

基于分类特征提取和深度学习的牛肉品质识别

Beef quality recognition based on classification feature extraction and deep learning

王新龙¹ 李翔²

WANG Xin-long¹ LI Xiang²

(1. 长治学院, 山西 长治 046011; 2. 陕西科技大学, 陕西 西安 710021)

(1. Changzhi College, Changzhi, Shanxi 046011, China;

2. Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China)

摘要:目的:降低数据差异性和光谱特征冗余度对牛肉品质识别的影响。方法:提出一种基于分类特征提取和深度学习的牛肉品质识别方法,采用改进的 DPeak 算法对光谱数据进行自适应聚类分析,实现对数据的差异性分析。定义牛肉光谱特征提取目标函数,采用离散狮群算法进行求解,提取每个分类的最佳光谱特征子集,最大限度降低特征冗余度。运用改进狮群算法(ILSO)对每个分类对应的支持向量机(SVM)模型参数进行优化,提出融合分类特征提取和 ILISO 优化 SVM 的牛肉品质识别模型,完成对牛肉品质的分类识别。结果:相比于 SSA-SVM、PCA-SVM 识别模型,该模型识别精度提高了约 12.3%~14.5%。结论:基于分类特征提取和深度学习的牛肉品质识别模型能够提高牛肉品质识别精度。

关键词:牛肉;特征;近红外光谱;狮群算法;支持向量机;品质识别

Abstract: Objective: To reduce the influence of data difference and spectral feature redundancy on beef quality recognition. **Methods:** A beef quality recognition method based on classification feature extraction and deep learning was proposed. The spectral feature extraction method of classified beef was designed, and the improved DPeak algorithm was used for adaptive clustering analysis of spectral data to realize the difference analysis of data. The objective function of beef spectral feature extraction was defined and solved by discrete lion swarm algorithm. The optimal spectral feature subset of each classification was extracted to minimize feature redundancy. The improved lion swarm algorithm (ILSO) was used to optimize the support vector

machine (SVM) model parameters corresponding to each classification, and a beef quality recognition model integrating classification feature extraction and ILISO optimized SVM was proposed to complete the classification and evaluation of beef quality. **Results:** Compared with SSA-SVM and PCA-SVM, the recognition accuracy of this model is improved about 12.3%~14.5%. **Conclusion:** The beef quality recognition model based on classification feature extraction and deep learning can improve the accuracy of beef quality recognition.

Keywords: beef; features; near infrared spectroscopy; Lion Swarm Optimization; support vector machines; quality identification

受屠宰方式、储存手段等因素影响,牛肉品质可以划分为正常肉、白肌肉、黑干肉和掺假肉等类型^[1-5],人工感官识别是目前常见的识别手段,其主观性强、劳动强度大、结果可信度低^[6]。随着近红外光谱技术的快速发展,光谱技术应用于品质识别与鉴定已成为热点研究领域之一。由于光谱中包含了大量被测物品信息,如何高效地对光谱特征进行提取筛选以实现高精度识别是当前亟需解决的难题。张保霞^[7]利用主成分分析法对猪肉光谱进行降维处理,提取出猪肉光谱特征,并采用支持向量机进行品质辨识,该方法识别精度可以达到 92.31% 以上,但是,该方法提取的特征能否保持原始数据分类能力有待进一步研究。吴叶兰等^[8]采用主成分分析法对高光谱成像特征进行筛选,并利用随机森林模型进行识别,有效分辨出了柑橘病虫害叶,但是主成分分析法对噪声鲁棒性不强,误差较大。孔德明等^[9]面对海量光谱数据,采用稀疏主成分分析法进行特征提取,得到的分类准确率比主成分分析法提高了 5%,但是该方法采用网格搜索法配置支持向量机模型参数,容易陷入局部最优,影响了识别精度。

基金项目:教育部产学研合作协同育人项目(编号:202002222004);
山西省大学生创新创业训练计划项目(编号:2018599)

作者简介:王新龙(1964—),男,长治学院副教授。

E-mail: czjxwzl@126.com

收稿日期:2021-12-25

牛肉品质识别过程需要进行大量数据样本训练,而以往研究往往忽略了数据差异性对识别精度的影响,尤其是当训练样本存在较多异常点时,得到的识别结果是不可靠的。为了降低数据差异性以及光谱特征冗余度对牛肉品质识别的影响,研究拟采用 DPeak^[10] 对牛肉光谱进行分类,使得同类内数据有更多的相似性,不同类数据具有更多的差异性;设计分类光谱特征提取机制,尽可能降低数据维度,并最大限度保持原始数据辨识能力;运用狮群算法^[11-12] 对分类支持向量机(SVM)模型参数进行优化,得到最优模型参数配置;并建立融合分类特征提取和 SVM 的牛肉品质识别模型,以提高牛肉品质分类识别精度。

1 材料与方法

1.1 材料与设备

正常肉、白肌肉、黑干肉试验样本:取生鲜牛肉股二头肌、左右侧背最长肌,每块样品去除结缔组织,切割厚度 30 mm,并取每块样品的局部进行化学分析以确定所属品质,市售;

掺假牛肉样本:指定合成牛肉生产商,按 5%,10%,15%,20%,25% 比例在牛肉中掺加鸭肉,并于 -30 ℃ 冷冻,冷藏解冻 24 h 后制成掺假牛肉样品。

1.2 试验方法

所有样品统一标签备注,按正常肉:白肌肉:黑干肉:掺假肉为 1:1:1:1 的比例选取 300 个样本组成训

练样本集,随机选取 4 类牛肉 80 个组成测试集。结合化学分析的 pH 值、乳酸含量、肌糖原、掺假比例等指标^[13],建立牛肉品质指数与品质类别对应表,并采用专家判定法^[14] 按表 1 对每个训练样本、测试样本品质指数进行赋值。

