

核极限学习机和激光诱导荧光技术在食用油识别中的应用

周孟然, 王锦国*, 宋红萍, 胡锋, 来文豪, 卞凯

安徽理工大学电气与信息工程学院, 安徽 淮南 232001

摘要 针对现有的食用油检测技术无法快速、准确地识别市售食用油的问题,提出了一种快速辨识食用油的方法。采用激光诱导荧光技术(LIF)获取油样的荧光光谱数据,然后采用主成分分析法提取光谱数据的特征信息,之后采用飞蛾-扑火优化器和核极限学习机相结合的算法建立多元分类学习模型,最后用该模型识别油样的类别。实验油样选取 5 种样本,每种样本采集 150 组荧光光谱,然后随机抽取 600 个样本用于学习模型的训练,剩余的 150 个用于测试训练好的模型。结果表明:在测试集上的平均分类准确率方面,该模型与极限学习机、反向传播神经网络相差不大,但该模型分类准确率的标准差远小于其他两种模型。这说明所建模型具有较稳定的分类性能,可以满足快速鉴别食用油的要求。

关键词 光谱学; 激光诱导荧光技术; 主成分分析法; 飞蛾-扑火优化算法; 核极限学习机

中图分类号 O657.3

文献标志码 A

doi: 10.3788/LOP57.203001

Application of Kernel Extreme Learning Machine and Laser Induction Fluorescence Technique in Edible Oil Identification

Zhou Mengran, Wang Jinguo*, Song Hongping, Hu Feng, Lai Wenhao, Bian Kai

*College of Electrical and Information Engineering, Anhui University of Science and Technology,
Huainan, Anhui 232001, China*

Abstract Existing edible oil detection technology cannot quickly and accurately identify edible oils sold in markets. Hence, in this paper, we propose a quick method of identifying edible oils. Fluorescence spectrum data of oil samples were obtained using the laser induction fluorescence(LIF) technique. Principal component analysis was used to extract characteristic information. Next, a multiclassification learning model was developed through the fusion algorithm of moth-flame optimization and kernel extreme learning machine (KELM) to identify the type of oil samples. Five types of oil samples were selected for experimental purposes, and 150 groups of fluorescence spectra were collected from each sample. Next, 600 samples were randomly selected to train the learning model, and the remaining 150 samples were used to test the trained model. Experimental results show that KELM model, extreme learning machine model and back propagation neural network model have similar average classification accuracy on the test set. However, the standard deviation of KELM model is less than those of other two models. This shows that KELM model has a stable classification performance and can quickly identify edible oils.

Key words spectroscopy; laser induction fluorescence technique; principal component analysis; moth-flame optimization algorithm; kernel extreme learning machine

OCIS codes 300.2530; 100.4994; 300.6170

收稿日期: 2019-12-04; 修回日期: 2020-01-02; 录用日期: 2020-01-09

基金项目: 国家重点研发计划(2018YFC0604503)、国家安全生产重大事故防治关键技术科技项目(anhui-0001-2016AQ)、安徽省青年科学基金(1808085QE157)

* E-mail: wangjinguo1023@163.com

1 引言

食用油对人们的身体健康至关重要,而近年来食用油的安全问题层出不穷,因此,对食用油进行鉴别和分类具有重要意义。传统的鉴别食用油的方法有液相色谱法和气相色谱法。液相色谱法^[1]具有准确快速和可重复性等特点;而气相色谱法^[2]具有简单、快速和准确性高等特点。传统的食用油鉴别方法需要在实验前对油样进行处理。目前已有采用近红外光谱法对食用油进行鉴定和分类的研究。近红外光谱法^[3-6]具有分析速度快以及非破坏性和无污染等优点,但灵敏度较低。

