

基于荧光光谱和堆栈自编码器的食用油快速无损检测

周孟然, 戴荣英*, 杨晨, 胡锋, 卞凯, 来文豪, 孔茜茜

安徽理工大学电气与信息工程学院, 安徽 淮南 232001

摘要 针对传统检测方式的食用油损耗大、操作烦琐、耗时长等缺陷, 提出了一种食用油种类快速无损检测的新思路。实验选用包括混合油样在内的 5 种待测油样本, 利用搭建出的激光诱导荧光系统采集数据 500 组, 随机选取其中 400 组光谱数据作为训练集, 余下的 100 组作为测试集。选用性能更为优异的堆栈自动编码器算法对获取的荧光光谱数据进行特征提取, 通过极限学习机进行分类识别, 最后利用不同时间测出的食用油样本验证模型的普适性。实验结果表明, 在所构建的识别模型下, 样本测试网络时间仅为 0.2 ms, 分类准确率可达到 100%, 用于验证的新油样同样可取得极好的分类效果, 分类速度快, 准确率高。所得结果证明所建立的模型是可靠的, 能够在确保精准识别的同时, 实现食用油类别的快速无损检测。

关键词 光谱学; 激光诱导荧光; 堆栈自动编码器; 特征提取; 极限学习机; 快速无损检测

中图分类号 O657.3

文献标志码 A

doi: 10.3788/LOP202259.0830001

Fast Nondestructive Detection of Edible Oil Based on Fluorescence Spectrum and Stack Autoencoder

Zhou Mengran, Dai Rongying*, Yang Chen, Hu Feng, Bian Kai, Lai Wenhao, Kong Xixi

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

Abstract In view of the defects of traditional test methods, such as large consumption of edible oil, cumbersome operation, and long time consumption, a new idea of fast nondestructive test of edible oil was put forward. In the experiment, five kinds of oil samples including mixed oil samples were selected. The laser induced fluorescence system built in the experiment was used to collect 500 groups of data, 400 groups of spectral data were randomly selected as the training set, and the remaining 100 groups of spectral data were used as the test set. After comparison, the stack autoencoder algorithm with better performance was selected to extract the features of the obtained fluorescence spectral data, and then the extreme learning machine was used for classification and recognition. Finally, the edible oil samples measured at different time were used to verify the generalization of model. The experimental results show that, under the recognition model constructed in this paper, the sample test network time is only 0.2 ms, and the classification accuracy can reach 100%. The sample test network used to validate the new sample can also achieve good classification effect, the classification is fast, and the accuracy is high. That is to say, the model established in this paper is reliable, and it can also realize fast nondestructive test of edible oil types while ensuring accurate identification.

收稿日期: 2021-05-26; 修回日期: 2021-06-20; 录用日期: 2021-07-27

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)、安徽省自然科学基金能源互联网联合基金重点支持项目

通信作者: *dry19970325@163.com

Key words spectroscopy; laser induced fluorescence; stack autoencoder; feature extraction; extreme learning machine; rapid nondestructive test

1 引言

在当前利益市场的驱动下,食用油掺假现象尤为严重^[1],各种劣质混合油以次充好,这扰乱了市场秩序,使得消费者身体健康受到影响。因此快速无损地鉴别食用油类别不但维护了消费人群的共同利益^[2],而且对食用油市场的稳定有着重大意义。

市场上常用的食用油检测手段一般是采用气相色谱或者其与气相色谱质谱相结合的方式^[3-4],但实验过程中样品损耗大,操作烦琐且耗时长^[5]。除此之外,作为热门检测方式之一的近红外光谱检测法虽然识别时间短、对样品的损耗小^[6],但却存在信噪比低、谱图重叠率高等缺陷,而拉曼光谱技术的信号强度相对较弱,在环境因素影响下易出现信号丢失现象^[7]。相较于传统的检测手段,激光诱导荧光(LIF)光谱分析技术具有较高的分析速度、精度和灵敏度^[8-10],在诸多领域都有着广泛的应用^[11-12],为食用油的快速无损精准分析提供了基础。如Hu等^[13]将LIF光谱学与一维卷积神经网络(CNN)相结合,以快速准确地实现矿井突水的识别。如来文豪等^[14]将LIF技术与深度学习算法相结合,实现了对白酒类别以及对应度数的判断等。

堆栈自动编码(SAE)神经网络本质上是由自动编码器堆叠而成的^[15],SAE算法相较于其他降维方式如主成分分析法(PCA)等^[16-17],可在保证可控降维的前提下单独对每一层进行训练,将复杂问题简单化,这有利于加快任务的完成,因此SAE具有极好的特征提取效果,已广泛应用到工业、农业、军事等领域^[18-19]。任俊等^[20]将堆栈自动编码器与支持向量机(SVM)算法相结合以实现军事装备体系效能的评估。饶利波等^[21]将高光谱成像技术结合堆栈自动编码器和极限学习机(ELM)来实现苹果硬度的检测等。

为实现食用油种类的快速无损检测,本文提出一种基于荧光光谱和堆栈自编码的食用油种类快速无损检测研究方法。先利用LIF技术获取原始荧光光谱图,再选取具有最优隐含层层数的无监督学习SAE算法进行特征提取,以有效提高检测的识别精度。在各个分类算法皆取得极佳的分类效果前

