

基于 PSO-LSSVM 和特征波长提取的 羊肉掺假检测方法

Detection method of mutton adulteration based on PSO-LSSVM
and characteristic wavelengths extraction

成甜甜^{1,2}

王克俭¹

韩宪忠¹

李 师¹

王 媛²

CHENG Tian-tian^{1,2} WANG Ke-jian¹ HAN Xian-zhong¹ LI Shi¹ WANG Yuan²

(1. 河北农业大学, 河北 保定 071000; 2. 国家羊肉加工技术研发专业中心

[衡水志豪畜牧科技有限公司], 河北 衡水 053000)

(1. Hebei Agricultural University, Baoding, Hebei 071000, China; 2. National Mutton

Processing Technology Research and Development Center [Hengshui Zhihao Animal

Husbandry Technology Co., Ltd.], Hengshui, Hebei 053000, China)

摘要:为解决羊肉—猪肉掺假快速检测这一问题,利用多光谱仪器对掺假羊肉进行光谱采集,得到样品在 350~1 100 nm 波段下的反射率。对数据预处理后,利用粒子群算法(PSO)对最小二乘支持向量机(LSSVM)进行优化,建立了基于粒子群优化的最小二乘支持向量机模型(PSO-LSSVM),与偏最小二乘(PLS)、反向传播神经网络(BPNN)和 LSSVM 3 种模型结果进行比较,结果表明,PSO 算法能有效优化 LSSVM 模型,预测的决定系数和均方根误差分别为 0.920 4 和 0.089 2。进一步采用随机青蛙算(RF)、无信息变量消除法(UVE)、竞争性自适应重加权法(CARS)提取特征波长并建立偏最小二乘模型,结果显示,UVE-PLS 模型预测集的决定系数和均方根误差分别为 0.996 7 和 0.016 2,UVE 优于其他特征波长提取方法。

关键词:羊肉;掺假;多光谱;粒子群算法;最小二乘支持向量机;特征波长

Abstract: In order to solve the problem of fast detection of adulteration of mutton and pork, the spectral collection of adulterated mutton was carried out by using a multi-spectral instrument, and the reflectivity of samples at the band of 350~1 100 nm was obtained. For data preprocessing, Particle Swarm Optimization (PSO) was used to optimize the Least Squares Support Vector

Machine (LSSVM), and a Least Squares Support Vector Machine (PSO-LSSVM) model based on Particle Swarm Optimization was established, compared with Partial Least Squares (PLS), Back Propagation Neural Network (BPNN) and LSSVM models. The result showed that PSO algorithm could effectively optimize LSSVM model, and the decision coefficient and root mean square error of prediction was 0.920 4 and 0.089 2. Furthermore, Random frog (RF), Uninformative Variable Elimination (UVE) and Competitive Adaptive Reweighted Sampling (CARS) were used to extract the characteristic wavelengths and establishing the model of PLS. The results showed that the UVE-PLS model's decision coefficient and the root mean square error of prediction set were 0.996 7 and 0.016 2, and UVE was better than other feature wavelengths extraction methods.

Keywords: mutton; adulteration; multispectral; particle swarm optimization; least Squares support vector machine; characteristic wavelengths

羊肉肉质细腻鲜嫩,蛋白质含量高,较其他肉类的脂肪、胆固醇含量低^[1],具有丰富的营养价值,不仅温脾养胃,还有补肝、益血、明目的功效,但由于羊肉价格较高,加工企业多且散乱难以管理,市场上一些不法商贩向羊肉掺入其他动物肉类以谋求暴利,严重损害消费者的利益甚至健康。传统用于检测掺假的方法大多是基于化学或物理的方法,例如酶联免疫吸附法(Enzyme Linked Immunosorbent Assay, ELISA)、聚合酶链反应(Polymerase Chain Reaction, PCR)和电子鼻技术等,但是这些方法操作复杂、灵敏度不高,甚至会破坏样品的完整性^[2]。

基金项目:河北省现代农业产业技术体系产业创新团队品牌与产品加工岗位项目(编号:HBCT2018140203)

