高效液相色谱法^[2]等化学检测方法.化学检测的精度较高但比较耗时,需要化学前处理,不适合快速检测^[3].

近年来,激光诱导荧光(Laser Induced Fluorescence, LIF)技术快速发展,广泛地应用在了诸多领域之中.LIF 技术具备良好的灵敏性和快速性,为食用油的快速准确识别指明了新的方向.文献[4]利用 LIF 技术获取了突水水源的光谱数据,在进行预处理后通过主成分分析进行降维,最后通过线性判别分析实现突水水源的快速准确识别.文献[5]通过基于液芯光纤的 LIF 装置获取食用油样本,并利用主成分分析和偏最小二乘法判别分析建立食用油种类鉴别模型,识别效果良好.文献[6]将 LIF 技术与主成分分析法相结合对掺假不同浓度煎炸油的植物油成功进行了分类,并通过偏最小二乘模型实现了掺假浓度预测.文献[7]通过 LIF 技术采集了不同品牌和黏度的机油样本,采用主成分分析对原始数据降维,结合聚类分析法可以方便快捷地对不同机油进行识别.然而目前将 LIF 技术用于食用油以及其他对象的识别研究中,大都采用主成分分析法把原始光谱数据线性组合成新的变量,选择部分新变量代替原始数据,实现降维.但这种方式完全破坏了荧光光谱曲线,而且主成分分析得到的若干主成分更加难以解释,不具备实际的物理意义.

连续投影算法(Successive Projections Algorithm, SPA)是一种常用的特征波长筛选算法,可以从光谱数据中筛选出冗余信息含量最少的特征组,大大降低建模所需要的特征数,使建模更加高效.在近红外光谱、高光谱等多种光谱的多元定性和定量分析中具有广泛的应用.文献[8]提出将 SPA 用于近红外光谱特征提取来分析小麦和啤酒提取物.文献[9]利用 SPA 结合 KNN 和 LDA 对查耳酮衍生物的抗反应活性进行分类.

收稿日期:2020-07-11;修回日期:2020-08-09.

基金项目:国家重点研发计划(2018YFC0604503);安徽省自然科学基金(1808085QE157).

作者简介:周孟然(1965—),男,安徽淮南人,安徽理工大学教授,博士,博士生导师,研究方向为光电信息处理、模式识别等.

通信作者:余道洋(1995—),男,安徽淮南人,安徽理工大学硕士研究生, E-mail:dyuuah@163.com.

文献[10]提出了竞争性自适应重加权抽样连续投影算法用于分析烟叶中尼古丁的近红外反射数据和农药制剂的近红外透射数据.文献[11]使用 SPA 算法对预处理后的高光谱数据进行特征波长选择,然后通过支持向量机建立小麦赤霉病籽粒识别模型,实现了染病籽粒可视化识别.

本文在对原始光谱进行降噪预处理的基础上,采用 SPA 提取食用油荧光光谱的特征波长,去除冗余信息降低光谱信息维度,并作为概率神经网络的输入用于建立食用油光谱识别模型,验证了 SPA 筛选食用油激光诱导荧光光谱特征波长的可行性.

1 材料与方法

1.1 材料与仪器

选择市面上在售的常见食用油为实验对象,采购自各大超市,种类包括菜籽油、大豆油、花生油、葵花籽油、玉米油,共 5 种食用植物油,其主要产品数据如表 1 所示.

实验所用仪器包括 USB2000+ 微型光纤光谱仪,由美国 Oceanoptics 公司生产,光谱检测范围为 340.472~1 021.086 nm,分辨率为 0.5 nm (FWHM),积分参数为 1 s/(1 000 nm);蓝紫光半导体激光器,北京华源拓达激光技术有限公司,设定入射激光波长 405 nm,入射激光的功率设置 150 mW;荧光探头,广东科思凯公司生产的 FPB-405-V3 可浸入式激光激发荧光探头.