提下,本文最终选取样本测试网络速度最高的ELM作为该研究的分类器,而将荧光光谱和SAE用于食用油种类的快速无损检测尚未见相关报道。

2 原理与算法

2.1 SAE降维算法

深度学习可以在深层体系结构中自动学习分层特征,并且能够基于大量无标签数据建立训练模型,该模型覆盖的范围广且适应性极强,无监督学习算法利用多层神经网络表征学习,不需要事先对训练样本进行标记处理。2006年,在自动编码器(AE)的原理之上,一种可用于高效学习编码的新型人工神经网络——无监督学习堆栈自动编码器(SAE)被提出^[22]。该算法采用了逐层贪婪式的学习过程,将多个自动编码层堆叠成一个新的深度神经网络层,再逐渐从繁杂的高阶输入数据中提取出同系列的简明特征信息。

实验将从每条光谱曲线上获取2048个原始油样光谱数据,因此选用SAE算法来降低初始数据维度,实现识别精度的有效提升。根据图1,SAE主要分为以下4个训练过程:1)在初始给定输入 X 的基础上,对第一层隐含层 V 进行无监督式的训练,通过重构误差的最小化操作使得输入 X 与重建后的输出 X' 的设定值保持不变;2)将从AE1中提取出的隐含层 V 作为下一个自动编码器AE2的输入,再以相

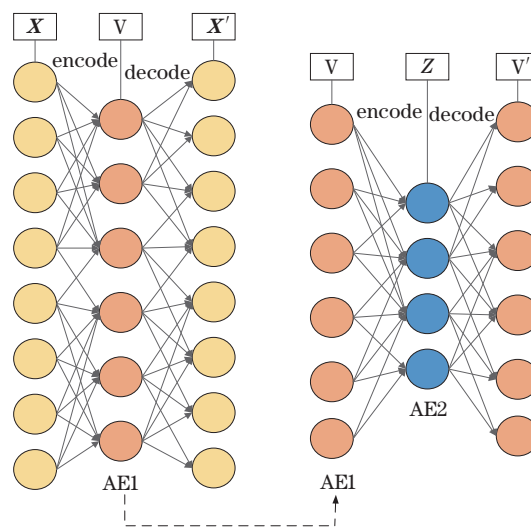


图1 SAE网络架构

Fig. 1 SAE network architecture

同的步骤对下一隐含层进行训练;3) 循环进行步骤 2), 直到所有自动编码器完成初始化任务;4) 将最终训练出的隐含层输出(图 1 中的 Z) 提取出来作为分类器的输入。

SAE 与 AE 内部结构相似, 大体可以分为输入层、编码层和重建层。其中将输入层映射到编码层的非线性映射函数可表示为

$$H_i = f(X_i) = \frac{1}{1 + e^{-(\omega_i X_i + b_i)}}, \quad (1)$$

式中: ω_i 表示编码器部分的权值; b_i 为对应的偏置向量; X_i 表示初始给定输入, 对应于图 1 中的 X ; f 表示映射。

用解码器重构的非线性函数可表示为

$$X'_i = g(H_i) = \frac{1}{1 + e^{-(\omega_2 H_i + b_2)}}, \quad (2)$$

式中: ω_2 为编码器部分的权值; b_2 为偏置向量。利用误差反向传播算法不断地调整权值与偏置大小, 使得重构误差取得最小值。最终实现的最小化重构误差可表示为

$$\partial = \min \frac{1}{N} \sum_{i=1}^N \|X_i - X'_i\|^2, \quad (3)$$

式中: N 表示训练样本集的大小; ∂ 为最小化重构误差。

在进行实验的过程中, 输入数据的降维特征表达是通过限制隐含层的维度实现的, 即将编码器维度设置为较输入数据维度更小的值。

2.2 ELM 算法

ELM 算法与传统的训练方式有所不同, 是从前馈神经网络(FNN)延伸出的机器学习方式, 适用于处理监督以及非监督学习问题, 其原理是任意选取输入权重大小以及隐含层偏置, 并训练具有单个隐含层的单层前馈神经网络(SLFN), 最后利用 Moore-Penrose(MP) 广义逆矩阵理论来求解输出层权重值。自 ELM 算法被提出起, 该算法被广泛地研究与应用。传统的 ELM 具有单隐含层, 相比于单层感知机或 SVM 等浅层学习系统, 可以减少模型训练参数与时长, 效率高, 非常适用于进行数据分类。

ELM 作为神经网络时, 一般会使用 SLFN 的结构, 隐含层的输出函数 $f_L(x)$ 为

$$f_L(x) = \sum_{i=1}^L \beta_i h(\omega_i * x_j + d_i), \quad (4)$$

$$j = 1, 2, \dots, N,$$

式中: L 代表隐含单元的数量; N 表示训练样本的数

量; x 为神经网络的输入向量; d_i 表示偏置向量; w_i 表示输入与输出之间的权重向量; β_i 表示第 i 个隐含层与输出之间的权重向量; $h(x)$ 为特征映射或激励函数, 可用于在 ELM 的特征空间中映射出初始输入层的数据。ELM 的特征映射具有一定随机性, 这是因为 ELM 中输入层至隐含层的特征映射是随机或人为设定的, 且无需进行调整。