作者简介:成甜甜,女,河北农业大学在读硕士研究生。

通信作者:韩宪忠(1965—),男,河北农业大学教授,硕士。

E-mail: 13832252366@126.com

收稿日期:2020-05-20

光谱仪可以获取物体的光谱数据信息,如光谱吸收、透射率、反射率、颜色等。将光谱数据与计算机技术结合,可以对肉类进行无损、快速检测。王飞翔^[3]建立偏最小二乘(PLS)模型,对调理肉在熟化过程中的水分含量进行预测,开发了基于多光谱成像技术的调理肉在线检测系统。范卉^[4]应用多光谱技术结合光学层析分析技术,对芝麻油、菜籽油、花生油、调和油、猪油、餐厨废弃油和花生煎炸油 7 类油进行甄别,为食用油和地沟油的鉴别提供了可靠方法。郝广等^[5-6]采用主成分分析法(Principal Component Analysis, PCA)、偏最小二乘法、最小二乘支持向量机(LSSVM)和反向传播神经网络(BPNN)建立定量模型,实现了多光谱成像技术在番茄酱中蔗糖掺假的快速检测。Ropodi 等^[7]针对肉类掺假的问题,提出利用多光谱成像技术对掺假牛肉进行检测,并建立了判别模型将所有样品正确分类。刘友华等^[8]针对羊肉掺假浓度检测问题,采用竞争性自适应重加权法(CARS)法建立特征波长预测模型,模型的预测集决定系数为 0.940 0,均方根误差为 0.076 6。白京等^[9]采用竞争性自适应重加权法提取特征波长并建立偏最小二乘模型,测定羊肉卷中猪肉掺假比例,其测试集的决定系数为 0.972 5,均方根误差为 0.057 7,相较于全波长模型结果有所提升。此外,还有一些学者^[10-13]对肉类的含水率和掺假量进行了预测。

目前中国利用多光谱成像技术对肉类掺假检测的研究较少,多数为对算法模型的简单应用,且未对模型进行改进以提高其预测能力,预测结果的准确度不高。为提高掺假肉检测准确度,试验拟将羊肉掺入不同比例的猪肉,利用多光谱系统提取样品表面的反射率,将掺假羊肉的反射率与羊肉掺入猪肉的浓度建立定量预测模型,对羊肉中猪肉的掺假浓度预测,并对模型进行优化,提升模型预测能力。进一步提取特征波长,简化模型。为掺假羊肉的快速无损识别提供切实可行的方法。

1 材料与方法

1.1 仪器

试验所用多光谱检测系统由电脑、光谱仪、可交换狭缝、WS-1 漫反射标准白板、海洋光学配套软件 Ocean View、光纤和探头等装置组成。光谱仪(上海蔚海光学仪器有限公司)型号为海洋光学(Ocean Optics)的 USB2000+,波长范围在 350~1 100 nm,共有 2 048 个光谱像素数。

1.2 样品制备

于保定市场购买的新鲜生肉,选取羊肉里脊和猪肉里脊各 1 kg。将羊肉和猪肉中的肥肉去掉,使用榨汁机将两种肉分别搅碎 1 min,直至羊肉和猪肉呈肉糜状,再分别按照 $m_{\text{羊肉}} : m_{\text{猪肉}}$ 分别为 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1 均匀混合并分装在玻璃培养皿

中,另外准备一份纯羊肉和一份纯猪肉的样本,共 11 份试验样本,每个样本 30 g。

1.3 多光谱数据采集与异常数据剔除

光谱仪的积分时间设置为 2.85 s,扫描次数为 100 次,滑动平均宽度设为 3,试验环境温度为 20 ℃。采集光谱数据时,数据结果容易受到外界光源或环境的干扰而产生噪声,令试验结果产生误差,使得模型效果变差,因此为系统设计了一个暗箱,保证环境无可见光的干扰,并且在采集光谱数据前,需要使用海洋光学光谱设备配套的 WS-1 漫反射标准白板进行白板校正,以此来减弱外界环境对数据的影响。采集光谱数据时,将探头垂直置于距样本 1 cm 的位置,对样本进行多次扫描。得到波长范围 350~1 100 nm 的反射光谱数据。

