文章编号:1004-5929(2024)04-0454-07

基于 PCA-ACO-SVM 算法和 FTIR 技术的植物油种类鉴别

贾丹[1*],明米娜[1],雷蕾[2],周锐[3],侯金亮[4]

(1. 河南省驻马店职业技术学院,河南驻马店,463000;
2. 河南财经政法大学,河南郑州,450046;
3. 河南农业大学,信息与管理科学学院,河南郑州 450003;
4. 郑州大学,信息工程学院,河南郑州 450001)

摘　要: 植物油种类的准确鉴别对油品的质量控制、欺诈检测、营养健康及粮油贸易等领域具有重要的意义。准确、快速鉴别植物油的种类对保证油品质量和维护市场监督至关重要。本文提出一种主成分分析结合蚁群优化支持向量机(Principal component analysis-Ant colony optimization-Support vector machine,PCA-ACO-SVM)算法结合傅里叶红外光谱(Fourier transform infrared spectroscopy, FTIR)技术快速识别植物油种类。实验采集了 6 种不同种类的植物油,并利用 FTIR 测量了样品的吸收、透射红外光谱。通过 PCA 对红外光谱数据进行降维,实现油品红外光谱特征的提取。运用 ACO-SVM 分类算法的核心参数进行参数优化,优化后的 SVM 分类模型核心参数 $C=1.1024043$ 和 $Gamma=0.1476193$ 。在该研究中使用 PCA-ACO-SVM 算法建立了植物油种类的识别模型。利用已知种类的油品进行分类模型训练和参数优化,进一步应用该模型对未知油品进行识别,通过与其他算法的对比,验证 PCA-ACO-SVM 算法在植物油种类识别中的准确性和高效率。结果表明 PCA-ACO-SVM 算法结合 FTIR 技术能够快速、准确地对植物油的种类进行识别。该方法不仅具有较高的分类准确率,而且在数据处理方面表现出较高的运算效率,非常适合大规模植物油种类识别的实际应用。综上所述,本文提出的 PCA-ACO-SVM 算法结合 FTIR 技术为植物油种类的快速识别提供了一种可行的解决方案,具有较高的效率和准确率。该方案在食品工业和质量监管等方面具有广阔的应用前景,并对植物油的质量监管具有重要意义。

关键词: 主成分分析;蚁群优化支持向量机;傅里叶变换红外光谱;植物油;分类识别
中图分类号: O433.4　　**文献标志码:** A　　**doi:** 10.13883/j.issn1004-5929.202404011

Identification of vegetable oil species based on PCA-ACO-SVM algorithm and FTIR technology

JIA Dan[1], MING Mina[1], LEI Lei[2], ZHOU Rui[3], HOU Jinliang[4]

(1. *Zhumadian Vocational and technical College*, *Zhumadian* 463000, *Henan*, *China*;
2. *Henan University of Economics Law*, *Zhengzhou* 450046, *Henan*, *China*;
3. *College of Information and Management Science*, *Henan Agricultural University*, *Zhengzhou* 450002, *Henan*, *China*
4. *School of Information and Engineering*, *ZhengZhou University*, *ZhengZhou*, 450001, *Henan*, *China*)

Abstract: Accurately identifying vegetable oil species is significant in oil quality control, fraud detection, nutrition and health, and grain and oil trade. Accurate and rapid identification of vegetable oils is essential to ensure oil quality and maintain market supervision. This paper proposes a Principal component analysis-Ant colony optimization-Support vector machine (PCA-ACO-SVM) algorithm combined with Fourier transform infrared spectroscopy (FTIR) technology for rapidly identifying vegetable oil species. Six different kinds of vegeta-