表 1 实验材料

Tab. 1 Experimental Materials

标签	名称	品牌	生产厂商
1	菜籽油	鸵鸟	南通中海食用油有限公司
2	大豆油	福临门	中粮东海粮油工业有限公司
3	花生油	鲁花	襄阳鲁花浓香花生油有限公司
4	葵花籽油	鲁花	内蒙古葵花仁油有限公司
5	玉米油	长寿花	山东三星玉米产业科技有限公司

1.2 数据采集与划分

食用油的荧光光谱数据采集实验系统如图 1 所示.实验时,由激光器发出高能激光射入实验样品中,再通过探头接收荧光发送到光谱仪中,获得其光谱数据.2019 年 11 月在实验室环境中将每种食用油样品等体积放入相同的玻璃瓶中.为了避免背景光以及其他人为因素对实验结果的影响,所有样品均未经过任何化学处理,并且荧光探头和存储食用油样本的玻璃瓶均在相同的温湿度的暗室中实验.其中每类食用油样品各采集 100 组光谱数据,然后采用顺序划分法对 5 类食用油的 500 组样品进行划分.参考文献[4-7]的样本划分方法,将划分比例设置为 4:1,每类油样取 80 组作为训练集,剩余 20 组作为测试集,共计 400 组训练集样品,100 组测试集样品.

1.3 LIF 技术

荧光是物体在吸收光线受到激发后,在去除激发时发出的一种辐射^[12].高能激光照射物质诱使其发出荧光即激光诱导荧光(LIF).而光谱技术大致上可以分成接收光谱和发射光谱,该实验使用的 LIF 技术属于发射光谱,即在激光光源的强度以及波长不变的情况下利用仪器获得的荧光强度跟随荧光接收波长的光谱图,简称荧光光谱.通过 LIF 技术获得的荧光光谱可以体现实验样品的一些结构特性,因此能够很好地分析样品的特异性.而液体的荧光强度与液体吸收的光强度、浓度等皆有关系^[13].由于每种食用油之间脂肪酸的成分与密度均不同,将 LIF 技术用于食用油油样识别,能充分利用不同油样的差异性,提高对不同油类的识别率.

1.4 特征波长选择

连续投影算法(SPA)是能够很好地消除波长数值间共线关系问题的前向循环波长筛选方法^[14],利用向量投影来优选出冗余度低、共线性小又能反映样本光谱关键信息的有效特征波段,利用降低输入特征的数目来提高建模速度并减少复杂程度,最大程度地减少光谱信息重叠^[15].它是一种前向筛选方法,即从一个波长为起点,每次迭代时合并一个新的波长,一直到指定数目的 n 个波长为止.该方法能够最大限度地降低测试样品的光谱数据维度,减少后期建模的计算量.

假设样本值 K 和波点值 M 构成的光谱矩阵 $X_{K \times M}$, $N(N < K - 1)$ 是所要提取的特征波点,SPA 的步骤

如下:(1) 在首次迭代($n=1$)之前,从 $X_{K \times M}$ 中任意选择一个列向量 x_j 设为 $x_{k(0)}$, $k(0)=j, j \in 1, 2, \dots, m$;
 (2) 将没有被选入的列向量位置的集合记为 $S, S = \{j, 1 \leq j \leq m, j \notin \{k(0)=j, j=1, 2, \dots, m\}\}$;
 (3) 计算剩余的列向量 $x_j (j \in S)$ 与当前所选向量 $x_{k(n-1)}$ 的投影 $P_{x_j} = x_j - (x_j^T x_{k(n-1)}) x_{k(n-1)} (x_{k(n-1)}^T x_{k(n-1)})^{-1}, j \in S$;
 (4) 选择投影最大值所对应波点的序号 $k(n) = \arg(\max \|P_{x_j}\|), j \in S$;
 (5) 令 $x_j = P_{x_j}, j \in S$;
 (6) 令 $n = n + 1$, 如果 $x_{k(n)} n < N$, 则返回第(2)步循环计算.

最终筛选的波长变量组合为 $k(n), n=0, 1, \dots, N-1$. 对于每一个初始化 $k(0)$, 循环一次后利用多元线性回归 (multiple linear regression, MLR) 进行交叉验证分析, 最小均方根误差 (root mean square error, RMSE) 值所对应的变量组合 $k(n)$ 即为最终选择的结果.

1.5 概率神经网络模型

概率神经网络 (Probabilistic Neural Networks, PNN) 是一种融合了概率密度估计和贝叶斯思想的径向基神经网络^[16]. 具有训练时间短、精准性和容错率高等优点. 在实际处理分类问题的过程中, PNN 的优势体现在用线性学习算法来完成非线性学习算法所要完成的工作, 同时又能保证非线性算法的高精度等特性^[17].