近些年,激光诱导荧光(LIF)技术^[7-8]逐渐兴起。该技术已被成功应用于化工、医疗和生物等领域,具有快速和灵敏度高等特点,为食用油的识别提供了一种新思路。来文豪等^[9]将LIF技术与深度学习结合起来应用于假酒的识别上,有效辨识出了白酒的真假;周孟然等^[10]将AdaBoost算法应用于矿井突水水源LIF光谱的识别和分类研究上,对矿井突水类型实现了有效识别;王亚等^[11]将LIF技术与极限学习机结合起来应用于矿井突水水源的识别上,实现了矿井突水类型的快速、准确识别。

本文将LIF技术与核极限学习机(KELM)结合起来应用于食用油的鉴别和分类上。首先采用LIF技术对食用油进行采样,然后运用主成分分析(PCA)法^[12]对数据进行降维,之后采用飞蛾-扑火(MFO)算法优化KELM参数,并将数据送入KELM中进行分类。

2 实验方法

2.1 实验仪器

油样的荧光光谱是在油样检测实验平台上采集

表1 食用油的种类和参数

Table 1 Types and parameters of edible oils

Type	Brand	Level	Production process	Producing area	Production time
Peanut oil	Luhua	A	Press	Shandong	2019-07-31
Soybean oil	Fulinmen	B	Leach	Jiangsu	2019-07-31
Corn oil	Changshouhua	A	Press	Shandong	2019-06-06
Rapeseed oil	Tuoniao	D	Leach	Jiangsu	2019-01-25
Sunflower seed oil	Luhua	A	Press	Inner Mongolia	2018-08-24

花生油品牌为鲁花,食用油等级为一级,生产工艺为压榨,原料产地为山东,生产日期为2019年7月31日。

大豆油品牌为福临门,食用油等级为一级,生产

得到的。油样检测实验系统简图如图1所示。

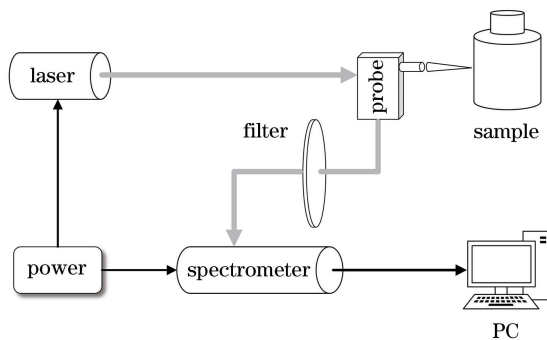


图1 油样检测实验系统简图

Fig. 1 Oil sample detection experiment system diagram

实验选用的半导体激光器(北京华源拓达激光技术有限公司)可发出波长为405 nm的激光,入射激光的最大功率为150 mW。实验选用USB2000+型激光诱导荧光光谱仪(美国海洋光学公司),它可以接受波段在340~1020 nm的激光,光谱采样间隔为0.5 nm,积分时间为500 ns。激光器和荧光光谱仪由五股照明光纤组成的光纤束连接至浸入式激光微型探头(广东科思凯公司,型号为FPB-405-V3),其中光纤接口为SMA905接头。采集油样的激光诱导荧光光谱时,用铁架台将微型探头固定(探头靠近待测油样),以保证采集过程中探头保持在同一位置。采集的油样光谱在经滤光片滤除无用波段荧光后,由光纤输送至荧光光谱仪。最后利用上位机中的光谱软件Spectra Suite采集并记录波段范围内的荧光光谱图样。

2.2 油样采集

本次实验选用的食用油分别为花生油、大豆油、玉米油、菜籽油和葵花籽油,每种油样采集150组光谱数据。上述5种食用油的特性如表1所示。

工艺为浸出,原料产地为江苏,生产日期为2019年7月31日。

玉米油品牌为长寿花,食用油等级为一级,生产工艺为压榨,原料产地为山东,生产日期为2019年

6月6日。

菜籽油品牌为鸵鸟,食用油等级为四级,生产工艺为浸出,原料产地为江苏,生产日期为2019年1月25日。

葵花籽油品牌为鲁花,食用油等级为一级,生产工艺为压榨,原料产地为内蒙古,生产日期为2018年8月24日。

2.3 实验方法

2.3.1 MFO 算法

MFO 算法^[13]是仿照自然界中飞蛾的飞行路径提出来的,根据飞蛾聚集在火焰周围的行为建立数学模型。飞蛾是候选解,飞蛾在解空间中的位置是所求问题的参数,因此飞蛾能通过改变它们的位置向量在一维、二维、三维以及更高维的空间中运动。