---

收稿日期: (2024-04-02), **修改日期:** (2024-04-17)
基金项目: 河南省终身教育专项课题和课程开发立项(202370216),国家自然科学基金(61806180)
作者简介: 贾丹(1982.8-),女,本科,讲师,研究方向:图像识别及人工智能算法,E-mail:MichaelJackson2008@126.com

ble oils were collected, and the absorption and transmission infrared spectra of the samples were measured by FTIR. PCA reduced the dimension of infrared spectral data, and the infrared spectral characteristics of oil products were extracted. The ACO algorithm optimizes the core parameters of the SVM classification algorithm. The optimized core parameters of the SVM classification model are $C = 1.1024043$ and $Gamma = 0.1476193$. This study uses the PCA-ACO-SVM algorithm to establish the identification model of vegetable oil species. The classification model was trained, and the parameters were optimized using the known types of oil products. The model was further applied to identify unknown oil products. By comparing with other algorithms, the accuracy and efficiency of the PCA-ACO-SVM algorithm in identifying vegetable oil types were verified. The results show that the PCA-ACO-SVM algorithm combined with FTIR technology can quickly and accurately identify the types of vegetable oil. This method has high classification accuracy and high computational efficiency in data processing, which is very suitable for the practical application of large-scale vegetable oil classification. In summary, the PCA-ACO-SVM algorithm proposed in this paper, combined with FTIR technology, provides a feasible solution for rapidly identifying vegetable oil species, which has high efficiency and accuracy. The scheme has broad application prospects in the food industry and quality supervision and is of great significance to the quality supervision of vegetable oil..

Keywords: Principal component analysis; Ant colony optimizationsupport vector machine; Fourier transform infrared spectroscopy; Vegetable oil; Classification recognition

0 引言

在现代社会中，植物油作为一种重要的食用油和工业原料，广泛应用于食品加工、制药、化妆品等领域。然而，随着市场上植物油种类的增多和质量标准的提高，快速准确地鉴别和分类不同种类的植物油变得越来越重要。传统的植物油鉴别方法，如色谱分析、质谱分析等，具有操作复杂、耗时长、成本高等缺点。

近年来，基于光谱技术的植物油种类鉴别方法备受关注。苏东斌等[1]采用拉曼光谱结合机器学习算法开展了 6 种纯植物油的种类鉴别研究，结果表明支持向量机模型优于正交偏最小二乘判别模型。接昭玮等[2]借助衰减全反射（Attenuated Total Refraction, ATR）-傅里叶变换红外光谱（Fourier Transform Infrared Spectroscopy, FTIR）技术对不同类别、品牌食用植物油进行了多层次分类识别工作，分析了利用竞争性自适应重加权算法结合布谷鸟搜索算法优化的极限学习机融合模型对植物油快速分类识别的可行性。王哲等[3]研究了基于二维相关近红外光谱的植物油分类鉴别方法，利用二维同步谱的自相关谱主成分之间的欧氏距离实现常见植物油分类判别。Htet 等[4]利用偏最小二乘回归法辅助拉曼光谱定性分析植物油种类并预测了生育酚含量，建立了对应的判别模型和定量模型。Jiménez-Carvelo 等[5]采用 ATR-FTIR 开展了橄榄油与其他植物油分类和掺假定量的研究工作，实现了纯橄榄油和其他植物油的准确分类，其中橄榄油实现了 100% 的识别，其他植物油分类识别准确率为 92%。

FTIR 作为一种广泛应用的光谱分析技术，具有非破坏性、高灵敏度、快速、定性定量等优点，被广泛用于植物油的分析。本文提出了一种结合主成分分析（Principal Component Analysis, PCA）、人工蚁群算法（Ant Colony Optimization, ACO）和支持向量机（Support Vector Machine, SVM）的算法来快速识别植物油种类。主成分分析作为一种常用的降维技术，能够将高维的光谱数据转化为低维的特征子空间，有效减少数据维度并提取最关键的特征信息。人工蚁群算法作为一种模拟蚁群觅食行为的优化算法，能够通过模拟蚁群的集体搜索行为，在特征子空间中找到最佳的判别特征子集。最后，支持向量机作为一种优秀的分类器，能够通过构建超平面实现植物油的分类。通过将 PCA、ACO 和 SVM 相结合，本文旨在提出一种快速、准确的植物油识别方法，以满足现代化生产和质量控制的需求。本研究对于提高植物油鉴别技术的精确性和效率，以及推动植物油行业的