表 1 牛肉品质指标与品质类别对应表

Table 1 Corresponding table of beef quality index and quality category

品质指数	[0.0,0.3)	[0.3,0.6)	[0.6,0.8)	[0.8,1.0]
品质类型	1 级 正常肉	2 级 白肌肉	3 级 黑干肉	4 级 掺假肉

利用近红外光谱仪采集样本全波段(FS)数据,得到训练样本 FS 数据集合和测试样本 FS 数据集合。对于训练样本 FS 数据集合,设计改进的 DPeak 算法(IDPeak)进行聚类分析,得到多个 FS 分类;定义分类牛肉光谱特征提取目标函数,采用改进 LSO(ILSO)进行求解,提取到每个分类最佳光谱特征子集;利用改进 ILISO 对分类 SVM 参数进行优化,建立融合分类特征提取和 SVM 的牛肉品质识别模型。对于测试样本 FS 数据集合,找到每个待测试样本光谱数据对应的分类,利用该分类识别模型提取特征参数,并进行牛肉品质识别,最终得到识别结果,牛肉品质识别方法示意图见图 1。

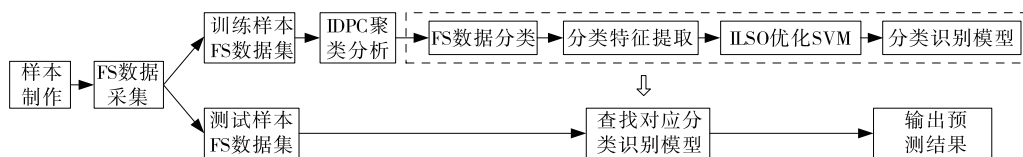


图 1 牛肉品质识别方法示意图

Figure 1 Schematic diagram of beef quality identification method

2 牛肉品质识别

2.1 改进狮群算法

狮群算法(LSO)设置狮王、母狮、幼狮 3 种类型个体,分别执行不同的更新进化策略,其算法核心在于平衡局部搜索与全局搜索,具有较好的收敛能力^[15-16]。LSO 缺陷主要有:幼狮等概率选取狮王、母狮、反向学习进行进化,导致算法初期不利于大范围搜索,算法后期影响了收敛速度;公狮、母狮只选择向自身历史最优解和当前公狮、母狮群体最优解进行学习,容易陷入局部最优;个体进化控制系数随机选取,不利于保持种群多样性。为此提出 ILISO,自适应调整幼狮不同学习进化的个体规模,增加狮王、母狮学习对象种类,设计改进的进化控制系数,并以概率的形式接受不好的解。

2.1.1 自适应学习 以最小值优化问题为例(目标优化函数为 f),LSO 种群规模为 Q 、最大迭代次数 $T_{\max}$ 、幼狮

个体数量为 N_c 。 t 时刻,种群最优解、最差解为 $X_b(t)$ 、 $X_w(t)$,采取向狮王、母狮和反向进化学的幼狮个体数量分别为 $N_{c,1}$ 、 $N_{c,2}$ 、 $N_{c,3}$ 。

$$N_{c,1} = N_{c,1}^{\min} \times \gamma_1 \times \frac{T_{\max}}{t + \epsilon}, \quad (1)$$

$$N_{c,3} = N_{c,3}^{\min} + \gamma_2 (N_{c,3}^{\max} - N_{c,3}^{\min}) \left\{ \exp \left(\frac{f[X_w(t)]}{f[X_b(t)] + \epsilon} \ln 2 \right) - 1 \right\}, \quad (2)$$

$$N_{c,2} = N_c - N_{c,1} - N_{c,2}, \quad (3)$$

式中:

$N_{c,1}^{\min}$ 、 $N_{c,3}^{\min}$ 、 $N_{c,3}^{\max}$ ——个体数量边界;

γ_1 、 γ_2 ——个体数量控制系数;

ϵ ——极小常数。

由式(1)~式(3)可知,算法初期,幼狮更多的个体向具有更优目标函数值的狮王位置进行移动,加速了算法收敛速度;随着迭代次数的增加,种群整体趋向全局最优

解,此时,幼狮更多的个体执行反向进化操作,提高了算法跳出局部极值的概率和收敛精度。

2.1.2 学习对象扩展与进化控制 t 时刻,对于 $X_{L,i}(t)$ 、母狮 $X_{M,j}(t)$,引入种群其他个体信息以扩展搜索空间:

$$X_{L,i}(t) = \omega X_b(t) + \text{rand}(-1, 1) [p_{L,i} - X_{L,i}(t)] + \text{rand}(-1, 1) [X_{L,i}(t) - X_{L,a}(t)], \quad (4)$$

$$X_{M,j}(t) = \omega X_b(t) + \text{rand}(-1, 1) [X_{M,j}(t) - X_{M,b}(t)] + \text{rand}(-1, 1) [X_{M,j}(t) - X_{M,c}(t)], \quad (5)$$

$$\omega = \begin{cases} \omega_{\min} - \frac{(\omega_{\max} - \omega_{\min})\{f - f[X_b(t)]\}}{f_{\text{avg}} - f[X_b(t)]} & f \leq f_{\text{avg}} \\ \omega_{\max} & f > f_{\text{avg}} \end{cases}, \quad (6)$$

式中:

$\omega(\omega_{\min}, \omega_{\max})$ ——进化控制系数(边界);

$p_{L,i} - X_{L,i}(t)$ ——历史最优解;

$X_{L,a}(t) [X_{M,b}(t), X_{M,c}(t)]$ ——公狮(母狮)其他个体;

f, f_{avg} ——LSO 当前个体目标函数值、种群目标函数平均值。

由式(4)~式(6)可知,狮王(母狮)迭代进化中引入种群最优解和其他公狮(母狮)个体信息,并根据自身目标函数值动态调整向种群最优解学习进化程度,有效扩展了种群搜索空间,提高了算法收敛精度。