对采集到的原始光谱数据进行筛选和归一化处理,剔除明显偏高或偏低的异常样本,尽量选取反射率无交叉,有明显区分度的区域作为分析数据。选出可用波段后,将试验数据分为两部分,训练集用于建立判别模型,测试集用于检验模型的准确度。

1.4 基于粒子群优化的最小二乘支持向量机模型(PSO-LSSVM)

最小二乘支持向量机方法结合粒子群算法,利用 PSO 对 LSSVM 的两个参数进行最优搜索,建立了一种用于羊肉掺假定量检测的粒子群优化最小二乘支持向量机的模型,并将该优化模型的预测结果与现有模型的预测结果进行对比分析。

1.4.1 最小二乘支持向量机 最小二乘支持向量机将支持向量机(SVM)优化问题的不等式约束替换为等式约束,对于给定的 m 组样本 $(\mathbf{x}_i, \mathbf{y}_i)$,样本数 $i=1, 2, \dots, m$, $\mathbf{x}_i$ 为 n 维输入向量, $\mathbf{y}_i$ 为输出向量。LSSVM 在回归时用超平面对 m 组数据进行拟合。

$$\min_{\mathbf{w}, \mathbf{b}, e} J(\mathbf{w}, e) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^m e_i^2, \quad (1)$$

约束条件为:

$$\mathbf{y}_i = \mathbf{w}^T \varphi(\mathbf{x}_i) + \mathbf{b} + e_i, \quad (2)$$

式中:

$\mathbf{w}$ ——超平面权重向量;

$\mathbf{b}$ ——超平面偏差向量;

e ——训练点的误差;

γ ——惩罚系数。

式(2)中的 γ 越高代表对误差的容忍度越小, γ 越低代表对误差的容忍度越大,合理取值可以提高模型的预测能力。

构造拉格朗日(Lagrange)函数求解:

$$L(\mathbf{w}, \mathbf{b}, e, a) = J(\mathbf{w}, e) - \sum_{i=1}^m a_i \{ \mathbf{w}^T \varphi(\mathbf{x}_i) - \mathbf{b} + e_i - \mathbf{y}_i \}, \quad (3)$$

式中:

a_i —— x_i 对应的拉格朗日乘子。

求解过程中引入高斯核函数 $K(x, x_i)$:

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (4)$$

高斯核函数中 σ 决定数据映射到新的特征空间后的分布, σ 越大支持向量越少, σ 越小支持向量越多。支持向量的个数与预测的速度有关。

推导最终得到 LSSVM 回归函数:

$$y(x) = \sum_{i=1}^m a_i K(x, x_i) + b \quad (5)$$

1.4.2 粒子群优化算法 粒子群算法是通过模拟鸟类飞行觅食而设计出的一种群体智能优化算法。已知在一个区域内有一块食物, 鸟群知道当前位置离食物还有多远, 找到食物最简单有效的方法就是搜索离食物最近的鸟的周围区域。采用粒子群算法优化最小二乘支持向量机的两个参数 γ 和 σ , 在问题中, 每一个解都是空间中的一只鸟, 称为粒子, 代表两个参数的不同组合, 食物代表最优的参数组合, 粒子通过迭代搜索调整自己的位置和速度寻找到最优解。