3 实验部分

3.1 实验材料

实验选取了市场上包括山东鲁花集团生产的花生油、菜籽油和中国粮油控股有限公司生产的福临门玉米油在内的三种油样作为实验材料。考虑到市场上花生油价值较高, 不良商贩可能利用混合油以次充好、谋取利益, 为实现油样的混合检测, 按照 1:1 的体积比对花生油与玉米油、花生油与菜籽油进行混合处理。表 1 选取菜籽油、玉米油、花生油、花生油与菜籽油的体积比为 1:1 的混合油(以下简称混合油样 A)、花生油与玉米油的体积比为 1:1 的混合油(以下简称混合油样 B)作为实验样本。分别从每种油中获取 100 组光谱样本, 共计 500 组样本, 将其中的 400 组油样随机选取出来作为训练样本, 其余的 100 组油样用作测试样本。为保证实验数据是真实且可靠的, 对所有油样进行避光保存。

表 1 实验样本材料

Table 1 Experimental sample materials

Label	Type	Volume ratio
1	Rapeseed oil	-
2	Corn oil	-
3	Peanut oil	-
4	Peanut oil mixed with rapeseed oil	1:1
5	Peanut oil mixed with corn oil	1:1

3.2 荧光光谱数据采集

一旦被测油样的成分不同, 它们所激发出的光谱强度也会有差异。如果油样成分差距很大, 则获取的荧光光谱图可表现出该差异, 且其空间距离很大。油样中的不同化学成分所吸收的光能量和散射出的光能量导致油样反映出不同的光谱数据, 这就为油样种类的无损检测提供了新的思路和方法。

为了实现食用油种类的有效鉴别, 实验选取了入射激光功率范围可达到 100~120 mW 且连续可调的蓝紫光半导体激光器(波长为 405 nm)。为了

提高耦合效率,选用了芯径为 600 μm 、孔径为 0.22NA 的 UV/VIS 石英光纤。为提高测量速度、降低成本,实现在线分析,实验选用了光谱范围为 340.472~1021.086 nm,通信模式为 232 的美国海洋公司生产的 USB 2000+ 微型光谱仪。为保证油类的快速无损检测,本实验实现了激光器与荧光探头的一体化,直接对油样进行激光激发,避免了侵入式荧光探头与油样的直接接触,再利用滤光片滤除无用荧光光谱信息后,将采集结果经由光纤传送至微型光谱仪中,最终获取的光谱数据由上位机中的 SpectraSuite 软件记录保存。为了避免光照条件对食用油的影响,实验过程皆在同一暗室下进行。油样检测设备如图 2 所示。

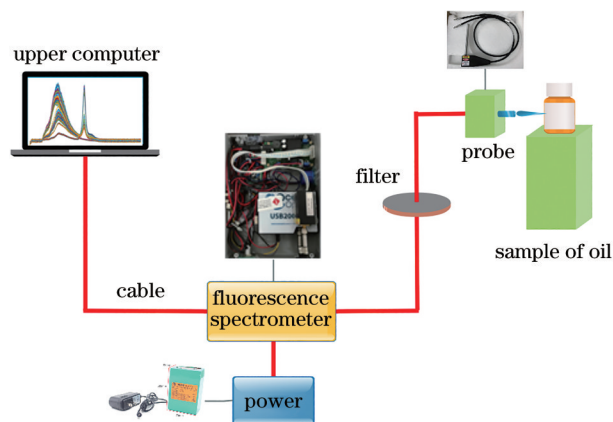


图 2 LIF 实验设备示意图

Fig. 2 Diagram of experimental equipment for LIF

4 结果与讨论

4.1 原始光谱图分析

借助图 2 所搭建出的实验平台可以获取待测油样的原始荧光光谱数据。图 3(a) 所示为全波段范围内 5 种待测油样分别对应的荧光光谱图,横坐标表示与光谱对应的波长大小,而纵坐标表示荧光强度区间,根据曲线的变化趋势,不难发现,不同油样荧光强度的差异主要集中在 400~800 nm 波段之间,当波长值超出 800 nm 后,光谱曲线几乎无变化。

图 3(b)~(f) 分别对应于菜籽油、玉米油、花生油、混合油样 A、混合油样 B 的光谱曲线变化图,可见菜籽油、玉米油、花生油的光谱曲线波峰差异明显,相对易于辨识,但是对于混合油样本来说,波峰与单一油样存在明显重叠的现象,无法被直观地辨别出。因此,需要对原始荧光光谱数据进行处理,本文采用了无监督 SAE 算法进行特征提取,再选用速度最优的 ELM 算法对油样进行分类识别。本实验的所有算法均是在 Matlab R2020b 环境下完成的,所用计算机处理器为 2.60 GHz 双核 Inter Core i7-9750 H,运行内存为 16G。