PNN 网络结构^[18]如图 2 所示. 输入层主要是将特征向量输入模型中; 第二层模式层是径向基层, 以连接权值来与上一层连接. 该层每个模式单元的输出为 $f(X, W_i) = \exp[-\frac{(X - W_i)^T (X - W_i)}{2\delta^2}]$, 式中, W_i 是连接第一层和第二层的权值; δ 是平滑因子, 对分类应用具有重要意义. 第三层是求和层, 负责统计同一类别的概率, 然后获得估计函数. 最后一层是竞争神经元, 对上一层的结果进行归一化并计算每个样品属于每个类别的概率, 最大概率所对应的那一类就是需要识别的样本类别.

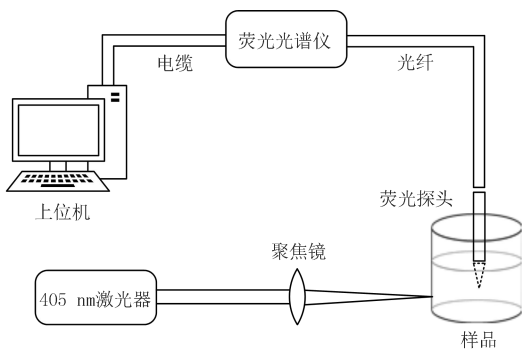


图1 LIF光谱系统示意图

Fig.1 Schematic diagram of LIF spectral system

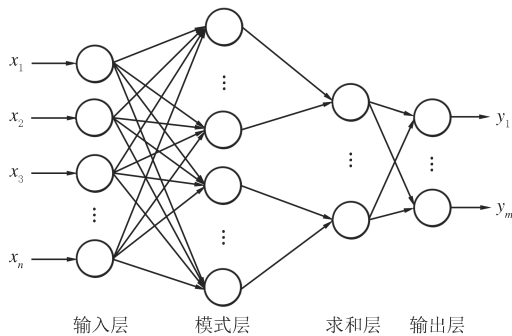


图2 概率神经网络结构图

Fig.2 Probabilistic neural network structure diagram

2 结果与分析

2.1 原始荧光光谱

通过上述实验设备采集的所有样本的荧光光谱曲线如图 3 所示. 花生油和葵花籽油有两个波峰, 其他 3 个只有一个波峰. 但 5 种油类之间波峰的位置和对应的荧光强度都差距不大, 仅通过观察无法直观的进行区分. 因此需要对实验得到的荧光光谱进行进一步处理, 通过算法建模实现油样的分类识别.

2.2 光谱预处理方法选择

为了消除光谱中噪声等因素的干扰, 采用 S-G 卷积平滑、移动平均平滑、标准正态变量变换、多元散射校正法等 4 种荧光光谱分析中较为常用的预处理算法进行测试. 4 种方法分别对原始光谱数据进行预处理, 其结果如图 4 所示.

从图 4 可以看出, 采用标准正态变量变换和多元散射校正处理后的光谱在波峰位置处出现了混乱, 另外两种算法的预处理效果良好, 为进一步确定不同预处理方法的效果差异, 选用信噪比 (Signal-Noise Ratio,

SNR)、相关系数(Correlation coefficient, RC)和均方差(Mean-Square Error, MSE)作为预处理结果的评价指标, SNR 和 RC 越高、MSE 越低代表该预处理方法的降噪效果越理想,其结果如表 2 所示.