(1) 对粒子群中粒子 i 的位置 $z_i = (\gamma_i, \sigma_i)$ 和速度 v_i 随机初始化, 生成大小为 n 的粒子种群。

(2) 将每个粒子代入 LSSVM 模型对训练集数据拟合, 得到模型的预测值 $f(x_i)$ 与期望输出 y_i , 训练结果的均方根误差决定每个粒子的适应度值 ($fitness$)。

$$fitness = \sqrt{\frac{\sum_{i=1}^m [f(x_i) - y_i]^2}{i}} \quad (6)$$

(3) 每个微粒根据适应度值更新自己的个体最优值 ($pbest_i$) 和群体最优值 ($gbest_i$)。

$$pbest_i = (pbest_{i1}, pbest_{i2}, \dots, pbest_{in}), \quad (7)$$

$$gbest_i = (gbest_{i1}, gbest_{i2}, \dots, gbest_{in}) \quad (8)$$

(4) 根据适应度值对粒子的速度和位置进行更新。

$$v_i = v_i + c_1 \times rand() \times (gbest_i - z_i) + c_2 \times rand() \times (gbest_i - z_i), \quad (9)$$

$$z_i = z_i + v_i, \quad (10)$$

式中:

c_1, c_2 ——学习因子。

在粒子群算法中, 学习因子的取值一般为 2, $rand()$ 为 0 和 1 之间的随机数。

(5) 通过终止条件判断是否结束迭代, 得到粒子的最优位置。

1.5 特征波长提取

提取特征波长不仅可以简化模型, 还能剔除无关变量, 提升模型性能和预测能力, 增强稳定性。试验分别采用随机青蛙算 (RF)^[14]、无信息变量消除法 (UVE)^[15-16]、竞争性自适应重加权法^[17] 提取特征波长, 以提取出的波长作为输入变量建立偏最小二乘特征波长模型, 对比预测结果, 得到最优的特征提取算法。

2 结果与讨论

2.1 原始光谱

利用多光谱检测系统对样本提取反射率, 图 1 为样本在 350~1100 nm 波段下的反射率。由图 1 可知, 相同波段下不同掺假比例的样本反射率走势相同, 在某些波段下有明显区分。样本中有一条数据反射率明显偏低, 考虑是由于试验误操作引起, 可以剔除。数据两侧的噪声较多不平滑, 走势密集不易区分, 不宜选用, 故选取波段中间 500~650 nm 下 427 个波长点的反射率作为可用数据。由于数据的量纲不同, 并且数据尺度不统一时对预测模型的结果影响很大, 故需要对光谱数据作归一化处理, 将数据映射在 -1 和 1 之间, 结果如图 2 所示。

2.2 全波长模型

选取了可用波段的数据后, 将 32 组数据按照 2:1 分为训练集和测试集, 有 21 组训练集数据 $(x_i, y_i) (i=1, 2, \dots, 21)$, x_i 为 427 维输入向量, 代表 427 个波点数, y_i 为羊肉掺入猪肉的浓度。

利用粒子群优化算法, 对最小二乘支持向量机的两个参数 γ 和 σ 进行寻优, 初始化粒子群的种群大小、学习因子、位置、速度、搜索范围和迭代次数。将粒子 i 的位置 $z_i = (\gamma_i, \sigma_i)$ 代入 LSSVM 模型对训练集数据进行拟合, 模型预测结果的均方根误差作为粒子群算法的适应度

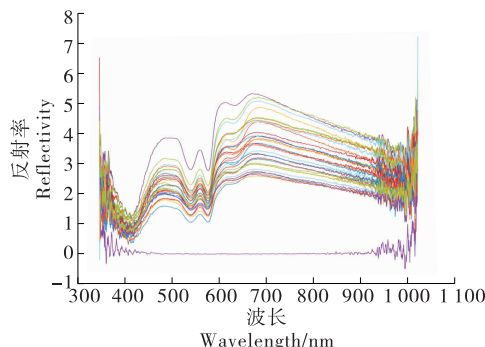


图 1 原始光谱图像

Figure 1 The image of original spectral

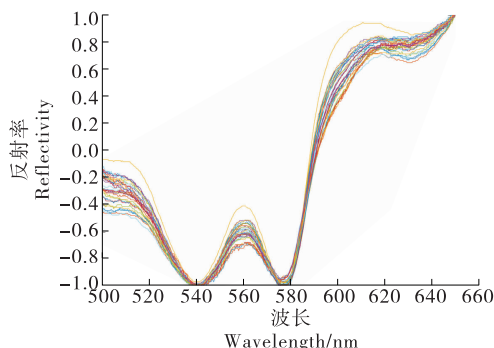


图 2 归一化后的光谱图像

Figure 2 The image of normalized spectral

值,每个粒子根据自身适应度值,得到 $pbest$ 和 $gbest$,计算更新粒子的速度 v_i 和 z_i ,直到迭代结束得到全局最优的位置,即为粒子群算法优化所得两参数 γ 和 σ 。