2.1.3 概率进化 当狮群内个体 $X_i(t)$ 执行完设定的更新策略后得到新的个体 $X_{i,\text{new}}(t)$,若 $X_{i,\text{new}}(t)$ 目标函数值优于 $X_i(t)$,则用 $X_{i,\text{new}}(t)$ 替代 $X_i(t)$;否则,以概率 p 接受 $X_{i,\text{new}}(t)$:

$$p = \exp\left\{-\frac{f[X_i(t)] - f[X_{i,\text{new}}(t)]}{\alpha}\right\}, \quad (7)$$

式中:

α ——比例系数。

根据马尔科夫链理论^[17],采用概率的形式接受不好的解,使得 ILSO 算法能够以概率 1 收敛于全局最优解。从 ILSO 算法实现可知,对于 D 维优化问题,算法种群初始化复杂度 $O(Q)$,算法迭代一次复杂度为 $O(Q \lg D)$,总计算复杂度为 $T_{\max} O(Q \lg D) + O(Q) \approx T_{\max} O(Q \lg D)$ 。

2.2 IDPeak 聚类分析

增加训练样本规模能改善模型训练效果,有利于提高识别精度,但是,训练样本规模增加的同时,也带入了大量噪声、孤立点等信息,若不妥善处理数据样本差异性,得到的识别结果往往是不可靠的。为此,引入改进 DPeak 算法(IDPeak)对训练样本进行聚类分析,降低数据差异性对识别精度的影响。DPeak 算法作为一种粒度计算模型,具有参数简单、鲁棒性强等特点,对于大部分数据类型都有着很好的适应性^[18]。对于牛肉光谱训练样本数据集 $\Theta = \{S_i\}_{i=1,\dots,N}$,每个样本 X_i 由 n 个波段组

成 $S_i = (s_{i1}, \dots, s_{in})$,设 S_i 到 S_j 的欧式距离为 d_{ij} ,DPeak 定义局部密度 ρ_i 、最近点距离 δ_i 、分类判定参数 γ_i :

$$\rho_i = \sum_j \chi(d_{ij} - d_c), \chi(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases}, \delta_i = \min_{j: \rho_j > \rho_i} d_{ij}, \gamma_i = \rho_i \delta_i, \quad (8)$$

式中:

d_c ——截断距离(DPeak 算法唯一设定参数), m 。

DPeak 基于 $\rho_i, \delta_i, \gamma_i$ 建立决策图,将数据点划分为离群点、密度峰值点和正常点,并选取决策图右上方的点为聚类(分类)中心。对于多孤立点复杂聚类问题,DPeak 聚类效果不佳,且 d_c 的取值大小直接影响了 $\rho_i, \delta_i, \gamma_i$,为此,设计改进 DPeak(IDPeak)算法,采用核距离 D_{ij} 对 S_i 到 S_j 的距离进行描述,并利用 ILSO 求解最佳截断距离:

$$D_{ij} = \|\Phi(S_j) - \Phi(S_i)\|^2, \quad (9)$$

式中:

$\Phi(\cdot)$ ——核函数(文中选取高斯函数)。

参照文献^[19]提出的邦费罗尼指数,定义聚类效果评价指标(CEEI):

$$C_{\text{EEI}} = \frac{N-1}{\sum_{i=1}^N \left\{ \left[i/N - \left(\sum_{j=1}^i \gamma_j \right) / \left(\sum_{j=1}^N \gamma_j \right) \right] / (i/N) \right\}}, \quad (10)$$

式中:

C_{EEI} ——聚类效果评价指标。

CEEI 取值越小,聚类效果越优^[19]。CEEI 涉及 σ, d_c 两个参数,采用 ILSO 对 CEEI 优化,将个体编码 X_i 等效为 $X_i = (\sigma_i, d_{c,i})$,目标函数为 $f(X) = \min C_{\text{EEI}}$ 。通过 ILSO 迭代进化,最终得到最佳截断距离 $d_{c,\max}$ 。图 2 为 IDPeak 对 4 个典型数据集的聚类结果,其与 FCM、DPeak 聚类正确率对比见表 2。由图 2 和表 2 可知,IDPeak 聚类结果更优。

2.3 分类特征提取

采用 IDPeak 对牛肉光谱训练数据 $\Theta = \{S_i\}_{i=1,\dots,N}$ 聚类分析,得到 C 个分类 $\Theta = \{C_{L,i}\}_{i=1,\dots,C}, C_{L,i} = \{S_{ij}\}_{j=1,\dots,C_i}, \sum_{i=1}^C C_i = N$ ($C_{L,i}$ 为第 i 个分类, C_i 为第 i 个分类内数据个数),对于分类 $C_{L,i}$ 内数据 $S_{ij} = (s_{j1}^i, \dots, s_{jn}^i)$,定义特征波段提取向量 $\mathbf{V} = (v_k)_{k=1,\dots,n}$,其中, v_k 取值“0”或“1”,当 $v_k = 1$ 时,表示 S_{ij} 内第 k 个波段 s_{jk}^i 被选取。定义全波段矩阵 $\mathbf{A}$ 、特征提取矩阵 $\mathbf{B}$ 和特征提取判定参数 T_{ZT} :

$$\mathbf{A} = [a_{ij}]_{n \times C_i} = [S_{i1} \ S_{i2} \ \dots \ S_{iC_i}] = \begin{bmatrix} s_{11}^i & s_{21}^i & \dots & s_{C_i 1}^i \\ s_{12}^i & s_{22}^i & \dots & s_{C_i 2}^i \\ \vdots & \vdots & \ddots & \vdots \\ s_{1n}^i & s_{2n}^i & \dots & s_{C_i n}^i \end{bmatrix}, \quad (11)$$