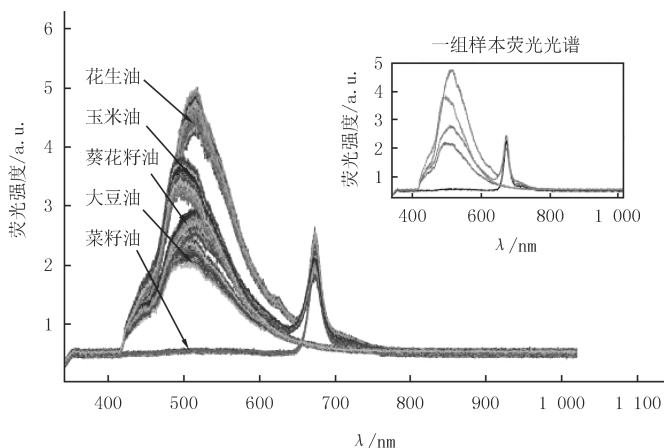


图3 原始荧光光谱图

Fig.3 Original fluorescence spectrum

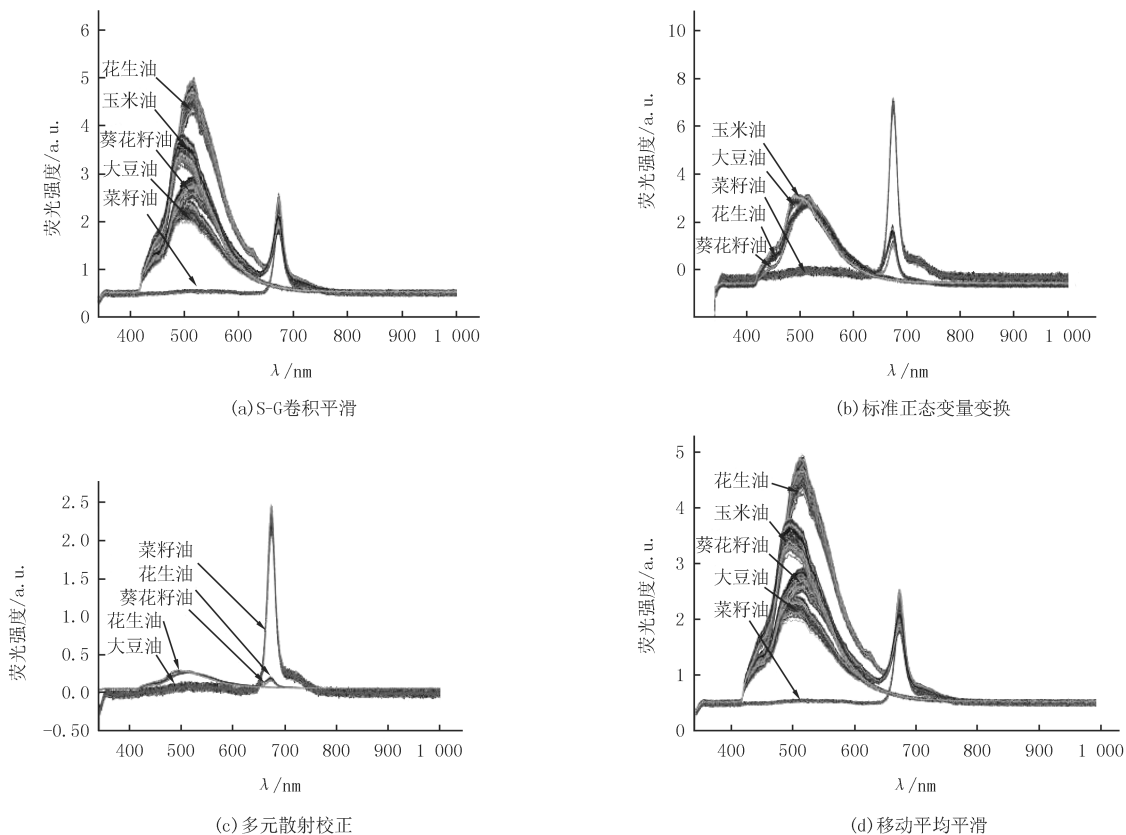


图4 不同预处理后的荧光光谱图

Fig.4 Fluorescence spectra after different pretreatments

从表 2 中可以看出,标准正态变量变换和多元散射校正处理后的 SNR 非常低,而 MSE 非常高,与光谱图看出的结果一致.S-G 卷积平滑处理后的 SNR 和 RC 最高、MSE 最低,说明其预处理效果最好,因此最终选择 S-G 卷积平滑法对样本数据进行预处理.

2.3 SPA 波长筛选

油样的原始荧光光谱数据量大,运算时间长,并且有很大的共线关系.原始光谱经过 S-G 卷积平滑预处理

理后使用 SPA 进行特征波长筛选,设置特征波段筛选的数目范围为 1~50.然后利用筛选的特征波点数据建立多元线性回归校正模型,通过均方根误差(Root Mean Square Error, RMSE)值决定最佳建模波长数目.如图 5 所示是提取不同个数的特征波长点数时, RMSE 值随之改变的情况.可以观察到,当提取 11 个波长波点时模型的 RMSE 为最小值 0.463 25.