将经过 PSO 算法优化的 LSSVM 模型与反向传播神经网络、偏最小二乘回归、最小二乘支持向量机的模型结果作对比。模型输出结果如图 3 所示,BPNN 预测输出结果较差,LSSVM 较好。模型训练集的决定系数(R_c^2)和均方根误差(RMSEC)、预测集的决定系数(R_p^2)和均方根误差(RMSEP)如表 1 所示,结果显示,BP 神经网络的预测结果整体较差,偏最小二乘模型对数据拟合能力较强,但预测结果不如最小二乘支持向量机模型,在经过粒子群算法优化后 LSSVM 模型对数据的拟合和预测能力显著提升,训练集的 $R_c^2=1.000\ 0$,RMSEC=0.001 2,预测集的 $R_p^2=0.920\ 4$,RMSEP=0.089 2。

2.3 特征波长提取和特征波长模型

2.3.1 随机青蛙 采用随机青蛙算法在对 500~650 nm 波段下的 32 组数据提取特征波长,为减少算法中随机因素的影响,将算法运行 1 000 次并以选择概率平均值作为波长选择的依据,结果如图 4。选择概率越大说明该变量对模型越重要,由图 4 可知,只有小部分的波长选择概率较大,最终选出前 10 个概率最大的波长作为特征波长,分别为 588.944,560.757,618.225,639.282,512.897,620.650,524.040,536.212,500.265,621.689 nm。

2.3.2 无信息变量消除法 UVE 将变量回归系数和标准偏差的比作为评判变量稳定性的值,稳定性绝对值越

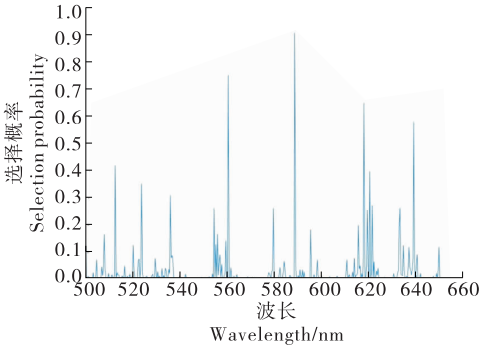


图 4 随机青蛙平均概率

Figure 4 The average probability of a random frog

大,证明该变量的可靠性越高。对所有波长点计算稳定性后结果见图 5。选择稳定性大于 4 的 17 条波长:504.963,536.927,578.408,579.815,582.979,583.330,584.032,584.383,584.735,611.630,612.325,612.673,613.020,513.367,615.450,615.797,633.087 nm。

2.3.3 竞争性自适应重加权法 通过自适应重加权采样法去掉 PLS 模型中回归系数绝对值权重较小的波长点,得到的结果见图 6,从 427 个波长点中提取出了 20 个波长点,分别为 500.265,513.977,523.681,557.213,560.757,568.184,584.383,588.944,596.644,606.064,