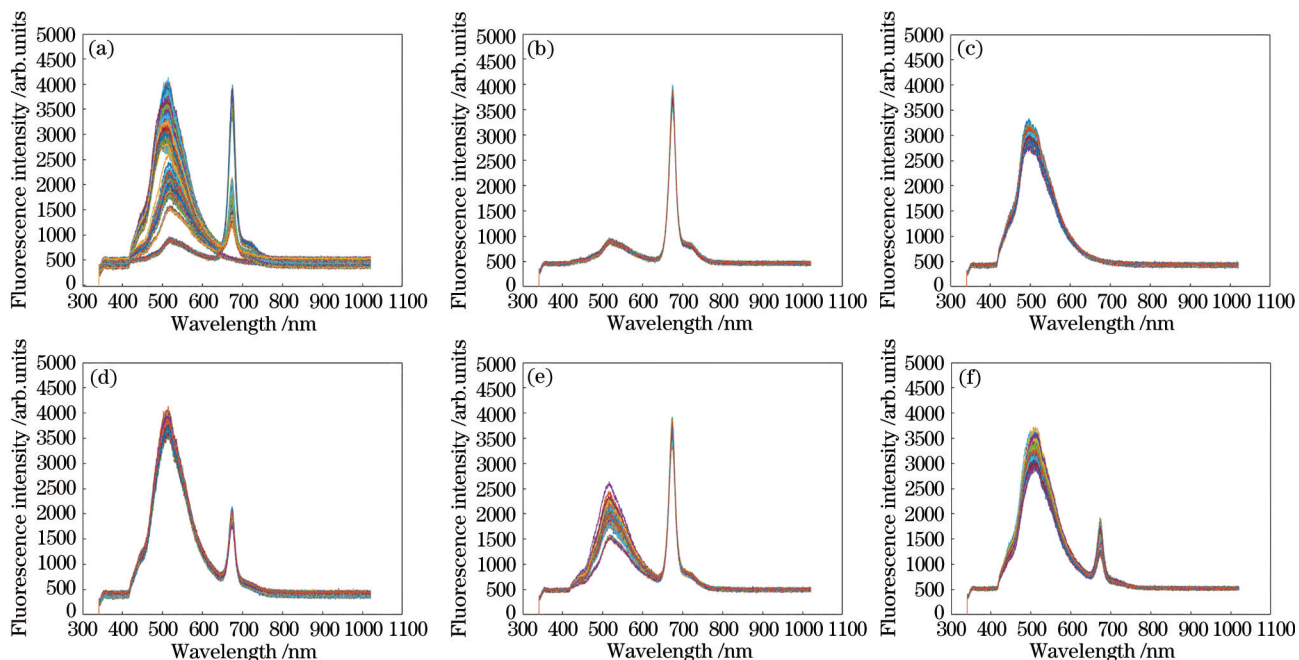


图 3 不同食用油样品的原始荧光光谱图。(a) 所有样品; (b) 菜籽油样品; (c) 玉米油样品; (d) 花生油样品; (e) 混合油样品 A; (f) 混合油样品 B

Fig. 3 Original fluorescence spectra of different edible oil samples. (a) All samples; (b) sample of rapeseed oil; (c) sample of corn oil; (d) sample of peanut oil; (e) mixed oil sample A; (f) mixed oil sample B

4.2 SAE降维

实验所获取的油样光谱数据维度为 2048, 维度较大, 包含的冗余信息自然较多, 这将会对最终分类结果产生干扰。为保证分类结果的准确性, 选用无监督学习 SAE 提取油样荧光光谱特征, 以降低原始数据的维度。

在 Matlab 仿真下, 本文对 5 种油样的 500 组样本按照 4:1 的比例进行随机划分, 将每种待测油样

中的 80 组数据随机选取出来用作训练集, 余下的 20 组数据作为实验的测试集, 共计 400 组测试样本与 100 组训练样本。在构建出的 SAE 模型中, 隐含层层数的多少会对分类结果产生影响, 为找出合适的层数, 实验比较了 4 个实验组, 隐含层层数从 1 开始依次递增。实验采集到的原始数据为 2048 维, 所以将输入层的节点数设置成 2048, 通过多轮参数迭代调优选出的各实验组隐含层层数见表 2。

表 2 各实验组的隐含层层数

Table 2 Number of hidden layers in each experimental group

Experimental group	Number of first hidden layer	Number of second hidden layer	Number of third hidden layer	Number of fourth hidden layer
1	300	-	-	-
2	300	50	-	-
3	300	50	10	-
4	300	50	10	5

每组实验运行 5 次, 选用系统自带的 softmax 分类器测试并记录每次的分类结果, 求取相应的方差, 以此作为最优隐含层层数选取的参数指标。4 个实验组的测试分类准确率如图 4 所示。可见隐含层层数的不同将对数据后续分类产生较大的影响, 第一层隐含层经过训练作为第二层隐含层的输入时, 可在原基础上再一次进行降维处理, 如此重复, 使得每次的训练过程都能随着层数的增加学习到新的特征, 最终提取出的有效特征便是多种效果的叠加, 不同的叠加结果在分类器中呈现出的分类效果自然有所差异。由图 4 可知: 第一组实验整体效果最差, 测试分类准确率最高为 78%, 最低为 72%; 第二组实验的分类准确率最高, 可达到 99%,

最低为 97%, 分类准确率总体较为稳定; 第三组实验的分类准确率最高可达到 87%, 最低为 83%; 第四组实验的分类准确率最高达到 85%, 最低为 79%, 该组实验的稳定性在四组实验中最差。