表 2 不同光谱预处理效果对比

Tab. 2 Comparison of different spectrum pretreatment effects

评价指标	S-G 卷积平滑法	标准正态变量变换	多元散射校正	移动平均平滑
SNR	40.007 5	0.006 5	-13.310 0	34.653 7
MSE	6.446 5	476.931 8	303.692 3	8.191 6
RC	0.999 8	0.874 9	0.463 6	0.999 7

通过上述分析选取 11 个波长作为最终特征波长,其在光谱图中的位置如图 6 所示,用方框标出.全波段光谱有 2 048 个光谱波长数据点,而通过 SPA 算法筛选出的波长数只有全波段的 0.54%.这 11 个波长点所对应的特征波长/nm 分别为 439.841、660.858、1004.967、901.962、783.373、927.6、1017.99、929.095、768.452、994.403、844.21.

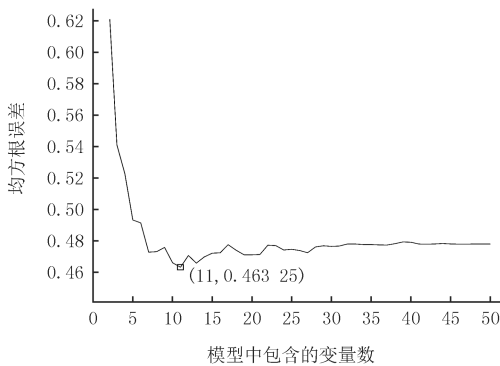


图5 不同波长变量下的均方根误差

Fig. 5 RMSE at different wavelength variables

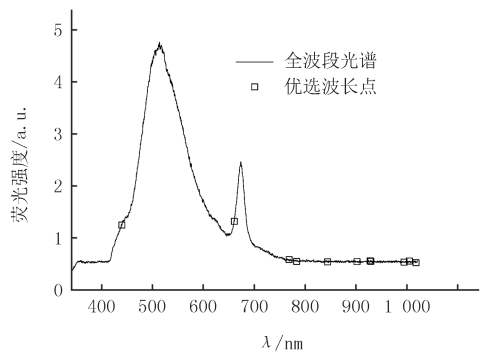


图6 SPA筛选的最佳特征波长

Fig. 6 The best characteristic wavelength for SPA screening

2.4 基于 SPA 特征波长的 PNN 模型

将 SPA 算法筛选出的 11 个特征波长数据作为输入,建立 SPA-PNN 模型进行食用油种类的识别和验证.训练样本数据共有 400 组,每组有 11 个特征变量,油样种类共 5 种,因此 PNN 模型的输入层、模式层、求和层、输出层的神经节点各为 11、400、5、5,其网络拓扑结构为 $11 \times 400 \times 5 \times 5$.

平滑因子 spread 的选取的对网络的性能起着重要作用,必须使得每个节点对输入变量涵盖的区域都能产生影响. spread 取得越大,模型得到的结果就越平滑,但是过大的 spread 值会使模型的运算变得复杂.根据上述分析,测试 spread 的取值区间为 $[0.1, 3]$,间隔为 0.1,共计 30 个值.试验结果如图 7 所示.

能够看到,在 $[0.1, 1)$ 的范围内,测试样品分类正确率基本不随 spread 值变化;在 $[1, 2.2)$ 的范围内,准确率迅速升高,说明 PNN 网络对该区域内的平滑因子变化比较敏感;当 $\text{spread} \geq 2.2$,样本分类正确率不在变化.综上所述,设置概率神经网络的平滑因子 $\text{spread} = 2.2$.根据以上分析,将训练集数据输入 PNN 网络中开始建模分析.

模型训练完成后,将训练样本再次回代到网络模型中,测得 PNN 网络对训练样本数据的识别率达到 100%,均未出现误差,输出类别标签与期望值一致.然后利用测试样本对模型进行验证,得到的结果如图 8 所示,有 5 个样本分类错误.从图 8 可以看出模型对第 1 类、第 2 类、第 5 类的分类效果与期望值一致,类别 3 和类别 5 共有 5 组数据未达到预期效果.通过反复试验得出该模型对测试集数据的识别正确率为 95%,可以实现很好地识别效果.

为了验证 PNN 网络建模的可行性,现将该网络的预测结果与径向基函数神经网络(Radial Basis Function, RBF)建模和 BP(Back Propagation)神经网络建模的两组预测结果对比,具体数据如表 3 所示.