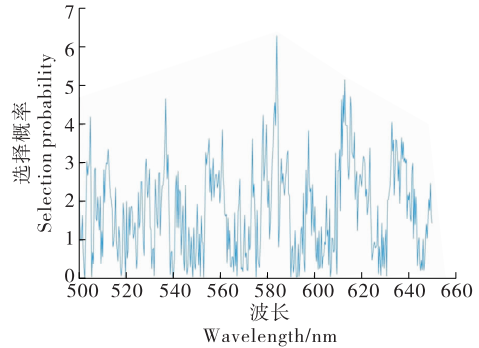


图 5 UVE 对样品稳定性值的计算结果

Figure 5 The calculation results of UVE for the stability value of the sample

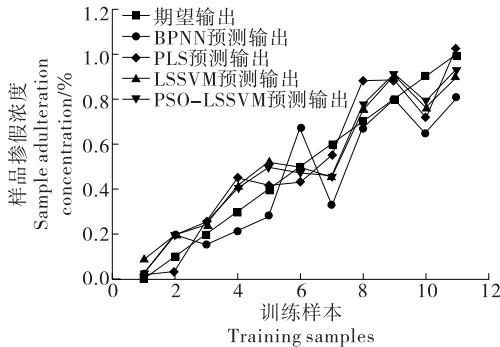


图 3 模型预测输出

Figure 3 The outputs of model prediction

表 1 不同模型的预测效果

Table 1 Prediction effects of different models

建模方法	R_c^2	RMSEC	R_p^2	RMSEP
PLS	0.994 2	0.024 7	0.899 0	0.100 5
BPNN	0.829 4	0.133 7	0.743 6	0.160 1
LSSVM	0.980 0	0.045 8	0.900 3	0.099 9
PSO-LSSVM	1.000 0	0.001 2	0.920 4	0.089 2

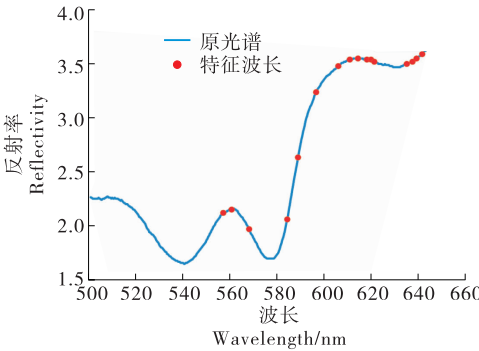


图 6 CARS 特征波长分布

Figure 6 The characteristic wavelengths of CARS

610.935, 614.409, 618.225, 619.957, 621.342, 635.153, 637.563, 639.282, 641.687, 649.574 nm。

2.3.4 建立特征波长模型 对 3 种波长选择方法提取出的特征波长建立偏最小二乘模型, 预测结果如表 2 所示, 对比全波长模型分析可知, 在提取特征波长后模型的决定系数均有所提升。3 种特征波长提取方法中, UVE 提取的波长所建模型预测效果更好, 误差最小, 训练集的 $R_c^2=0.999\ 1$, RMSEC=0.011 9, 预测集的 $R_p^2=0.996\ 7$, RMSEP=0.016 2。

表 2 不同特征波长提取方法下的模型预测效果

Table 2 The prediction effects of different feature wavelengths extraction methods

建模方法	R_c^2	RMSEC	R_p^2	RMSEP
RF-PLS	0.976 8	0.049 3	0.939 9	0.077 5
UVE-PLS	0.999 1	0.011 9	0.996 7	0.016 2
CARS-PLS	0.998 3	0.013 4	0.987 6	0.035 2

3 结论

应用多光谱图像技术获取羊肉和掺假羊肉在 350~1 100 nm 波段下的反射率, 建立偏最小二乘、BP 神经网络、最小二乘支持向量机 3 种定量判别模型, 并通过粒子群算法对最小二乘支持向量机的两个参数进行优化后建立模型, 通过比较预测结果可知, 最小二乘支持向量机在 3 种常用模型中的预测结果最优, 使用粒子群算法优化最小二乘支持向量机后, 模型预测效果显著提升。采用随机青蛙、无信息变量消除法、竞争性自适应重加权法对 500~650 nm 波段下的数据提取特征波长后建立偏最小二乘模型, 结果显示模型预测效果整体提升, 其中基于无信息变量消除法提取特征波长建立的模型预测结果最好。