本算法选用对数螺旋作为飞蛾的主要更新机制,MFO 算法的对数螺旋 $S(M_i, F_j)$ 定义为

$$S(M_i, F_j) = D_i \cdot \exp(bt) \cdot \cos(2\pi t) + F_j, \quad (1)$$

式中: D_i 表示第 i 个飞蛾到第 j 个火焰的距离; b 定义为对数螺旋形状的常数; t 是一个随机数, $t \in [-1, 1]$; F_j 表示第 j 个火焰; M_i 表示第 i 个飞蛾。为了更加强化开发能力, t 被设在 $[r, 1]$ 区间的随机数,其中 r 为一个变下限参数。变量 t 能够使飞蛾根据迭代次数并利用相应的火焰来更好地改变自身的位置。

2.3.2 ELM 和 KELM 模型

极限学习机(ELM)^[14]是一种新的单隐层前馈神经网络(SLFNs)。在 ELM 中,隐含层的参数是随机选择的,不需要通过多次迭代进行更新。ELM 的网络结构图如图 2 所示,其中, $f(x)$ 为 ELM 的输出函数, m 为输出层的节点数, $G(a_i, b_i, x)$ 为 ELM 的激活函数, n 为输入层的节点数, L 为隐含层的节点数。

ELM 的输出函数定义为

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta, \quad (2)$$

式中: β_i 为第 i 个隐含层节点和输出层节点之间的输出权重向量; $h_i(x)$ 为第 i 个隐含节点产生的输出向量。

在二分类应用中,对于 N 个训练样本 $(x_i, t_i) \in \mathbf{R}^d \times \mathbf{R}^m$, t_i 表示第 i 个训练样本的类比向量, $t_i \in \{-1, 1\}^m$ 。因此,ELM 的激活函数 $G(w, b, x)$ 由 $H\beta = T$ 进行数学建模,其中 w 为输入权重, H 为隐含层产生的输出矩阵。

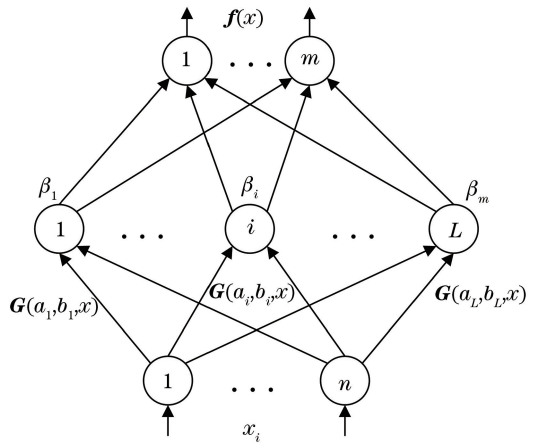


图 2 ELM 网络结构

Fig. 2 Network structure of ELM

可以用 $\hat{\beta} = H^+ T$ 来计算得出 β 的最优输出权值 $\hat{\beta}$,其中 H^+ 是 H 的 Moore-Penrose 广义逆。本文用核矩阵来替代 ELM 中的随机初始数,所以 ELM 的输出函数可以写成

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T B, \quad (3)$$

其中,

$$B = \left(\frac{I}{C} + HH^T \right)^{-1} T, \quad (4)$$

式中: C 为正则化系数; $K(x, x_N)$ 为核矩阵。

相对于 ELM, KELM^[15]虽然没有显著地缩短时间,但精度更高。

2.4 算法流程

本文将 MFO 算法与 KELM 分类器结合起来对油样的荧光光谱进行分类。算法流程如图 3 所示。算法步骤如下:

- 1) 数据预处理。利用主成分分析法对油样的荧光光谱数据进行降维。
- 2) 初始化飞蛾数、火焰数、最大迭代次数、飞蛾的初始位置、全局最优位置等,初始化 KELM 参数 C 和径向基核函数的参数 g 。
- 3) 更新飞蛾位置和火焰位置及其适应度值,并记录最优位置和最佳适应度值。
- 4) 若达到最大迭代次数,则算法结束,转向步骤 5),否则转向步骤 3)。
- 5) 输出 KELM 参数 C 和径向基核函数的参数 g 的全局最优值以及最佳适应度值,算法结束。

利用 MFO 算法自动选择 KELM 模型中的参数,克服了人为选择的主观性,避免了人为尝试,节

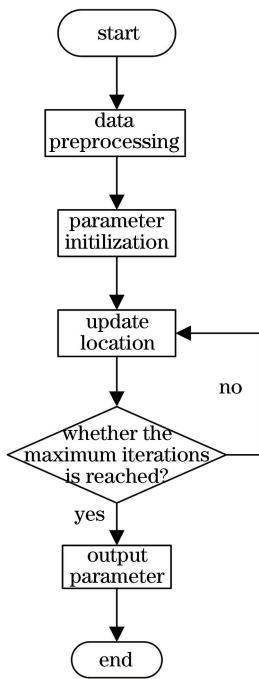


图3 算法流程图

Fig. 3 Algorithm flowchart

约了时间。

3 实验结果与讨论

笔者采用 KELM 分类器,并结合 PCA 降维算法和 MFO 优化算法建立食用油分类学习模型,采用该模型对 5 种油样的荧光光谱进行分类,并将 MFO 算法与灰狼优化(GWO)算法和粒子群优化(PSO)算法的参数寻优性能进行比较,将所建模型与 ELM 和 BP 神经网络模型的分类性能进行比较。三种模型都在 MATLAB 2017b 环境中运行,硬件环境为 Core i5-5200U 2.2 GHz CPU, RAM 4 GB。实验运行了 100 次。

3.1 原始荧光光谱

在图 1 所示的实验平台上获取原始荧光光谱数据。为了更清晰地显示谱图,选取 340~1021 nm 波段间的荧光强度绘制成图。图 4 中给出了花生油、玉米油、葵花籽油、大豆油和菜籽油的原始荧光光谱。

3.2 PCA 降维后的数据

将 340~1021 nm 波段的所有荧光光谱数据作为 PCA 的输入。当设置累计贡献率为 95% 时,可获得主成分的个数为 2;当设置累计贡献率为 99% 时,可获得主成分的个数为 3,如图 5 所示。