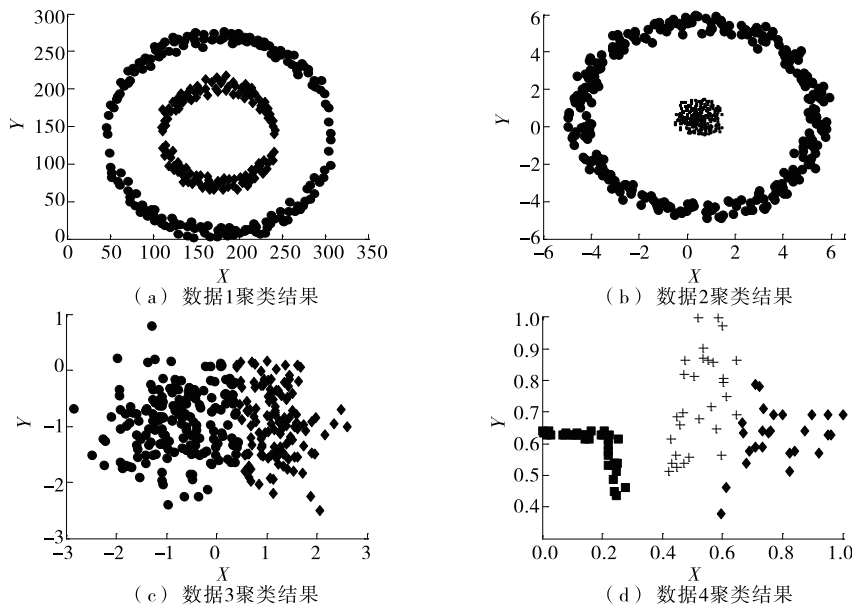


图 2 IDPeak 典型数据集聚类结果

Figure 2 Clustering results of typical IDPeak data sets

表 2 评价指标对比结果

Table 2 Comparison results of evaluation indexes

%

数据 1			数据 2			数据 3			数据 4		
IDPeak	DPeak	FCM	IDPeak	DPeak	FCM	IDPeak	DPeak	FCM	IDPeak	DPeak	FCM
100.00	91.23	80.14	98.82	87.55	76.21	99.11	86.24	72.36	95.41	85.24	66.47

$$\mathbf{B} = [\mathbf{V}^T \quad \mathbf{V}^T \quad \cdots \quad \mathbf{V}^T] = \begin{bmatrix} v_1 & v_1 & \cdots & v_1 \\ v_2 & v_2 & \cdots & v_2 \\ \vdots & \vdots & \vdots & \vdots \\ v_n & v_n & \cdots & v_n \end{bmatrix}, \quad (12)$$

$$T_{ZT} = \|\mathbf{H}^T(\mathbf{A}\mathbf{B})(\mathbf{A}\mathbf{B})^T\mathbf{H} - \mathbf{C}_i^2\mathbf{U}\|_F^2, \quad (13)$$

式中:

$\mathbf{H}$ ——类间相似度矩阵且为常数矩阵;

$\mathbf{U}$ ——相关性矩阵。

分类特征提取的目的是通过确定 $\mathbf{V}$ 表达形式,使得提取到的特征波段尽可能保持原始数据的分类能力,为此,利用 IDPeak 对 $C_{L,i}$ 进行聚类分析,以验证提取后波段特征的分类能力。式(13)中 $\mathbf{H}$ 反映了 IDPeak 分类后的类间相似程度, $\mathbf{U}$ 反映了 S_{ij} 与类的相关性,可以采用最大信息系数法^[20]求解。

当 T_{ZT} 取最小值时对应的 $\mathbf{V}$ 即为最佳特征波段提取向量 $\mathbf{V}_{\text{best}}$,为此,采用 ILSO 对 T_{ZT} 最小值问题进行优化,将个体编码 X_i 等效为 $X_i = \mathbf{V}$,目标函数为 $f(X) = \min T_{ZT}$ 。由于 T_{ZT} 最小值问题属于离散问题,对 ILSO 对应的式(4)、式(5)进化方式进行离散化理解,即 $\omega X_b(t)$ 理解为随机选取 $X_b(t)$ 内 ω 个编码位进行替代

操作, $\text{rand}(-1,1) [X_{M,j}(t) - X_{M,b}(t)]$ 理解为随机选取 $[\text{rand}(-1,1) \times m]$ 个编码位进行替代操作 [m 为 $X_{M,j}(t)$ 、 $X_{M,b}(t)$ 不同编码位的个数]。

通过 ILSO 迭代进化,最终得到 $C_{L,i}$ 对应的 $\mathbf{V}_{\text{best},i}$,利用 $\mathbf{V}_{\text{best},i}$ 对 S_{ij} 进行特征波段提取,得到最佳特征波段集合 $S'_{ij} = (s'_{ij1}, \dots, s'_{ijM_i})$ (M_i 为 $\mathbf{V}_{\text{best},i}$ 非零元素个数),此时,问题维度由 n 降到 M_i (一般的有 $M_i \ll n$),大大降低了特征冗余度。

2.4 分类识别模型

对于牛肉光谱训练数据 C 个分类,分别建立牛肉品质识别模型:以分类 $C_{L,i}$ 为例,采用一阶求导 (1st Der)^[8] 等数据预处理方法对 $C_{L,i}$ 进行预处理,利用 $\mathbf{V}_{\text{best},i}$ 进行特征波段提取,得到特征数据集 $C_{L',i} = \{S'_{ij}\}_{j=1, \dots, C_i}$, $S'_{ij} = (s'_{ij1}, \dots, s'_{ijM_i})$ 。识别模型采用支持向量机 (SVM),引入拉格朗日乘子后的 SVM 模型表达形式为:

$$y(\lambda, \theta) = \sum_{j \in \text{sv}} (\alpha_j - \alpha_j^*) K(S'_{ij}, S'_i) + b, \quad (14)$$