发展具有积极的推动作用。

1 材料与方法

1.1 仪器条件

仪器设备:ABB 公司 FTPA2000 傅里叶变换红外光谱仪,光谱采集范围 400～4000 cm^{-1},光谱分辨率 4 cm^{-1},扫描次数为 16 次。使用长方体石英比色皿测量纯净植物油,其中光谱测量的比色皿光程为 2 cm,每个植物油种类均随机采集 500 条光谱,如图 1 所示。采集光谱时,参考背景选用空比色皿的光谱作为基准。

1.2 油品来源

本文共研究了 6 种纯植物油,分别为特级初榨橄榄油(Extra virgin olive oil,EVOO),大豆油(Soybean oil,SBO),玉米油(Corn oil,CO),稻米油(Rice oil,RO),菜籽油(Rapeseed oil,RSO)和葵花油(Sunflower oil,SFO),每种植物油各有 5 种不同品牌。如表 1 所示。由于实验对象是纯植物油,所以每种植物油每个品牌单独制备 5 个独立样本。

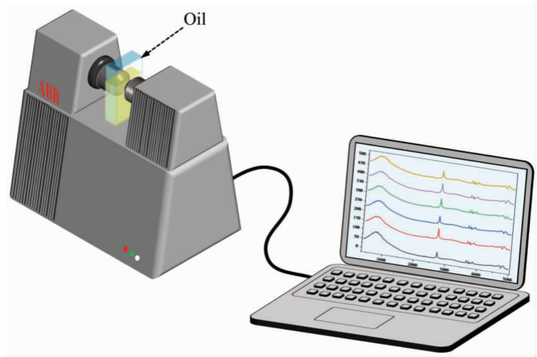

图 1 实验仪器示意图

Fig. 1 Schematic diagram of experimental instrument

表 1 实验所用的植物油种类及品牌

Table 1 Types and brands of vegetable oils used in the experiment

类别	EVOO	SBO	CO	RO	RSO	SFO
品牌	欧丽薇兰	金龙鱼	金龙鱼	金龙鱼	金龙鱼	金龙鱼
	蓓琳娜	福临门	福临门	得乐康	福临门	福临门
	品利	鲁花	长寿花	井江	鲁花	北大荒
	欧贝拉	长寿花	西王	长寿花	道道全	多力
	贝蒂斯	九三	刀唛	得乐康	长安花	刀唛
数量	25	25	25	25	25	25

1.3 数据处理

本文采用 Matlab R2019a 软件处理数据,PCA-ACO-SVM 算法基于此数学软件编写和运行,文中所有光谱图通过 Origin 9.0 软件绘制。

2 结果与分析

2.1 植物油的 FTIR 吸收、透射光谱

利用 FTIR 法测得的植物油的吸收、透射红外光谱如图 2 所示。图 2(a)为 6 种植物油的 FTIR 红外吸收光谱,图 2(b)为植物油的 FTIR 红外透射光谱。FTIR 技术中 ATR-FTIR 是应用最广泛的技术手段,由于缺少 ATR 晶体材料本文使用比色皿直接测量纯植物油的吸收和透射光谱。

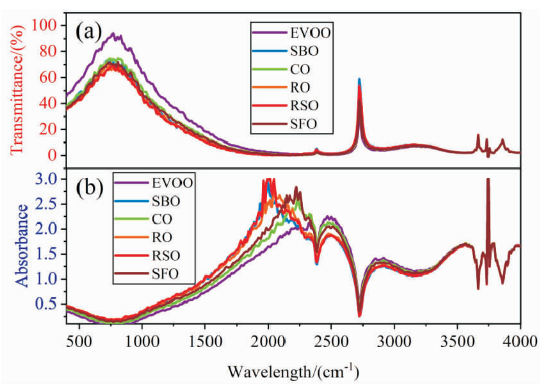

图 2 6 种植物油的 FTIR 红外光谱。(a)吸收光谱;(b) 透射光谱

Fig. 2 FTIR spectra of six vegetable oils. (a) Absorption spectra; (b) Transmission spectra