本文将 2048 维数据降到 3 维,在不损失数据主要信息的前提下,降低了数据维度,简化了数据结构。

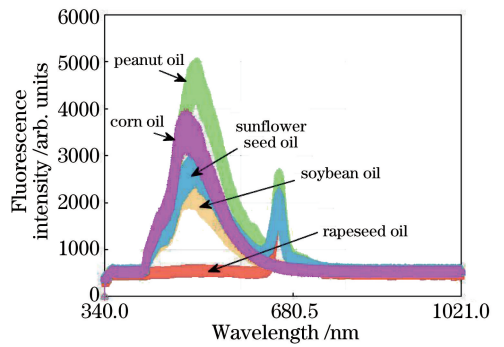


图4 5种油样的原始荧光光谱

Fig. 4 Original fluorescence spectra of five oil samples

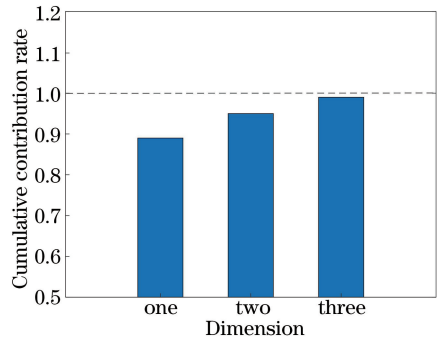


图5 累计贡献率图

Fig. 5 Figure of cumulative contribution rate

3.3 MFO 算法选取的参数

将主成分数为 3 的降维数据作为输入数据,可以大大简化网络结构。本文利用 MFO 算法寻找 KELM 分类器中的最佳参数。MFO 算法中的飞蛾数和火焰数都设置为 5,参数 C 的寻优范围为 $[0.25, 4]$,参数 g 的寻优范围为 $[1, 300]$ 。MFO 寻优算法的测试分类准确率如图 6 所示,可以看出,MFO 算法在参数寻优时的表现比较好,可以快速收敛到全局最优值。在算法开始迭代时,油样的分类准确率就达到了 90% 以上,在迭代次数达到 5 时已经收敛到全局最优值,分类准确率达到 95.33%。

为了比较 MFO 算法与 GWO 算法、PSO 算法

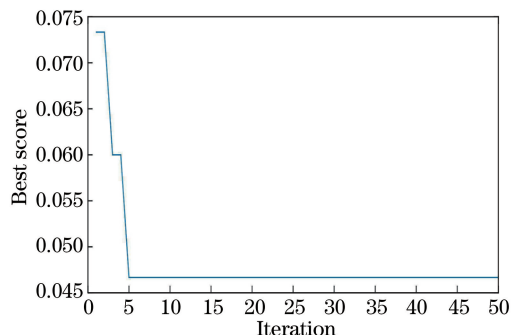


图6 MFO 算法的测试分类准确率

Fig. 6 Test classification accuracy of MFO algorithm

在寻优时间和寻优性能上的差异,将输入数据统一划分成 600 个训练样本和 150 个测试样本,参数 C 的范围固定为 $[0.25, 4]$, 参数 g 的范围固定为 $[1, 300]$, 飞蛾数目、狼群数目和粒子群数目都设置为 5。实验结果如表 2 所示,可以看出,三种寻优算法的平均时间都在 195 s 以上并且相差不到 2 s, 分类准确率都在 90% 以上。虽然 PSO 算法的寻优时间短于 MFO 算法和 GWO 算法,但其平均分类准确率却比其他两种算法低了约 4 个百分点,这说明在 150 个测试样本中 PSO 算法相比其他两个算法多分错了 6 个样本。此外, MFO 算法和 GWO 算法的寻优时间只差了 1 s 左右,两者的平均分类准确率都在 95% 以上,说明在 150 个测试样本中,这两种算法错分的样本不超过 8 个。

表 2 三种算法的性能比较

Table 2 Performance comparison of three algorithms

Algorithm	Model practical time /s	Average recognition rate /%
MFO	196.43	95.33
GWO	195.29	95.33
PSO	194.77	91.16

三种优化算法的分类错误率和迭代次数的关系图如图 7 所示,纵坐标为分类错误率,横坐标为迭代次数。由图 7 可以看出: PSO 算法的寻优性能明显低于其他两种算法, PSO 算法在迭代次数为 10 时基本上趋于稳定,分类准确率达到 90% 以上,但却远远低于 MFO 算法和 GWO 算法; MFO 算法和

表 3 三种分类模型的性能比较

Table 3 Performance comparison of three classification models

Model	Model practical time /s	Average recognition rate /%	Recognition rate standard deviation
KELM	0.0255	95.62	0.0157
ELM	0.0025	95.16	6.5940
BP	0.2137	94.77	0.0640