每组实验的平均准确率以及方差见表 3。通过观察可知, 表 3 的前两项呈现出平均准确率递增、方差递减的现象, 而随着层数的继续增加, 平均准确率与方差的变化规律皆出现逆转现象, 即平均准确率递减、方差递增。可见, 并非隐含层层数越多, 呈现出的效果越好, 这可能是由于当隐含层层数超过 2 层后, 深度学习网络在从底层到高层特征的深入过程中, 每增加一层都会学习到新的特征, 有用特征增加的过程中也会导致部分冗余特征的增加, 从而对结果产生一定的干扰; 在此基础上, 2 层隐含层以上的堆栈自动编码器神经网络出现过学习现象, 从而影响最终的识别效果。当隐含层层数为 2 时, 平均分类准确率可达到 98.4%, 方差为 0.64, 此时的分类效果最佳, 故实验选取具有 2 层隐含层的 SAE 进行 LIF 光谱数据的特征提取。

表 3 不同隐含层数下的平均分类准确率和方差

Table 3 Average classification accuracy and variance under different number of hidden layers

Model	Number of hidden layers	Mean accuracy / %	Variance
SAE 1	1	75.4	3.84
SAE 2	2	98.4	0.64
SAE 3	3	85.4	3.84
SAE 4	4	80.6	5.04

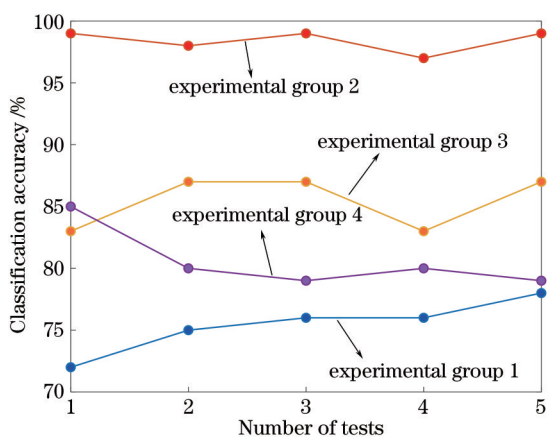


图 4 隐含层层数与 Softmax 分类准确率的关系

Fig. 4 Relationship between number of hidden layers and Softmax classification accuracy

4.3 分类模型选取与建立

为了实现不同种类油样的分类识别,在 Matlab 仿真下,选用合适的分类算法对上述经降维后维数为 50 的数据进行处理,以获取分类准确率。本文选用了 BP(back propagation)神经网络、ELM 算法、SVM 以及决策树(decision tree)算法,基于这 4 种典型且应用范围广的方法构建出 4 种不同的分类模型,以选择最佳分类识别模型。由表 3 可见,经 SAE 算法特征提取后,4 种分类算法对数据的识别效果极佳,无论是对于训练集还是测试集皆可实现 100% 的分类效果。

为进一步进行验证,选取数据挖掘中常用的一种主成分分析(PCA)降维算法进行比较,该算法可将变量通过正交变换的方式由线性相关转化为线性不相关。在将 PCA 算法用于原始光谱数据的特征提取过程中,设置 PCA 的累计贡献度为 95%,可得到 2 个主成分数。由表 4 可见,经过 PCA 降维后,在 BP、ELM、SVM 以及决策树分类算法下,无论是对于训练集还是测试集,其分类准确率皆难以达到 100%。在经 BP 神经网络分类后,训练集与测试集的分类准确率分别可达到 94%(训练集 400 组数据被正确识别出 376 组)与 97%(测试集 100 组数据被正确识别出 97 组);ELM 算法对训练集与测试集的分类准确率分别为 97%(训练集 400 组数据被正确识别出 388 组)和 99%(测试集 100 组数据被正确识别出 99 组);SVM 算法对训练集与测试集的分类准确率分别达到 97%(训练集 400 组数据被正确识别出 386 组)和 97%(测试集 100 组数据被正确识别出 97 组);而决

策树算法对训练集与测试集的分类准确率分别为 100%(训练集 400 组数据被正确识别出 400 组)和 97%(测试集 100 组数据被正确识别出 97 组)。推测其原因可能在于:PCA 算法进行特征提取只是近似地舍弃一部分特征,而其对影响最终识别准确率的部分相关特征信息是不予考虑的,这导致分类准确性在一些重要特征信息丢失后会变差。PCA 的降维效果明显差于 SAE,这也许是因为在确保可控降维的过程中,PCA 降维并不能够较好地保留数据信息。通常 PCA 需要对数据进行预处理操作,首先对数据进行中心化,当数据的尺度难以统一时,还需要对其进行标准化。那么,这种情况就会导致问题的产生,比如标准差很小,接近于零,特别是对于被噪声污染后的数据,噪声的标准差将会对这些数据产生显著的放大作用,而标准化的过程对未被噪声污染过的数据则几乎不产生放大作用。在实验过程中,荧光光谱的采集不可避免地会引起噪声等干扰因素,同时方差小的非主成分也可能含有使样本产生差异的重要信息,在进行降维后将舍弃则可能会干扰后续的数据处理,这些都导致 PCA 各个特征维度的含义具有一定的模糊性,不如原始样本特征的解释性强。而 SAE 通过压缩初始数据来产生较低维度的数据,该降维过程是通过消除重要特征上的噪声和冗余,找到数据在较低维度中的表征来实现的,相比于 PCA,SAE 无需预处理操作便可以取得极好的特征提取效果,可以有效简化数据的繁杂性,提高检测的准确率,使得任务得以顺利完成。