不同植物油透射光谱的差异集中在 500～1500 cm^{-1} 和 2500～3000 cm^{-1},而吸收光谱的差异集中在 1500～2500 cm^{-1},因此透射光谱和吸收光谱的融合可以最大程度的提供更多的植物油种间的差异特征。本文将吸收和透射光谱拼接,并定义为吸透谱(Absorption Transmission Spectroscopy,ATS),形成的新的一维矩阵包含的两种光谱特征信息,光谱的维度也从 600 扩增至 1200,如图 3 所示。

2.2 基于 PCA 数据降维

采用经典的数据降维算法 PCA 对高维(1200 维)光谱数据进行降维处理[6]。选择前三个主成分

绘制 6 种植物油的 3000 条数据的聚类效果,如图 4 所示。图 4 表明 PCA 算法不仅降低的数据维度,还提取了光谱特征。

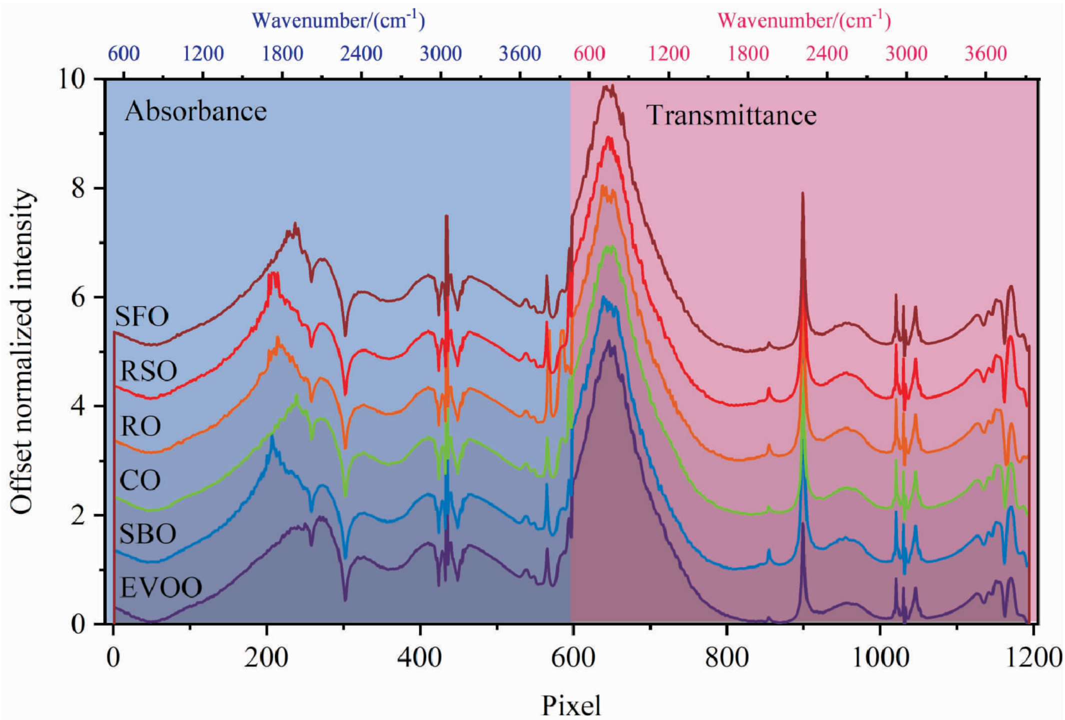

图 3　6 种纯植物油的 ATS

Fig. 3　ATS of 6 kinds of pure vegetable oils

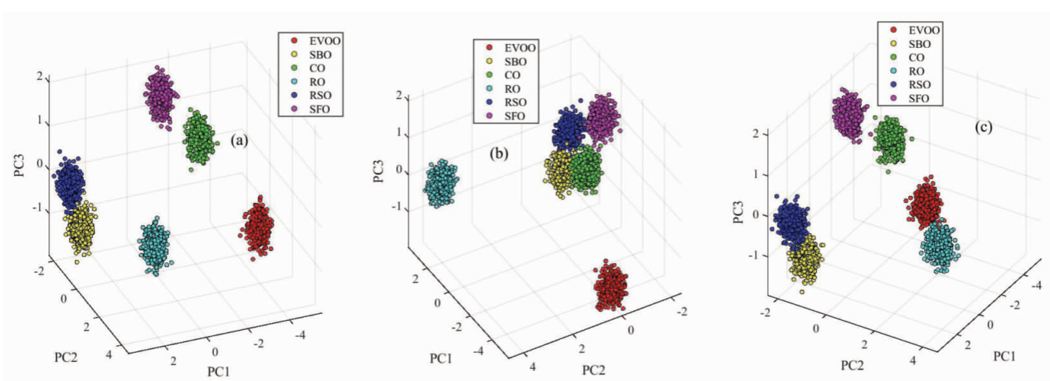

图 4　基于 PCA 算法的 FTIR 光谱降维

Fig. 4　Dimensionality reduction of FTIR spectra based on PCA algorithm