式中:

y ——模型输出;

$K(\cdot)$ ——核函数,参数为 θ ;

α_j, α_j^* ——拉格朗日乘子;

λ ——惩罚参数;

b ——超平面偏差。

SVM 需要对参数 λ 、 θ 进行配置^[8],为此采用 ILSO 求解最佳分类 SVM 参数,个体编码 X_i 等效为 $X_i = (\lambda, \theta)$,目标函数为 SVM 模型输出值与 $C_{L',i} = \{S'_{ij}\}_{j=1,\dots,C_i}$ 实际牛肉品质值的均方根误差:

$$f(X) : \min \frac{1}{W} \sqrt{\frac{1}{C_i} \sum_{j=1}^{C_i} (y_j - y'_j)^2}, \quad (15)$$

式中:

W ——训练次数;

y_j —— S'_{ij} 对应的 SVM 模型输出值;

y'_j —— S'_{ij} 对应的实际牛肉品质值。

ILSO 循环迭代进化,最终得到分类 $C_{L,i}$ 的 SVM 模型最佳参数组合 $(\lambda, \theta)_{\text{best},i}$ 。对于测试样本 FS 数据集 $\Theta_s = \{Z_i\}$,牛肉品质识别实现过程为:根据 Z_i 与 $\Theta = \{S_i\}_{i=1,\dots,N}$ 每个分类中心的距离判定其所属分类,采用 1st Der 等方法数据预处理后,利用该分类对应的最佳特征波段提取向量特征波段,将其作为分类 SVM 模型输入,通过 SVM 模型识别预测,最终完成 Z_i 牛肉品质识别。

3 结果与分析

采用 Thermo 公司生产的 Antaris II 型近红外光谱仪进行全波段扫描,光谱采集软件为 Avasoft7.4 版本,数据

处理软件为 MATLAB2017a,图 3 为试验方法示意图。

3.1 分类特征提取

采用 IDPeak 对训练数据集进行聚类分析,对每个分类提取特征波段,ILSO 参数设置: $Q=200$ 、 $T_{\max}=500$ 、 $N_{c,1}^{\min}=5$ 、 $N_{c,3}^{\min}=8$ 、 $N_{c,3}^{\max}=25$ 、 $\omega_{\min}=0.15$ 、 $\omega_{\max}=0.75$,表 3 为原始数据和经过 1st Der 处理后的数据分类结果、分类特征波段提取结果,图 4 为某个分类下,ILSO 优化 C_{EEI} 、 T_{ZT} 收敛曲线。表 4 为经过 1st Der 预处理后,训练样本数据集对应的每个分类样本数量以及样本所属牛肉品质类型情况。

由表 3 可知,IDPeak 算法将训练数据集划分为 7 个分类,对于每个分类,无论是直接进行特征提取,还是先经过 1st Der 预处理再进行特征提取,每个分类提取到的特征波段是不同的,表明不同类数据具有较大的差异性,且数据经过预处理后,特征波段也发生了变化,这种差异性和特征变化会影响最后识别精度。由图 4 可知,相比于 LSO、PSO 算法,ILSO 收敛速度更快,收敛精度更高。由表 4 可知,不同分类中同时包含了多种品质类型的牛肉样本,例如,对于“分类 7”,同时包含了正常肉、黑干肉和掺假肉 3 种牛肉品质类型,表明不同牛肉品质的样本在光谱特性上有相似性,进一步验证了采用分类进行牛肉品质识别的重要性。

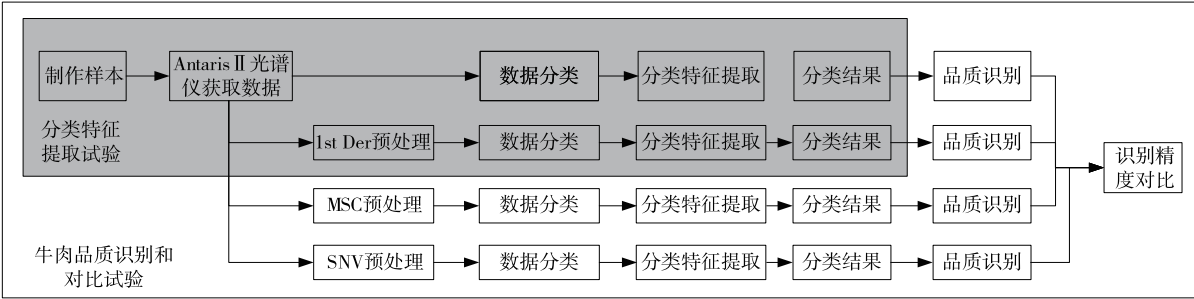


图 3 试验方法示意图

Figure 3 Schematic diagram of test method

表 3 分类结果与分类特征波段提取结果

Table 3 Classification results and classification feature band extraction results

分类	未经数据处理的原始数据		1 st Der 处理	
	特征波段	M_i	特征波段	M_i
1	431,468,552,586,611,724,768,841,904	9	424,489,541,567,598,614,622,714,729,941	10
2	441,476,489,502,547,584,613,784,882,912	10	416,428,478,516,578,603,619,752,741,882,952	11
3	403,456,514,547,582,661,674,701,789,813	10	442,478,496,522,576,618,713,768,846	9
4	554,578,589,615,667,716,884,903,915,967,992	11	507,563,594,615,662,664,734,809,884,913,928,964	12
5	407,422,465,489,503,518,582,609,635,694,774,803	12	423,471,531,557,589,606,632,678,702,881,956	11
6	407,522,547,648,678,718,882,917,906,945,985	11	487,501,526,557,608,664,708,814,922,991	10
7	422,456,513,552,589,641,656,732,758,841,903,955	12	406,418,483,517,548,613,629,675,715,831,856,915,955	13