在 SAE 较优的降维效果下,无论选用 BP 神经网络、ELM、SVM 还是决策树算法皆可取得极高的训练集与测试集分类准确率。但为了选取最优分类算法并实现对不同种类油样的辨别,本文记录了 4 种模型针对所有样本的测试网络时间,见表 5。在经 SAE 特征提取后,利用 ELM 分类算法进行识别,其测试网络时间要明显短于其他 3 种算法,仅为 0.2 ms。因此选择 ELM 算法作为该油样快速无损检测的分类算法可使得系统的识别效果最佳。

图 5 展示了在隐含层神经元个数变化的情况下,ELM 模型的测试准确率与测试网络时间的变化情况。可以看出,当隐含层神经元的数目达到 11

表 4 不同识别模型下的分类结果

Table 4 Classification results for different recognition models

Modeling method	Accuracy of training set / %	Accuracy of test set / %
SAE+BP	100(400/400)	100(100/100)
SAE+ELM	100(400/400)	100(100/100)
SAE+SVM	100(400/400)	100(100/100)
SAE+decision tree	100(400/400)	100(100/100)
PCA+BP	94(376/400)	97(97/100)
PCA+ELM	97(388/400)	99(99/100)
PCA+SVM	97(386/400)	97(97/100)
PCA+decision tree	100(400/400)	97(97/100)

表 5 不同识别模型下的测试网络时间

Table 5 Test network time for different recognition models

Modeling method	SAE+BP	SAE+ELM	SAE+SVM	SAE+decision tree
Testing network time	197.5	0.2	1.6	16.4

时,测试集的预测精度已经达到 100%,测试网络时间随着神经元数量的增加而变化,在相同精度下,将隐含层神经元的个数设置为 11,能够在确保高精度的基础上,最大程度地降低 ELM 模型的复杂度。图 6 为 SAE-ELM 分类模型的结果。

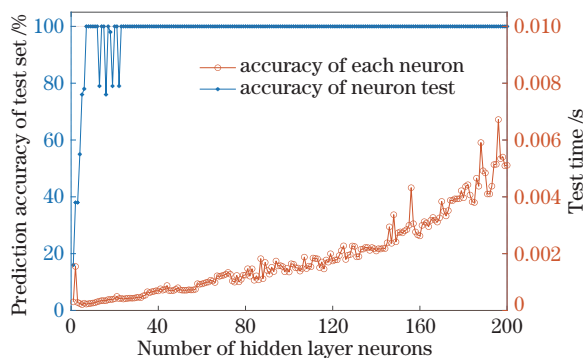


图 5 隐含层神经元对 ELM 性能的影响

Fig. 5 Effect of hidden layer neurons on ELM performance

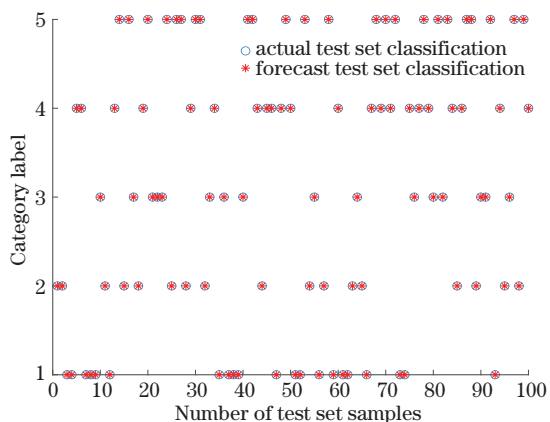


图 6 经 SAE 处理后的光谱测试集分类结果

Fig. 6 Classification results of spectrum test set processed by SAE

5 普适性的验证

良好的适应能力可以保证模型的可靠性,若同种算法对于不同食用油数据仍具有较高的分类准确性,说明该算法的分类可靠性高。为验证 SAE-ELM 模型的可靠性,实验选取了 2019 年 9 月同样利用 LIF 技术采集的花生油、大豆油、玉米油、菜籽油和葵花籽油 5 种油样作为实验样本,从 750 组样本(每种油 150 组样本)中随机选取 600 组油样作为训练样本,余下的 150 组则作为测试样本。与前文介绍的处理方式相同,首先利用具有 2 层隐含层的 SAE 算法对原始数据进行特征提取,再将提取后的数据放入 ELM 模型中进行分类,如图 7 和图 8 所示,测试网络