由表 3 可以看出: 三种分类模型的平均分类准确率相差不大,基本都保持在 94% 以上,说明这三种分类模型对油样样本都具有良好的分类学习能力;在多次实验中, BP 神经网络模型的平均训练时间是最长的,远长于其他两种分类模型,一次训练时间达到了 0.2 s 左右,而 ELM 模型的平均训练时间是三种模型中最低的,只有 0.0025 s。但是由图 8(b) 可以看出, ELM 模型分类准确率的稳定性较差,最高的分类准确率可以达到 100%, 而最低的准确率却只有 72% 左右。这是由于 ELM 模型在隐含层随机选取权重,导致该算法的泛化能力较差。由表 3 还可以看出, KELM 模型的平均训练时间只

GWO 算法最终的准确率基本一样,都达到了 95% 以上,但 MFO 算法的收敛速度比 GWO 算法快,在迭代次数为 5 时, MFO 算法已收敛到全局最优值,而 GWO 算法在迭代次数为 9 时才达到最优值。由此可以看出,在 KELM 参数寻优上, MFO 算法的寻优性能要优于 PSO 算法和 GWO 算法。

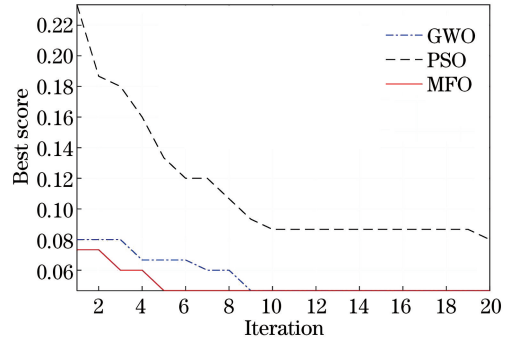


图 7 三种寻优算法测试分类准确率的比较

Fig. 7 Comparison of test classification accuracy of three algorithms

3.4 KELM 分类模型及与其他模型比较

在 KELM 分类学习之前,先使用 PCA 法进行特征降维,以有效降低网络的复杂程度,提高运算效率。参数 C 、 g 取 MFO 寻优得到的最优参数,以保证分类器能够得到较好的分类效果。

对油样进行 100 次实验,每次随机抽取 600 个样本作为训练集,剩余的 150 个样本作为测试集,然后求取平均训练时间、平均分类准确率和标准差,比较三种分类模型的性能,比较的结果如表 3 所示。

有 0.0255 s,而且其分类准确率也比较稳定,如图 8(a) 所示。

KELM、ELM、BP 模型的标准差分别为 0.0157、6.5940、0.0640,如表 3 所示。虽然 KELM 分类模型的平均训练时间稍微长于 ELM,但其分类性能的稳定性远大于 ELM。这说明, KELM 分类模型具有良好的泛化能力。

结合图 9 和表 3 可知: BP 网络模型分类稳定性较好,但平均训练时间过长,无法满足油样分类的快速性要求; ELM 模型的平均训练时间最短,可以满足油样分类的快速性要求,但其分类稳定性较差,无法满足分类的稳定性要求;而 KELM 模型不仅可

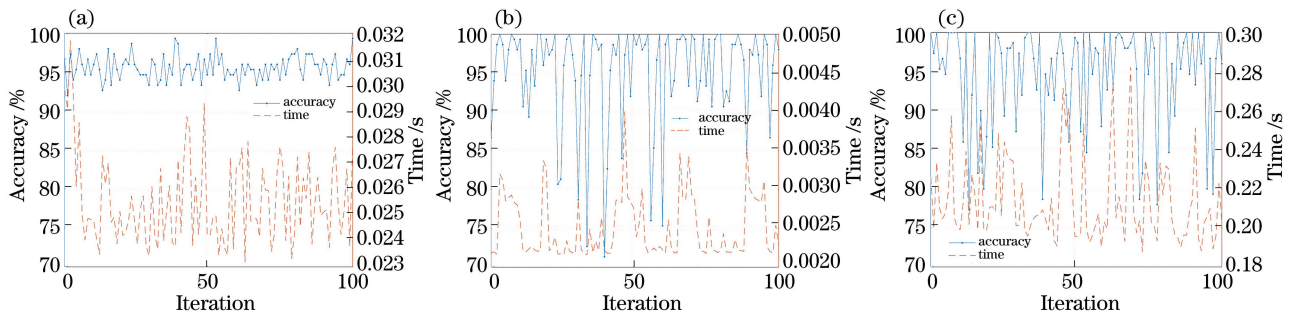


图 8 三种模型分类准确率和训练时间的统计。(a) KELM 模型;(b) ELM 模型;(c) BP 模型

Fig. 8 Statistics of classification accuracy and training time of three models.