为评价 PCA 数据降维前后变量间相关性的变化,采用皮尔逊相关系数(Pearson correlation coefficient,Pearson's r)作为统计度量,将 Pearson's r 简写为 r:

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{Y})^2}} \tag{1.1}$$

其中,x_i 和 y_i 分别代表两个变量第 i 个值,$\bar{X}$ 和 $\bar{Y}$ 分别代表两个变量的均值。当 r 为正且数值越趋近于 1,说明表明两个变量存在线性正相关;当 r 为负且数值越趋近于-1,说明表明两个变量存在线性负相关;当 r 数值越趋近于 0,说明表明两个变量不存在相关性。根据式(1.1),分别计算了 ATS 原始数据集和 PCA 降维数据集的 r,根据计算结果绘制图 5 所示的热力图。

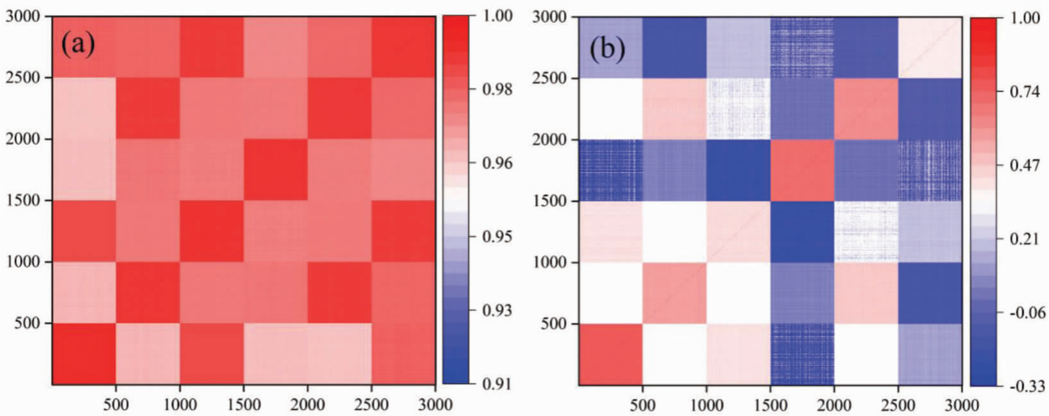

图 5 皮尔逊相关系数热力图,(a) ATS 原始数据集皮尔逊相关系数热力图,(b) PCA 数据集皮尔逊相关系数热力图

Fig. 5 Heatmap of r, (a) heatmap of r of ATS original data set, (b) heatmap of r of PCA data set