时间为 0.2 ms,测试集准确率同样可达到 100%,该模型的速度快且准确率高,因此本文建立的模型对单一油样和混合油样都能够实现准确分类。

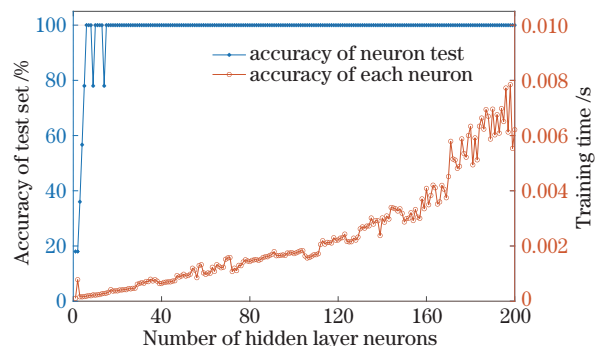


图 7 隐含层神经元对 ELM 性能的影响

Fig. 7 Effect of hidden layer neurons on ELM performance

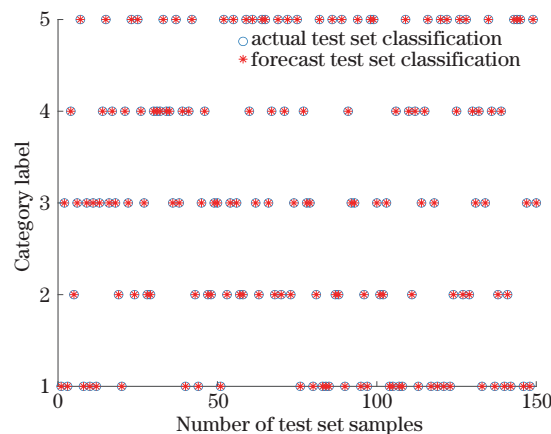


图 8 独立验证集合的分类结果

Fig. 8 Independent validation of classification results of set

6 结 论

针对传统检测方式的不足,提出了 SAE-ELM 结合荧光光谱用于食用油种类快速无损检测的新思路。实验利用分析速度快、精度和灵敏度高的 LIF 技术获取了包括花生油、玉米油、菜籽油、混合油样 A、混合油样 B 在内的三种单一油样以及两种混合油样的荧光光谱数据。经比较可得,选用较 PCA 算法更优的 SAE 算法对采集到的油样光谱数据进行特征提取,再利用 ELM 算法对特征提取出的数据进行分类识别,分类准确率达到 100%,且样本测试网络时间仅为 0.2 ms。除此之外,SAE 算法对于利用 LIF 技术采集的花生油、大豆油、玉米油、菜籽油和葵花籽油 5 种单一油样数据也可取得 100% 的分类识别率。实验结果表明,利用 SAE 算法对原始数据进行特征提取,可将原始数据从 2048

维降到 50 维,该算法可以有效降低数据维度,提高识别准确率;而所选用的 ELM 算法中模型训练参数调节简单,可以有效提高学习速率,非常适用于数据分类。所提出的 SAE-ELM 模型分类准确率高,且模型测试速度快,说明设计出的油样快速无损检测系统是有意义的,且可靠性极高。