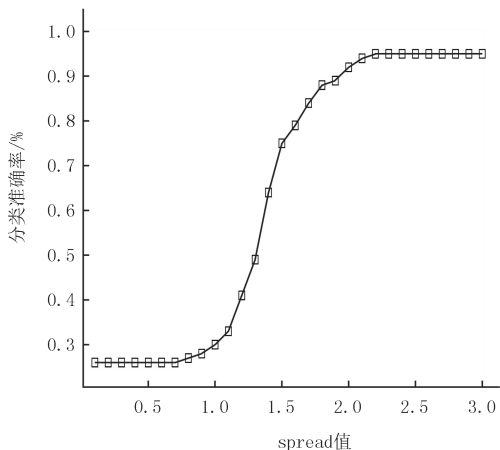


图7 分类准确率随spread变化曲线

Fig.7 Classification accuracy rate curve with spread

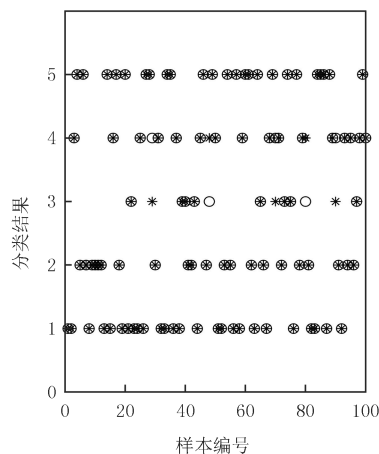


图8 测试样本预测结果

Fig.8 Test sample prediction results

从对比结果可知,在训练样本预测准确率上,PNN 网络建模的预测效果比 BP 建模效果更好,和 RBF 建模效果相同都达到了 100%。在测试样本准确率上比其他两种模型都稍好,并且在运行时间上可以看出,相比其他两种模型,PNN 模型的收敛速度更快,在 1 s 以内就可以完成运算。通过以上实验可以证明 SPA-PNN 食用油分类预测模型是合理可靠的。

表 3 不同模型准确率对比结果

Tab. 3 Comparison results of accuracy of different models

建模方法	训练样本准确率/%	测试样本准确率/%	运行时间/s
BP	91.25	88	2.495
RBF	100	93	1.736
PNN	100	95	0.907

同时,为了验证构建的 SPA-PNN 油样识别模型的通用性,另外选取 2020 年 1 月在大致相同的实验室环境下采集的 5 种同品类的食用油样本,每组各采集 20 个,共 100 组荧光光谱数据用作独立验证集。采用 S-G 卷积平滑对光谱预处理,然后将 SPA 筛选的 11 个特征波长信息输入到已经建立的模型中,识别结果如表 4 所示。在 100 组样本中,类别 3 有 5 组错误的预测为类别 4,类别 4 有 3 组错误的预测为类别 3,类别 5 有 1 组错误的预测为类别 1,其余

表 4 独立验证集预测结果

Tab. 4 Independent validation set prediction results

真实类别	预测类别				
	类别 1	类别 2	类别 3	类别 4	类别 5
类别 1	20	0	0	0	0
类别 2	0	20	0	0	0
类别 3	0	0	15	5	0
类别 4	0	0	3	17	0
类别 5	1	0	0	0	19

均识别正确,预测准确率为 91%。说明本文建立的识别模型具有通用性,同时也说明采用 SPA 筛选的特征波长数据能够充分代表原光谱的有效信息。

3 结 论

通过分析食用油的激光诱导荧光光谱数据信息,提出了一种更具策略的食用油光谱处理方法,并结合 PNN 分类模型,可以实现食用油种类的快速、准确识别。1)使用 S-G 卷积平滑进行光谱预处理,有效地消除了原始光谱中的部分随机噪声。2)在此基础上使用 SPA 算法进行波段筛选,从全波段 2 048 个波长范围中优选出 11 个特征波长点,大大减少了光谱的冗余信息,提高后续模型分析的速度和精度。3)通过构建 PNN 分

类识别模型进行验证,训练集和测试集识别准确率分别为100%和95%,独立验证集预测准确率为91%,效果良好.实验验证了将SPA用于食用油荧光光谱筛选是完全可行的,且保留了实际的物理意义,便于解释,并且采用SG-SPA-PNN分类识别模型对进一步设计专用的在线式食用油检测仪具有重要的意义.