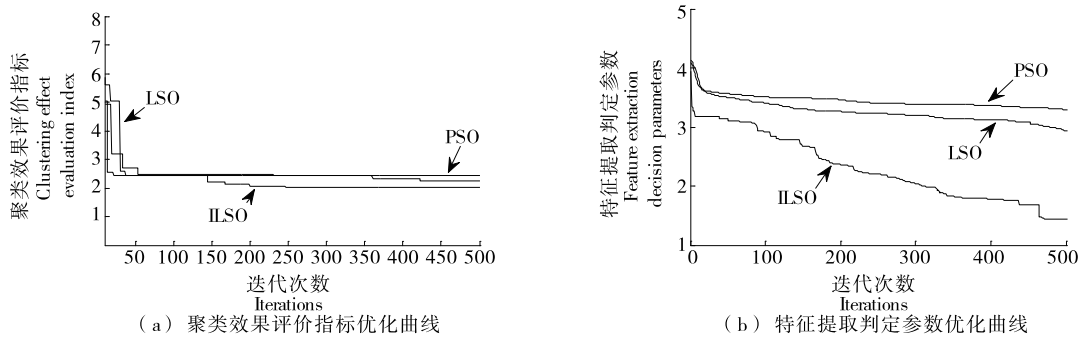


图 4 ILSO 优化 C_{EEI} 、 T_{ZT} 收敛曲线
Figure 4 ILSO optimization and convergence curve

3.2 牛肉品质识别和对比试验

利用测试样本数据集验证文中所提融合分类特征提取和 ILSO 优化 SVM 识别模型 (IDPeak-ILSO-SVM) 精度,表 5 给出了识别精度结果。

设置对比试验场景 1:不对训练样本进行聚类,采用主成分分析法 (PCA) 和文中提出的最佳特征提取方法 (CEEI) 对训练样本进行特征提取,采用文中提出的 ILSO 优化 SVM 模型 (ILSO-SVM) 进行识别,表 6 给出了识别精度对比结果。

设置试验场景 2:利用 DPeak、IDPeak 对训练样本进行聚类,采用 CEEI 提取特征波段,并运用 RF、ILSO-SVM 进行识别,表 6 给出了识别精度对比结果。

由表 5~表 7 可知,不同数据预处理方法对识别精度影响不同,相比于多元散射校正 (MSC)、标准正态变换

(SNV) 预处理方法以及原始数据,经 1st Der 预处理后的识别精度更高,而 MSC 表现最差。未进行聚类分析时,经 1st Der 预处理后,ILSO-SVM 对 4 种牛肉品质的识别精度分别为 92.4%,91.7%,92.8%,91.1%;采用 DPeak 算法进行聚类分析时,ILSO-SVM 的识别精度分别为 93.1%,92.8%,93.4%,93.5%;采用 IDPeak 算法进行聚类分析时,ILSO-SVM 的识别精度分别为 99.6%,99.1%,99.9%,99.5%,表明利用 IDPeak 算法进行聚类分析,降低了数据差异性对识别精度的影响。

由表 6 可知 (以 1st Der 预处理数据为例),未提取特征波段,ILSO-SVM 的识别精度分别为 82.6%,81.5%,79.6%,80.5%;采用 PCA 提取特征波段时,ILSO-SVM 的识别精度分别为 88.4%,89.3%,86.2%,87.5%;采用 CEEI 进行特征提取时,ILSO-SVM 的识别精度分别为 92.4%,91.7%,92.8%,91.1% (未进行聚类分析的情况下),表明利用 CEEI 提取到的特征波段的识别能力优于 PCA 算法,识别精度更高。由表 7 可知,采用 IDPeak 算法聚类分析、CEEI 分类特征提取,RF 的识别精度分别为

表 4 每个分类样本对应牛肉品质类型及个数
Table 4 Number of samples and beef quality type of each classification

分类	正常肉	白肌肉	黑干肉	掺假肉
1	47	7	0	10
2	0	30	0	3
3	2	0	0	24
4	0	44	0	4
5	0	4	52	0
6	25	3	9	0
7	5	0	4	27

表 5 IDPeak-ILSO-SVM 模型识别精度
Table 5 IDPeak-ILSO-SVM recognition accuracy %

预处理方法	正常肉	白肌肉	黑干肉	掺假肉
原始	94.2	95.3	94.9	94.7
1 st Der	99.6	99.1	99.9	99.5
MSC	95.3	96.2	97.4	98.4
SNV	97.2	98.4	97.8	98.2

表 6 试验场景 1 下 ILSO-SVM 模型识别精度对比
Table 6 Comparison of recognition accuracy under comparative test scenario 1 %

预处理方法	不提取				CEEI				PCA			
	正常肉	白肌肉	黑干肉	掺假肉	正常肉	白肌肉	黑干肉	掺假肉	正常肉	白肌肉	黑干肉	掺假肉
原始	78.1	77.7	75.6	76.2	88.4	86.9	87.3	86.4	81.4	82.6	81.9	82.3
1 st Der	82.6	81.5	79.6	80.5	92.4	91.7	92.8	91.1	88.4	89.3	86.2	87.5
MSC	68.1	70.2	66.9	69.7	77.2	75.8	74.9	73.7	65.1	64.7	66.2	64.0
SNV	75.2	74.6	74.9	75.3	81.2	80.6	82.5	81.1	70.8	71.2	73.3	71.5

94.2%, 93.6%, 92.7%, 94.6%, 低于 ILSO-SVM 识别精度, 表明采用 ILSO 对 SVM 进行参数配置, 提高了模型的识别精度。