试验还需进一步完善, 扩充样本数量, 对羊肉不同部位的肉作进一步研究和区分, 增加不同种类动物的肉与羊肉掺杂, 扩大模型的应用范围。

参考文献

- [1] 赵从民, 宋桂芳. 吃羊肉与保健[J]. 肉品卫生, 2001(5): 20.
- [2] 施姿鹤, VOGLMEIR Josef, 刘丽. 肉及其加工制品的掺假鉴别技术研究进展[J]. 食品科学, 2019, 40(23): 319-326.
- [3] 王飞翔. 调理肉多光谱图像在线检测设备的设计与试验[D]. 洛阳: 河南科技大学, 2019: 33-43.
- [4] 范卉. 基于多光谱技术与光学层析方法的地沟油特征研究[D]. 西安: 西安工业大学, 2017: 37-62.
- [5] 郝广. 基于多光谱成像技术的番茄酱品质与安全快速检测研究[D]. 合肥: 合肥工业大学, 2018: 29-34.
- [6] LIU Chang-hong, HAO Guang, SU Min. Potential of multi-spectral imaging combined with chemometric methods for rapid detection of sucrose adulteration in tomato paste[J]. Journal of Food Engineering, 2017, 215: 78-83.

- [7] ROPODI A I, PAVLIDIS D E, MOHAREB F. Multispectral image analysis approach to detect adulteration of beef and pork in raw meats[J]. Food Research International, 2015, 67: 12-18.
- [8] 刘友华, 白亚斌, 邱祝福, 等. 基于高光谱图像技术和波长选择方法的羊肉掺假检测方法研究[J]. 海南师范大学学报(自然科学版), 2015, 28(3): 265-269.
- [9] 白京, 李家鹏, 邹昊, 等. 近红外特征光谱定量检测羊肉卷中猪肉掺假比例[J]. 食品科学, 2019, 40(2): 287-292.
- [10] LIU Jin-xia, CAO Yue, WANG Qiu. Rapid and non-destructive identification of water-injected beef samples using multispectral imaging analysis[J]. Food Chemistry, 2016, 190: 938-943.
- [11] 刘锦霞. 基于多光谱成像技术快速检测注水肉及冷冻肉的品质安全[D]. 合肥: 合肥工业大学, 2016: 21-28.
- [12] 王迪, 张敬埔, 张珏, 等. 基于高光谱技术的羊肉含水率无损检测[J]. 食品工业科技, 2018, 39(4): 215-218.
- [13] 邢素霞, 王九清, 陈思, 等. 基于 K-means-RBF 的鸡肉品质分类方法研究[J]. 食品科学技术学报, 2018, 36(4): 93-99.
- [14] 郑剑, 周竹, 仲山民, 等. 基于近红外光谱与随机青蛙算法的褐变板栗识别[J]. 浙江农林大学学报, 2016, 33(2): 322-329.
- [15] 于慧春, 付晓雅, 殷勇, 等. 拉曼光谱结合 UVE-SVR 算法预测加热食用油反式脂肪酸的含量[J]. 核农学报, 2020, 34(3): 582-591.
- [16] 王怡森, 朱金林, 张慧, 等. 基于 MC-UVE/GA 算法及因子分析对葡萄酒酒精度近红外定量模型的优化研究[J]. 发光学报, 2018, 39(9): 1 310-1 316.
- [17] 王海龙, 杨国国, 张瑜, 等. 竞争性自适应重加权算法和相关系数法提取特征波长检测番茄叶片真菌病害[J]. 光谱学与光谱分析, 2017, 37(7): 2 115-2 119.

(上接第 3 页)

复合消毒剂的杀菌作用、食品生产方式的溯源分析等内容展开讨论。

食品化学与生物技术的飞速发展延长食品贮藏时间、优化食品风味质构、提高食品理化性能等提供了新手段。在食品化学与生物技术及其应用专题中, 报告人围绕食品的分子结构及互作机制, 从多方面、多角度、多层次探讨食品化学与生物技术创新对食品工业的推动作用, 并就混合鱼糜凝胶协同增效作用、食品热风干制过程中的脂质氧化等应用研究进行了分享。

年会上, 还举行了年会会旗的交接仪式, 中国食品科学技术学会第十八届年会将由天津科技大学、天津市食品学会联合津京冀共 12 家单位共同承办和协办, 于 2021 年在天津召开。

2021——相会于天津!

(中国食品科学技术学会秘书处 供稿)