参 考 文 献

- [1] 范桂芳,李佩佩,齐立松,等.液液萃取-气相色谱法测定发酵液中的有机酸与乙醇[J].农业工程学报,2018,34(23):227-231.
FAN G F,LI P P,QI L S,et al.Determination of organic acids and ethanol in fermentation broth by liquid-liquid extraction and gas chromatography[J].Transactions of the Chinese Society of Agricultural Engineering,2018,34(23):227-231.
- [2] KONG Q,BAI X Z,LIN H J,et al.Improved Determination of salicylaldehyde in Water Samples by Liquid-liquid Extraction Followed by High Performance Liquid Chromatographic Analysis[J].Journal of Central South University,2018,25(4):701-708.
- [3] PENG X T,LI Y N,XIA H,et al.Rapid and Sensitive Detection of Fipronil and Its Metabolites in Edible Oils by Solid-phase Extraction Based on Humic Acid Bonded Silica Combined with Gas Chromatography with Electron Capture Detection[J].Journal of Separation Science,2016,39(11):2196-2203.
- [4] 闫鹏程,尚松行,周孟然,等.基于激光诱导荧光技术的煤矿水源识别研究[J].光谱学与光谱分析,2020,40(7):2176-2181.
YAN P C,SHANG S H,ZHOU M R,et al.Research on identification of Coal Mine Water Source Identification Based on Laser Induced Fluorescence Technology[J].Spectroscopy and Spectral Analysis,2020,40(7):2176-2181.
- [5] 范苑,吴瑞梅,艾施荣,等.基于液芯光纤的激光诱导荧光食用油种类鉴别研究[J].光谱学与光谱分析,2016,36(10):3202-3206.
FAN Y,WU R,AI S R,et al.Identification study of edible oil species with laser induced fluorescence technology based on liquid core optical fiber[J].Spectroscopy and Spectral Analysis,2016,36(10):3202-3206.
- [6] 朱泉水,郝仕国,罗宁宁,等.基于激光诱导荧光的植物油掺假检测与量化分析[J].中国激光,2019,46(12):1211002.
ZHU Q S,HAO S G,LUO N N,et al.Detection and Quantitative Analysis of Vegetable Oil Adulteration Based on Laser-induced Fluorescence[J].Chinese Journal of Lasers,2019,46(12):1211002.
- [7] 刘晓华,陈思颖,张寅超,等.基于激光诱导荧光的常见机油快速识别方法[J].光谱学与光谱分析,2014,34(8):2148-2151.
LIU X H,CHEN S Y,ZHANG Y C,et al.Common Motor Oil Rapid Identification Method Based on Laser-induced Fluorescence[J].Spectroscopy and Spectral Analysis,2014,34(8):2148-2151.
- [8] DE ARAÚJO GOMES A,GALVÃO R K H,DE ARAÚJO M C U,et al.The Successive Projections Algorithm for Interval Selection in PLS[J].Microchemical Journal,2013,110:202-208.
- [9] GOODARZI M,SAEYS W,DE ARAUJO M C U,et al.Binary Classification of Chalcone Derivatives with LDA or KNN Based on Their Antileishmanial Activity and Molecular Descriptors Selected Using the Successive Projections Algorithm Feature-selection Technique[J].European Journal of Pharmaceutical Sciences,2014,51:189-195.
- [10] TANG G,HUANG Y,TIAN K D,et al.A New Spectral Variable Selection Pattern Using Competitive Adaptive Reweighted Sampling Combined with Successive Projections Algorithm[J].The Analyst,2014,139(19):4894-4902.
- [11] 刘爽,谭鑫,刘成玉,等.高光谱数据处理算法的小麦赤霉病籽粒识别[J].光谱学与光谱分析,2019,39(11):3540-3546.
LIU S,TAN X,LIU C Y,et al.Recognition of Fusarium head blight wheat grain based on Hyperspectral Data Processing Algorithm[J].Spectroscopy and Spectral Analysis,2019,39(11):3540-3546.
- [12] 来文豪,周孟然,李大同等.无监督学习 AE 和 MVO-DBSCAN 结合 LIF 在煤矿突水识别中的应用[J].光谱学与光谱分析,2019,39(8):2437-2442.
LAI W H,ZHOU M R,LI D T,et al.Application of Unsupervised Learning AE and MVO-DBSCAN Combined with LIF in Coal Mine Water Inrush Recognition[J].Spectroscopy and Spectral Analysis,2019,39(8):2437-2442.
- [13] 周孟然,胡锋,闫鹏程,等.基于 FCM 的煤矿突水激光诱导荧光光谱分析[J].光谱学与光谱分析,2018,38(5):1572-1576.
ZHOU M R,HU F,YAN P C,et al.Laser induced fluorescence spectrum analysis of water inrush in coal mine based on FCM[J].Spectroscopy and Spectral Analysis,2018,38(5):1572-1576.
- [14] 李佳平,沈国康,欧耀明,等.应用连续投影算法及最小二乘支持向量机的单组分纺织品识别[J].纺织学报,2018,54(2):023001.
LI J P,SHEN G K,OU Y M,et al.Knot deflection on coniferous wood surface by near infrared spectroscopy and successive projections algorithm[J].Journal of Textile Research,2018,54(2):023001.
- [15] 周竹,尹建新,周素茵,等.基于近红外光谱与连续投影算法的针叶材表面节子缺陷识别[J].激光与光电子学进展,2017,54(2):311-319.
ZHOU Z,YIN J X,ZHOU S Y,et al.Defect Recognition of Coniferous Wood Surface Knots Based on Near-infrared Spectroscopy and Continuous Projection Algorithm[J].Laser & Optoelectronics Progress,2017,54(2):311-319.
- [16] 刘嘉蔚,李奇,陈维荣,等.基于概率神经网络和线性判别分析的 PEMFC 水管理故障诊断方法研究[J].中国电机工程学报,2019,39(12):3614-3621.
LIU J W,LI Q,CHEN W R,et al.Research on PEMFC Water Management Fault Diagnosis Method Based on Probabilistic Neural Net-