图 5(a)中 $0.9 < r < 1$,这表明不同品种植物油的 ATS 存在很大的形似性,而图 5(b)中 $-0.33 < r < 1$,这表明不同品种植物油的 PCA 数据存在较大差异,即特征提取后的矩阵更好的反映了种间的差异性和种内的相似性。结合 PCA 降维数据集通过设定 r 的阈值也可实现对油品种类的鉴别,但阈值的设定很可能引入人为判断错误,因此只用 r 衡量 PCA 降维算法的效果和统计学意义。数据集拆分过程中将 PCA 降维数据集按照 3∶1∶1 随机生成训练集、验证集及测试集。

2.3 基于 ACO 算法的 SVM 核心参数优化

蚁群优化算法(Ant Colony Optimization,ACO)是一种源于大自然生物世界的仿生进化算法,它是上世纪九十年代科学家 Maniezzo 和 Colorni 等[7]人通过模拟自然界中蚂蚁集体寻径行为而提出的一种基于种群的启发式随机搜索算法。支持向量机算法(Support Vector Machine,SVM)是一种经典的机器学习算法,可应用于解决分类和回归问题,是由科学家 Vapnik 和 Coworkers[8] 于 1992 年提出的监督算法。SVM 的核心思想是通过寻找新的超平面,使得两个类别间的间隔最大,从而达到分类的最优化。本文的 SVM 分类算法选用的核函数为高斯核函数,高斯核函数是一种局部性强的核函数,相比于多项式核函数拥有更少的参数,ACO 优化基于高斯核函数的 SVM 效率更高,高斯核函数主要的两个参数分别为惩罚因子 C 和 $Gamma$,其中 C 是分类误差的宽容度,$Gamma$ 是支持向量影响半径的倒数。ACO 算法通过调节 SVM 中高斯核函数的 C 和 $Gamma$ 寻找最优参数值,并建立基于最优参数值的分类模型。

ACO-SVM 算法程序可概括为:参数初始化,构建解空间,更新信息素及迭代收敛终止程序。其中,参数初始化是对蚁群规模、信息素重要程度因子,启动函数重要程度因子、最大迭代次数等的初始化;构建解空间是将所有蚂蚁随机生成不同的始发地,并计算访问地点;更新信息素是计算所有蚂蚁的总路径并记录最优路径;迭代收敛终止程序是 ACO 实现 SVM 核心参数的最优化。

经过 ACO 优化后的 SVM 分类模型核心参数 $C = 1.1024043$ 和 $Gamma = 0.1476193$,分类模型训练集、验证集及测试集的混淆矩阵如图 6 所示。

评价分类模型的三个重要参数分别为准确率(Precision, P)、召回率(Recall, R)和 F 值:

$$P = \frac{TP}{TP + FP} \tag{1.2}$$

$$R = \frac{TP}{TP + FN} \tag{1.3}$$

$$F = \frac{2PR}{P + R} \tag{1.4}$$

其中,TP 表示正样本正确预测为正的数量,FP 表示负样本错误预测为正的数量,FN 表示正样本错误预测为负的数量。PCA-ACO-SVM 分类模型的评价参数结果如表 2 及图 7 所示。

图 7 为 PCA-ACO-SVM 分类模型训练集、验证集及测试集的评价参数电线图,图 7 表明 EVOO、CO、RO 及 SFO 在分类模型的预测中未出现预测错误的情况,而 SBO 和 RSO 在分类模型中出现负样本错误预测和正样本错误预测的情况,这主要源于两种纯植物油的 ATS 相似度较高导