(a) KELM model; (b) ELM model; (c) BP model

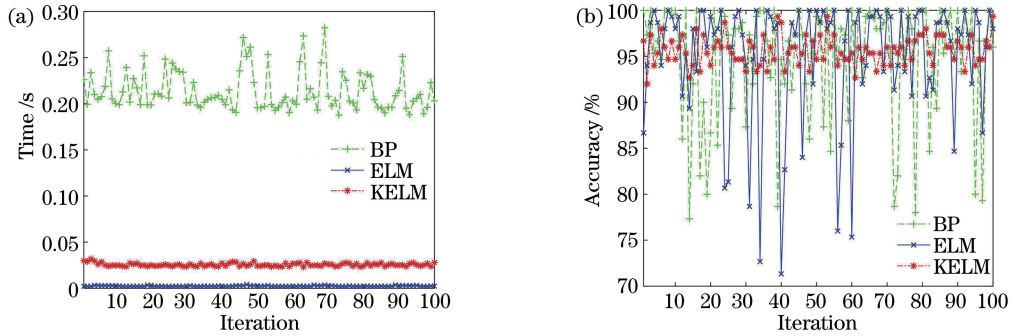


图 9 三种模型的平均训练时间和分类准确率的对比。(a)平均训练时间;(b)分类准确性

Fig. 9 Comparison of average training time and classification accuracy of three models.

(a) Average training time; (b) classification accuracy

以满足油样鉴定的实时性要求,还可以满足油样类别鉴定的准确性要求和泛化要求。

三种算法的分类准确率比较如图 9(b)所示,可以清楚地看到 KELM 模型分类准确率密集地分布在 95%左右,再次说明了 KELM 模型分类稳定性远高于其他两种算法,可以满足食用油检测的稳定性要求。

4 结 论

利用 LIF 技术获取油样的荧光光谱数据,然后采用 PCA 方法对油样的光谱数据特征进行降维,之后采用 MFO 算法对 KELM 的参数进行寻优,建立 KELM 的分类判别模型,并将 KELM 模型与传统的分类学习模型 ELM 和 BP 进行比较。实验结果表明,与传统的分类学习模型相比,KELM 模型具有较高的分类准确率,并且分类准确率的标准差非常小,能够保持较好的分类效果。KELM 模型在不损失网络训练时间的前提下,保证了分类的泛化能力,满足了油样鉴定的实时性、准确性以及稳定性。

参 考 文 献

- [1] Luo J G. Determination of F-2 toxin in corn oil by immune-affinity column purification and HPLC [J]. China Oils and Fats, 2018, 43(5): 88-91.
- [2] 罗建光. 免疫亲和和柱净化-高效液相色谱法测定玉米油中 F-2 毒素含量的研究 [J]. 中国油脂, 2018, 43(5): 88-91.
- [2] Qiu R F, Huang Z P, Wang L L. Analysis of fatty acid composition in cottonseed by gas chromatography with on-line pyrolytic methylation [J]. Chinese Journal of Chromatography, 2018, 36(9): 925-930.
- 邱若凤, 黄忠平, 王丽丽. 在线热辅助甲基化-气相色谱法分析棉籽仁中的脂肪酸组成 [J]. 色谱, 2018, 36(9): 925-930.
- [3] Zuo Q Q, Sun J M, Wang Q Y, et al. Optimization of identification model of adulterated peanut oil based on near infrared spectroscopy [J]. The Food Industry, 2019, 40(4): 318-322.
- 左倩倩, 孙金梦, 王倩玉, 等. 基于近红外光谱的掺伪花生油鉴别模型优化 [J]. 食品工业, 2019, 40(4): 318-322.
- [4] Yan H, Zhang J X, Gao J X, et al. Towards improvement in prediction of iodine value in edible oil system based on chemometric analysis of portable vibrational spectroscopic data [J]. Scientific Reports,