为进一步验证所提 IDPeak-ILSO-SVM 性能, 与 SSA-SVM^[7]、1st Der-PCA-SVM^[8] 进行对比, 每种方法独

立运行 30 次, 选取识别精度均值、均方根误差均值为评价指标, 对比结果见表 8。由表 8 可知, 相比于其他 2 种识别模型, IDPeak-ILSO-SVM 的识别精度均值、均方根误差均值更优, 识别精度提高了约 5.02%~8.30%, 更适用于牛肉品质识别。

表 7 试验场景 2 下模型算法识别精度对比

Table 7 Comparison of recognition accuracy under comparative test scenario 2 %

预处理 方法	ILSO-SVM 模型+DPeak 算法				RF 模型+IDPeak 算法			
	正常肉	白肌肉	黑干肉	掺假肉	正常肉	白肌肉	黑干肉	掺假肉
原始	90.2	89.4	90.8	91.6	81.4	82.4	81.9	82.2
1 st Der	93.1	92.8	93.4	93.5	94.2	93.4	92.7	94.6
MSC	81.5	82.7	83.1	81.9	76.3	77.1	75.8	78.6
SNV	88.6	84.3	85.1	87.8	80.4	79.6	78.5	81.3

表 8 不同识别方法评价指标对比

Table 8 Comparison of evaluation indexes of different identification methods

方法	原始		1 st Der		MSC		SNV	
	识别精度	均方根误	识别精度	均方根误	识别精度	均方根误	识别精度	均方根误
	均值/%	差均值	均值/%	差均值	均值/%	差均值	均值/%	差均值
试验方法	98.41	1.23e-006	99.21	3.84e-006	97.16	4.21e-005	96.82	3.72e-004
SSA-SVM	96.52	4.28e-005	97.16	5.14e-004	90.24	6.15e-003	92.44	7.14e-004
1 st Der-PCA-SVM	96.18	6.12e-004	91.64	7.19e-005	92.51	3.74e-004	96.11	1.07e-004

4 结论

对正常肉、白肌肉、黑干肉、掺假肉 4 种牛肉品质类型识别问题进行研究, 提出了基于分类特征提取和深度学习的牛肉品质识别方法。结果表明, 通过设计改进 DPeak 算法、最佳分类特征波段提取策略和 ILSO 优化 SVM 模型, 降低了数据差异性以及特征冗余度对识别精度的影响, 识别精度更高。下一步, 将围绕提高在线牛肉品质识别精度问题进行研究。

参考文献

[1] BABAEI F, LASHKARI Z B, SAFARI A, et al. Salp swarm algorithm-based fractional-order PID controller for LFC systems in the presence of delayed EV aggregators[J]. IET Electrical Systems in Transportation, 2020, 10(3): 259-267.

[2] 毛衍伟, 张一敏, 朱立贤, 等. 中国牛羊肉的供需现状及消费者对牛羊肉的态度和品质需求[J]. 食品与发酵工业, 2016, 42(2): 244-251.

MAO Yan-wei, ZHANG Yi-min, ZHU Li-xian, et al. Beef and mutton production and consumers' requirement in China[J]. Food and Fermentation Industries, 2016, 42(2): 244-251.

[3] 熊德皓, 王东阳, 胡志全. 我国牛肉供需及国际竞争力研究[J]. 中国农业资源与区划, 2020, 41(10): 89-100.

XIONG S H, WANG D Y, HU Z Q. Research on China's beef supply, demand and international competitiveness[J]. Chinese Journal

of Agricultural Resources and Regional Planning, 2020, 41(10): 89-100.

[4] 明丹丹, 张一敏, 董鹏程, 等. 牛肉肉色的影响因素及其控制技术进展[J]. 食品科学, 2020, 41(1): 284-291.

MING D D, ZHANG Y M, DONG P C, et al. Recent progress in studies of factors influencing beef color and technologies for controlling it[J]. Food Science, 2020, 41(1): 284-291.

[5] 张花菊, 徐亚铂, 王琳琳, 等. 郟县红牛与不同品种肉牛杂交后代肉用性能比较研究[J]. 中国畜牧兽医, 2017, 44(12): 3 482-3 490.

ZHANG H J, XU Y B, WANG L L, et al. Comparatives study on the beef traits between Jiaxian red cattle and crossbred beef cattle[J]. China Animal Husbandry & Veterinary Medicine, 2017, 44(12): 3 482-3 490.

[6] 吴澎, 贾朝爽, 孙东晓. 食品感官评价科学研究进展[J]. 饮料工业, 2017, 20(5): 58-63.

WU Peng, JIA Chao-shuang, SUN Dong-xiao. Progress of food sensory evaluation science research[J]. Beverage Industry, 2017, 20(5): 58-63.

[7] 张保霞. 基于主成分分析和改进支持向量机的猪肉品质识别[J]. 食品与机械, 2022, 38(1): 146-151.

ZHANG Bao-xia. Pork quality identification based on principal component analysis and improved support vector machine[J]. Food & Machinery, 2022, 38(1): 146-151.