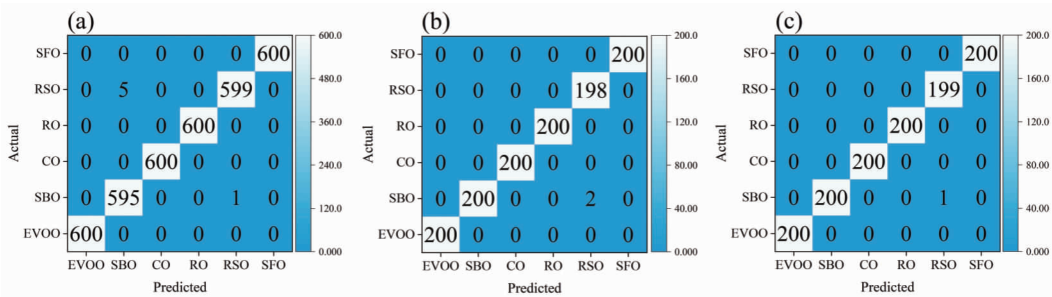

图 6　PCA-ACO-SVM 分类模型的混淆矩阵

Fig. 6　Confusion matrix of PCA-ACO-SVM classification model

表 2　PCA-ACO-SVM 分类模型预测效果

Table 2　Prediction effect of PCA-ACO-SVM classification model

评价参数	Training set			Validation set			Test set		
	P_{tr}	R_{tr}	F_{tr}	P_v	R_v	F_v	P_{te}	R_{te}	F_{te}
EVOO	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SBO	0.9917	0.9983	0.9950	1.0000	0.9901	0.9950	1.0000	0.9950	0.9975
CO	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RO	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RSO	0.9983	0.9917	0.9950	0.9900	1.0000	0.9950	0.9950	1.0000	0.9975
SFO	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

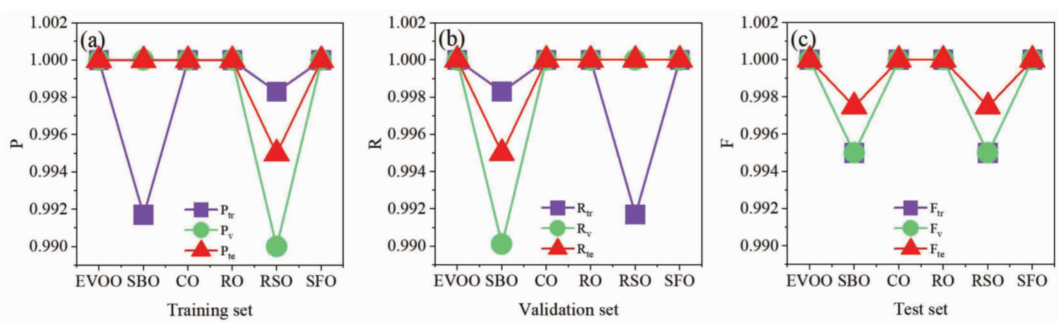

图 7　PCA-ACO-SVM 分类模型评价参数的点线图

Fig. 7　Point line diagram of evaluation parameters of PCA-ACO-SVM classification model

致的结果。为了对比 PCA-ACO-SVM 分类模型的预测法效果,将 PCA 与 K 邻近分类算法(k-nearest neighbor classification)相结合,并使用训练集、验证集和测试集构建 PCA-KNN 分类模型,模型分类效果如图 8 所示。

对比图 7 和 8 的 P 、R 及 F 评价参数可以发现 PCA-ACO-SVM 分类模型与 PCA-KNN 分类模型在对 EVOO、CO、RO 及 SFO 纯植物油预测时均达到 100%的分类识别。而在 SBO 和 RSO 纯植物油的分类效果中,PCA-ACO-SVM 分类模型中三个评价参数 P 、R 及 F 均达到 0.99 以上,而 PCA-KNN 分类模型中评价参数 P 、R 及 F 均小于 0.90。总之,本文将 PCA-ACO-SVM 分类算法与 FTIR 技术相结合实现了对 6 种纯植物油的快速无损鉴别。

3　结论

在缺少 ATR 晶体的情况下,本文采用传统的 FTIR 技术结合 PCA-ACO-SVM 分类模型开展了 6 种植物油的分类识别。通过谱线拼接的方式提高了不同种类植物油的光谱差异,并基于 PCA 算法实现了数据降维和特征提取,为后续 ACO-SVM 分类模型提供可靠的特征信息。结果表明在缺少 ATR 配件的情况下,植物油的吸收和透射光谱质量一般,但通过结合 PCA-ACO-SVM 分类模型仍可以实现对植物油种类的鉴别。因此,在光谱质