- 2018, 8(1): 14729.
- [5] Bi X, Jin Y B, Li S F, et al. Rapid and sensitive determination of fatty acids in edible oil by liquid chromatography-electrospray ionization tandem mass spectrometry[J]. *Science China Chemistry*, 2014, 57(3): 447-452.
- [6] Chen Y, Luo Q S, Wang J, et al. Rapid identification and characterization of recovered edible oil, based on Raman and near-infrared spectroscopy [C]//Proceedings of the 2018 3rd International Conference on Modelling. Paris, France: Atlantis Press, 2018: 4.
- [7] Hu F, Zhou M R, Yan P C, et al. Influence of temperature on laser induced fluorescence spectroscopy of mine goaf water [J]. *Spectroscopy and Spectral Analysis*, 2018, 38(8): 2583-2587.
胡锋, 周孟然, 闫鹏程, 等. 温度对矿井老空水激光诱导荧光光谱的影响[J]. *光谱学与光谱分析*, 2018, 38(8): 2583-2587.
- [8] Zhou M R, Song Q, Wang Y, et al. Application of LIF technology and XGBoost algorithm in counterfeit wine recognition [J]. *Applied Laser*, 2019, 39(1): 130-135.
周孟然, 宋奇, 王亚, 等. LIF技术和XGBoost算法在假酒识别中的应用[J]. *应用激光*, 2019, 39(1): 130-135.
- [9] Lai W H, Zhou M R, Wang Y, et al. Application of counterfeit liquor recognition based on deep learning and laser induced fluorescence [J]. *Laser & Optoelectronics Progress*, 2018, 55(4): 043001.
来文豪, 周孟然, 王亚, 等. 深度学习与激光诱导荧光在假酒识别中的应用[J]. *激光与光电子学进展*, 2018, 55(4): 043001.
- [10] Zhou M R, Li D T, Hu F, et al. Research of the AdaBoost arithmetic in recognition and classifying of mine water inrush sources fluorescence spectrum [J]. *Spectroscopy and Spectral Analysis*, 2019, 39(2): 485-490.
周孟然, 李大同, 胡锋, 等. AdaBoost算法在矿井突水水源的荧光光谱识别中的研究[J]. *光谱学与光谱分析*, 2019, 39(2): 485-490.
- [11] Wang Y, Zhou M R, Yan P C, et al. A rapid identification model of mine water inrush based on extreme learning machine [J]. *Journal of China Coal Society*, 2017, 42(9): 2427-2432.
王亚, 周孟然, 闫鹏程, 等. 基于极限学习机的矿井突水水源快速识别模型[J]. *煤炭学报*, 2017, 42(9): 2427-2432.
- [12] Wang Y, Zhou M R, Yan P C, et al. Identification of coalmine water inrush source with PCA-BP model based on laser-induced fluorescence technology [J]. *Spectroscopy and Spectral Analysis*, 2017, 37(3): 978-983.
王亚, 周孟然, 闫鹏程, 等. PCA-BP模型在判别基于LIF技术煤矿突水水源的应用[J]. *光谱学与光谱分析*, 2017, 37(3): 978-983.
- [13] Mirjalili S. Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm [J]. *Knowledge-Based Systems*, 2015, 89: 228-249.
- [14] Deng C W, Huang G B, Xu J, et al. Extreme learning machines: new trends and applications [J]. *Science China Information Sciences*, 2015, 58(2): 1-16.
- [15] Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2012, 42(2): 513-529.