参 考 文 献

- [1] He W X, Lin Q. Identification of sesame oil based on Fourier transform infrared spectroscopy combined with chemometrics[J]. Spectroscopy and Spectral Analysis, 2021, 41(1): 319-326.
何文绚, 林棋. 基于高效分离不皂化物的傅里叶红外光谱结合化学计量学鉴定芝麻油[J]. 光谱学与光谱分析, 2021, 41(1): 319-326.
- [2] Zhou M R, Wang J G, Song H P, et al. Application of kernel extreme learning machine and laser induction fluorescence technique in edible oil identification [J]. Laser & Optoelectronics Progress, 2020, 57(20): 203001.
周孟然, 王锦国, 宋红萍, 等. 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[J]. 激光与光电子学进展, 2020, 57(20): 203001.
- [3] Liu J, Liu Y F, Li J W. Thermal oxidation characteristics of main fatty acids in vegetable oils at high temperature by GC method[J]. China Oils and Fats, 2019, 44(1): 35-40.
刘剑, 刘元法, 李进伟. 基于气相色谱的植物油特征脂肪酸高温热氧化特性研究[J]. 中国油脂, 2019, 44(1): 35-40.
- [4] Shen W J, Wang H, Lu H Y, et al. Determination of eight vitamin E in vegetable oils by gas chromatography-mass spectrometry and its application on authentication of sesame oil[J]. Chinese Journal of Chromatography, 2020, 38(5): 595-599.
沈伟健, 王红, 陆慧媛, 等. 气相色谱-质谱法测定植物油中 8 种维生素 E 及其在芝麻油真伪鉴别方面的应用[J]. 色谱, 2020, 38(5): 595-599.
- [5] Ochiai M, Komiya Y. Detection of edible insect derived phospholipids with polyunsaturated fatty acids by thin-layer chromatography, gas chromatography, and enzymatic methods[J]. Journal of Food Composition and Analysis, 2021, 99: 103869.
- [6] Peng D, Li L Q, Liu Y L, et al. A general model for the peroxidation values of two vegetable oils based on near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2020, 40(6): 1828-1832.
彭丹, 李林青, 刘亚丽, 等. 基于近红外光谱两种植物油过氧化值通用模型研究[J]. 光谱学与光谱分析, 2020, 40(6): 1828-1832.
- [7] Yu Y T, Wang J F, Sun Y Y, et al. Identification of adulterated olive oil by two-dimensional Raman correlation spectroscopy with cooming as a perturbation factor[J]. Spectroscopy and Spectral Analysis, 2020, 40(12): 3727-3731.
于迎涛, 王季锋, 孙玉叶, 等. 采用降温扰动二维相关拉曼光谱鉴别掺假橄榄油[J]. 光谱学与光谱分析, 2020, 40(12): 3727-3731.
- [8] Zhu J J, Wan M G, Wu G, et al. Research progress of laser-induced fluorescence technology in combustion diagnostics[J]. Chinese Journal of Lasers, 2021, 48(4): 0401005.
朱家健, 万明罡, 吴戈, 等. 激光诱导荧光技术燃烧诊断的研究进展[J]. 中国激光, 2021, 48(4): 0401005.
- [9] Zhang J F, Chen L H, Yu J H, et al. Detection of polycyclic aromatic hydrocarbons in diffusion flame of propane by laser induced fluorescence[J]. Chinese Journal of Lasers, 2020, 47(4): 0411002.
张健夫, 陈玲红, 余佳涵, 等. 丙烷扩散火焰中多环芳烃的激光诱导荧光测量研究[J]. 中国激光, 2020, 47(4): 0411002.
- [10] Chen Z K, Guo R, Cheng P F. Application of LIF technology-based spectral feature extraction in oil detection[J]. Laser & Optoelectronics Progress, 2020, 57(13): 133002.
陈至坤, 郭蕊, 程朋飞. 基于 LIF 技术的光谱特征提取在油类检测中的应用[J]. 激光与光电子学进展, 2020, 57(13): 133002.
- [11] Li X, Hu J P, Li Y H, et al. Analysis of DNA restriction enzyme fragments of recombinant human herpes simplex virus type 1 by capillary electrophoresis with laser-induced fluorescence detection[J]. Chinese Journal of Pharmaceutical Analysis, 2020, 40(1): 37-42.
李响, 胡金盼, 李永红, 等. 应用 CE-LIF 方法分析重组人 1 型单纯疱疹病毒的 DNA 限制酶酶切片段[J]. 药物分析杂志, 2020, 40(1): 37-42.
- [12] Qin T, Zu L L. The study on the laser induced fluorescence spectroscopy of cyclohexyl dinitrite[J]. Spectroscopy and Spectral Analysis, 2018, 38(S1): 279-280.
秦泰, 祖莉莉. 亚硝酸环己二酯光解离过程的激光诱导荧光光谱研究[J]. 光谱学与光谱分析, 2018, 38(S1): 279-280.
- [13] Hu F, Zhou M R, Yan P C, et al. Identification of

- mine water inrush using laser-induced fluorescence spectroscopy combined with one-dimensional convolutional neural network[J]. RSC Advances, 2019, 9(14): 7673-7679.
- [14] 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.
- [15] Ghosh S K, Biswas B, Ghosh A. SDCA: a novel stack deep convolutional autoencoder: an application on retinal image denoising[J]. IET Image Processing, 2019, 13(14): 2778-2789.
- [16] Niftiyev I. Understanding principal component analysis (PCA) in the Azerbaijan economy: case studies of vegetable and fruit sectors[EB/OL]. (2021-02-09)[2021-04-05]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3782426.
- [17] Song H S, Ma L Z, Wang Y F, et al. Recognition of formaldehyde, methanol based on PCA-BP neural network[J]. Laser & Optoelectronics Progress, 2020, 57(7): 071201.
宋海声, 麻林召, 王一帆, 等. 基于 PCA-BP 神经网络对甲醛和甲醇的识别研究[J]. 激光与光电子学进展, 2020, 57(7): 071201.
- [18] Zhang G X, Lü F P. Fault-selection method of transmission line based on stacked autoencoder[J]. Water Resources and Power, 2019, 37(6): 173-177.
张国星, 吕飞鹏. 基于堆叠自动编码器的输电线路故障选相方法[J]. 水电能源科学, 2019, 37(6): 173-177.
- [19] Zhao R Z, Wen Y F, Ye X, et al. Research on frequency indicators evaluation of disturbance events based on improved stacked denoising autoencoders[J]. Proceedings of the CSEE, 2019, 39(14): 4081-4093.
赵荣臻, 文云峰, 叶希, 等. 基于改进堆栈降噪自动编码器的预想事故频率指标评估方法研究[J]. 中国电机工程学报, 2019, 39(14): 4081-4093.
- [20] Ren J, Li N. Effectiveness prediction of weapon system of systems based on stacked autoencoder dimension reduction[J]. Military Operations Research and Systems Engineering, 2017, 31(1): 61-67.
任俊, 李宁. 基于堆栈自编码降维的武器装备体系效能预测[J]. 军事运筹与系统工程, 2017, 31(1): 61-67.
- [21] Rao L B, Pang T, Ji R S, et al. Firmness detection for apples based on hyperspectral imaging technology combined with stack autoencoder-extreme learning machine method[J]. Laser & Optoelectronics Progress, 2019, 56(11): 113001.
饶利波, 庞涛, 纪然仕, 等. 基于高光谱成像技术结合堆栈自动编码器-极限学习机方法的苹果硬度检测[J]. 激光与光电子学进展, 2019, 56(11): 113001.
- [22] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.