- work and Linear Discriminant Analysis[J].Proceedings of the CSEE,2019,39(12):3614-3621.
- [17] 张燕,许京,陈玲玲,等.基于激光距离传感器的路况识别系统的设计[J].激光与红外,2016,46(3):265-270.
ZHANG Y,XU J,CHEN L L,et al.Design of terrain recognition system based on laser distance sensor[J].Laser & Infrared,2016,46(3):265-270.
- [18] 梁晴晴,韩华,崔晓钰,等.基于主元分析-概率神经网络的制冷系统故障诊断[J].化工学报,2016,67(3):1022-1031.
LIANG Q Q,HAN H,CUI X Y,et al.Fault Diagnosis of Refrigeration System Based on Principal Component Analysis and Probabilistic Neural Network[J].CIESC Journal,2016,67(3):1022-1031.

Continuous projection algorithm for screening characteristic wavelength of laser induced fluorescence spectrum of edible oil

Zhou Mengran, Yu Daoyang, Hu Feng, Bian Kai, Lai Wenhao, Sun Lei, Yan Pengcheng

(College of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232000, China)

Abstract: Aiming at the difficulty to quickly and accurately identify the type of edible oil in the current quality supervision field, a laser-induced fluorescence technology combined with continuous projection algorithm is proposed to identify the edible oil spectrum. In the experiment, a laser-induced fluorescence system is set up to collect a total of 500 sets of fluorescence spectrum data of 5 edible vegetable oils. First, the Savitzky-Golay convolution smoothing algorithm is selected through experimental comparison to preprocess the fluorescence spectrum; then the continuous projection algorithm is used to filter the characteristic wavelengths; finally, the selected characteristic wavelengths are sent to the established probabilistic neural network model for training and testing, re-verifying through the independent verification set of oil samples collected at different times. The results show that through the screening of characteristic wavelengths, 11 are selected from the 2 048 wavelengths in the full spectrum to reduce redundant information and the number of wavelengths is reduced to 0.54% of the original. And the accuracy of training samples and test samples in the probabilistic neural network model reached 100% and 95%, respectively, which are better than radial basis function neural networks and Back Propagation neural networks. The prediction accuracy rate in the independent validation set also reached 91%. Therefore, using the continuous projection algorithm to screen the characteristic wavelengths of the edible oil laser-induced fluorescence spectrum combined with the probabilistic neural network model can realize the rapid and accurate classification of edible oils and is universal, and provide further design of a dedicated online edible oil type identification instrument The theoretical basis.

Keywords: laser induced fluorescence; characteristic wavelength; successive projections algorithm; probabilistic neural networks; edible oil

[责任编辑 杨浦 刘洋]