[8] 吴叶兰, 陈怡宇, 廉小亲, 等. 高光谱成像的柑橘病虫害叶片识

- 别方法[J]. 光谱学与光谱分析, 2021, 41(12): 3 837-3 843.
- WU Ye-lan, CHEN Yi-yu, LIAN Xiao-qin, et al. Study on the identification method of citrus leaves based on hyperspectral imaging technique[J]. Spectroscopy and Spectral Analysis, 2021, 41(12): 3 837-3 843.
- [9] 孔德明, 陈红杰, 陈晓玉, 等. 三维荧光光谱结合稀疏主成分分析和支持向量机的油类识别方法研究[J]. 光谱学与光谱分析, 2021, 41(11): 3 474-3 479.
- KONG De-ming, CHEN Hong-jie, CHEN Xiao-yu, et al. Research on oil identification method based on three-dimensional fluorescence spectroscopy combined with sparse principal component analysis and support vector machine[J]. Spectroscopy and Spectral Analysis, 2021, 41(11): 3 474-3 479.
- [10] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6 191): 1 492-1 496.
- [11] RAJAKUMAR B R. The Lion's algorithm: A new nature-inspired search algorithm[J]. Procedia Technology, 2012, 6: 126-135.
- [12] DENG Z, LI D, KE Y H, et al. An improved SVM algorithm for high spatial resolution remote sensing image classification[J]. Remote Sensing for Land & Resources, 2016, 28(3): 12-18.
- [13] 肖洪兵, 郭培源, 王瑜. 基于深度学习的高光谱腊肉营养安全分级[J]. 中国食品学报, 2022, 22(1): 275-281.
- XIAO Hong-bing, GUO Pei-yuan, WANG Yu. Nutritional health risk grading of bacon hyper-spectrum based on deep learning[J]. Journal of Chinese Institute of Food Science and Technology, 2022, 22(1): 275-281.
- [14] 朱永国, 邓斌, 霍正书, 等. 检测数据和专家知识混合驱动的小样本飞机结构件装配粗差判定[J]. 计算机集成制造系统, 2021, 27(12): 3 462-3 474.
- ZHU Yong-guo, DENG Bin, HUO Zheng-shu, et al. Assembly gross error identification of small sample aircraft structure driven by inspection data and expert knowledge[J]. Computer Integrated Manufacturing Systems, 2021, 27(12): 3 462-3 474.
- [15] 刘生建, 杨艳, 周永权. 一种群体智能算法: 狮群算法[J]. 模式识别与人工智能, 2018, 31(5): 431-441.
- LIU S J, YANG Y, ZHOU Y Q. A swarm intelligence algorithm-lion on swarm optimization[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(5): 431-441.
- [16] 刘振, 郭恒光, 任建存. 一种局部搜索能力增强的狮群算法[J]. 河南师范大学学报(自然科学版), 2019, 47(3): 35-41.
- LIU Z, GUO H G, REN J C. An enhanced local search lion optimization algorithm[J]. Journal of Henan Normal University (Natural Science Edition), 2019, 47(3): 35-41.
- [17] 田瑾. 高维多峰函数的量子行为粒子群优化算法改进研究[J]. 控制与决策, 2016, 31(11): 1 967-1 972.
- TIAN Jin. Improvement of quantum-behaved particle swarm optimization algorithm for high-dimensional and multi-modal functions[J]. Control and Decision, 2016, 31(11): 1 967-1 972.
- [18] 丁世飞, 徐晓, 王艳茹. 基于不相似性度量优化的密度峰值聚类算法[J]. 软件学报, 2020, 31(11): 3 321-3 333.
- DING Shi-fei, XU Xiao, WANG Yan-ru. Optimized density peaks clustering algorithm based on dissimilarity measure[J]. Journal of Software, 2020, 31(11): 3 321-3 333.
- [19] BARCENA-MARTIN E, SILBER J. The Bonferroni index and the measurement of distributional change[J]. Metron, 2017, 75(1): 1-16.
- [20] 孙广路, 宋智超, 刘金来, 等. 基于最大信息系数和近似马尔科夫毯的特征选择方法[J]. 自动化学报, 2017, 43(5): 795-805.
- SUN Guang-lu, SONG Zhi-chao, LIU Jin-lai, et al. Feature selection method based on maximum information coefficient and approximate markov blanket[J]. Acta Automatica Sinica, 2017, 43(5): 795-805.
- (上接第 90 页)
- [19] 唐雪梅, 张薇, 李慧. 卷烟真伪鉴别的近红外定性分析方法[J]. 烟草科技, 2008(11): 5-8.
- TANG Xue-fei, ZHANG Wei, LI Hui. Application of near-infrared reflectance spectroscopy in discriminating counterfeit cigarettes from genuine[J]. Tobacco Science & Technology, 2008(11): 5-8.
- [20] YANG S Y, HOU Y, YANG L C, et al. Automatic identification of cigarette brand using near-infrared spectroscopy and sparse representation classification algorithm[J]. Journal of the Brazilian Chemical Society, 2018, 29(7): 1 480-1 486.
- [21] 谢有超, 彭黔荣, 杨敏, 等. 基于烟丝近红外光谱的卷烟品牌识别方法[J]. 烟草科技, 2021, 54(3): 72-77.
- XIE You-chao, PENG Qian-rong, YANG Min, et al. Cigarette brand recognition based on NIR spectral data of cut filler[J]. Tobacco Science & Technology, 2021, 54(3): 72-77.
- [22] 曹妙玲. 基于卷烟常规化学检测数据及近红外信息的卷烟牌号识别[D]. 上海: 华东理工大学, 2012: 36-41.
- CAO Miao-ling. Discrimination of cigarettes brands based on their routine testing chemical indicators and near infrared spectra[D]. Shanghai: East China University of Science and Technology, 2012: 36-41.
- [23] VAPNIK Vladimir. The Nature of statistical learning theory[M]. Berlin: Springer, 1995: 133-161.
- [24] VAPNIK Vladimir. An overview of statistical learning theory[J]. IEEE Transactions on Neural Networks, 1999, 10(5): 988-999.
- [25] 杨志民, 刘广利. 不确定性支持向量机: 算法及应用[M]. 北京: 科学出版社, 2012: 74-84.
- YANG Zhi-min, LIU Guang-li. Uncertainty support vector machines: Algorithms and applications[M]. Beijing: Science Press, 2012: 74-84.
- [26] YANG Xin-she. Nature-inspired algorithms and applied optimization[M]. Gewerbestrasse: Springer International Publishing AG, 2018: 288-308.
- [27] FISTER Iztok, FISTER Iztok Jr, YANG Xin-she, et al. A comprehensive review of firefly algorithms[J]. Swarm and Evolutionary Computation, 2013(13): 34-46.