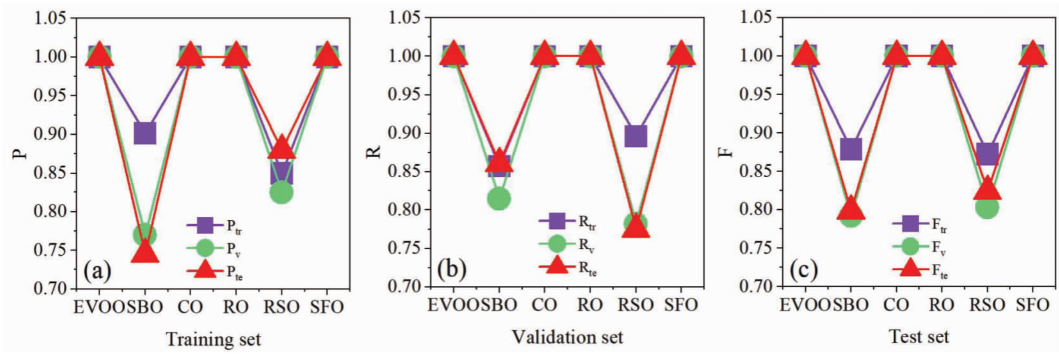

图 8 PCA-KNN 分类模型评价参数的点线图

Fig. 8 Point line diagram of evaluation parameters of PCA-KNN classification model

量一般的情况下，通过先进或更复杂的机器学习算法可以仍可以实现植物油种类鉴别的定性分析。本研究对光谱仪器的再度开发应用和植物油种类监督具有一定的借鉴价值。

参考文献

[1] 苏东斌，秦嘉桧，李开开. 拉曼光谱结合机器学习对植物油的分类鉴别[J]. 食品与发酵工业，2024，50(6)：274-281.（Su D B，Qin J H，Li KK. Classification and identification of vegetable oils by Raman spectroscopy combined with machine learning[J]. Food and Fermentation Industry，2024，50(6)：274-281.）

[2] 接昭玮，李绅，汪睿璇，等. 红外光谱快速识别食用植物油种类的研究[J]. 中国油脂，2023，48(07)：56-61.（Jie Z W，Li S，Wang R X，et al. Study on rapid identification of edible vegetable oils by infrared spectroscopy[J]. China Fats & Oils，2023，48(07)：56-61.）

[3] 王哲，李晨曦，钱蕊，等. 二维相关近红外光谱的植物油鉴别方法[J]. 光谱学与光谱分析，2020，40(10)：3230-3234.（Wang Z，Li C X，Qian R，et al. Vegetable oil identification by two-dimensional correlation near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis，2020，40(10)：3230-3234.）

[4] Htet T T M，Cruz J，Khongkaew P，et al. PLS-regression-model-assisted raman spectroscopy for vegetable oil classification and non-destructive analysis of alpha-tocopherol contents of vegetable oils[J]. Journal of Food Composition and Analysis，2021，103，104119.

[5] Jiménez-Carvelo A M，Osorio M T，Koidis A，et al. Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FT-IR-ATR and Raman spectroscopy[J]. LWT，2017，86：174-184.

[6] Shi W Y，Wu G. Perturbation analysis on PCA plus graph embedding methods and PCA plus exponential graph embedding methods[J]. Journal of Computational and Applied Mathematics，2024，444：115788.

[7] Dorigo M，Maniezzo V，Colorni A. Ant system：optimization by a colony of cooperating agents[J]. IEEE Trans Syst Man Cybern B Cybern. 1996，26(1)：29-41.

[8] Boser B，Guyon I，Vapnik V. A training algorithm for optimal margin classifiers[J]. Proceedings of the Fifth Annual Workshop on Computational Learning Theory，Pittsburgh. 1992，5：144-152.