

IVISSA 算法冷鲜滩羊肉嫩度的高光谱模型优化

刘贵珊, 张 翀, 樊奈昀, 程丽娟, 余江泳, 袁瑞瑞

宁夏大学农学院, 宁夏 银川 750021

摘 要 高光谱成像可同时获取被检测对象的图像信息和光谱信息, 并对其内部成分进行定性和定量分析。国内外学者采用高光谱对肉品质研究多集中在水分、菌落总数、色泽、pH、挥发性盐基氮等方面, 在肉品嫩度检测中应用区间变量迭代空间收缩法优选特征波长的研究鲜有报道。利用可见-近红外(400~1 000 nm)和近红外(900~1 700 nm)高光谱结合化学计量学方法对冷鲜滩羊肉嫩度进行无损预测, 优选最佳建模波段。首先, 采集羊肉的高光谱图像, 提取样本感兴趣区域的光谱反射值, 采用 TA-XTplus 质构仪测量滩羊肉嫩度; 其次, 将两个波段下的原始光谱数据进行多元散射校正(multiple scattering correction, MSC)、去趋势(de-trending)、基线校准(baseline)、标准正态变量(standard normal variable, SNV)、归一化(normalize)和卷积平滑(Savitzky-Golay)等预处理; 分别采用连续投影算法(successive projection algorithm, SPA)、竞争性自适应加权算法(competitive adaptive reweighted sampling, CARS)、变量组合聚类分析法(variables combination population analysis, VCPA)和区间变量迭代空间收缩法(interval variable iterative space shrinkage approach, IVISSA)对最佳预处理的光谱数据优选特征波长; 最后, 建立冷鲜滩羊肉嫩度的偏最小二乘回归(partial least squares regression, PLSR)预测模型, 优选最佳建模波段。结果表明: (1)滩羊肉嫩度的近红外高光谱模型的预测效果优于可见-近红外高光谱; (2)经过多种预处理方法所建立的滩羊肉嫩度的模型中, 近红外区域的原始光谱(original spectra, OS)模型效果最优, 其 $R_c = 0.83$, $R_p = 0.79$, $RMSEC = 874.94$, $RMSEP = 1 465.97$; (3)近红外高光谱的原始光谱经 SPA, CARS, VCPA, IVISSA 四种方法共挑选出 15, 16, 13 和 123 个特征波长, 占总波长的 7%, 6%, 5% 和 54%; (4)近红外高光谱结合 OS-IVISSA-PLSR 建立的冷鲜滩羊肉嫩度预测模型最好, 其 $R_c = 0.85$, $RMSEC = 850.86$, $R_p = 0.79$, $RMSEP = 1 497.11$ 。IVISSA 算法不仅可大幅度减少模型运算次数, 还可以保证模型的精准和稳定性。研究表明, OS-IVISSA-PLSR 模型对冷鲜滩羊肉嫩度进行高光谱的快速无损检测是可行的。

关键词 冷鲜滩羊肉; 嫩度; 高光谱成像技术; 区间变量迭代空间收缩法; 偏最小二乘回归

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)08-2558-06

引 言

盐池滩羊是宁夏优势特色畜种, 肉质鲜嫩、低脂无膻、脂肪均匀^[1]。嫩度是评价滩羊肉品质极为重要的指标之一^[2], 决定了肉在烹饪后的口感^[3]。肉品嫩度的传统检测方法为 WBS 剪切仪, 检测过程复杂且破坏样本, 不能满足肉品嫩度快速无损检测的需求。近年来, 高光谱成像技术作为新型快速无损检测手段, 广泛应用于畜产品内外部品质的检测^[4-5], 为肉品嫩度的快速无损检测提供了理论依据^[6-8]。高光谱成像技术中所采集的样品光谱波段较多, 这些波段不仅包含样品信息还包含了大量的无用信息, 如噪音、背景干扰

等, 通常会优选特征变量对原始光谱数据降维, 提取有效信息, 减少模型的运算次数, 提高模型的预测精度。因此, 特征变量的选择对光谱数据分析和模型简化有着重要的意义。国内外学者主要采用 SPA^[9]、CARS^[10] 和 VCPA^[11] 等常规算法来进行特征变量选取, 其弊端在于样本中校正集出现变换或波动时, 会影响集合内变量的判断, 导致所构建模型预测能力显著降低, 而 IVISSA^[12] 算法在弥补这一缺陷的同时还可以最大限度的提取样品的有效信息, 保证变量特征信息的完整性, 降低运算次数, 提高预测模型的鲁棒性。

以冷鲜滩羊肉为研究对象, 利用高光谱成像系统采集图谱信息, 采用六种预处理方法对可见-近红外和近红外区域的波段进行预处理, 并采用 SPA, CARS, VCPA 和 IVISSA

收稿日期: 2019-07-16, 修订日期: 2019-11-10

基金项目: 国家自然科学基金项目(31760435)资助

作者简介: 刘贵珊, 1979 年生, 宁夏大学农学院副教授 e-mail: liugs@nxu.edu.cn

算法进行降维处理, 比较以上算法提取出的特征波长所构建模型的预测效果, 为盐池滩羊肉品质的在线快速无损检测提供理论依据。

1 实验部分

1.1 材料

6 月龄的盐池滩羊背部最长肌样本 213 个, 采集于宁夏盐池县大夏牧场食品有限公司。

1.2 仪器与设备

可见-近红外高光谱(400~1 000 nm), 采用美国 Headwall 公司生产的 HyperSpec VNIR N 型设备, 光谱分辨率为 2.8 nm, 狭缝宽度为 25 μm , 狭缝长度 18 mm; CCD 相机像素尺寸 8.0 μm , 光源为卤素灯双线光源。

近红外高光谱(900~1 700 nm), 采用芬兰 Specimen 公司的 Inspector N17E-NIR 型成像光谱仪, 光谱分辨率 5 nm, 狭缝宽度 30 μm , 狭缝长度 9.6 mm, CCD 型号为 Models XC-130 100 Hz, 混合像元尺寸 30 \times 30 pix, 全幅帧数 60 帧 \cdot s⁻¹, 光源为 Hyper-SIS 系统功率, 35 W, 光通量 3 200 lm 的 4 个卤素灯。

TA-Xtplus 质构仪(英国 SMS 有限公司), 刀具为 HDP-BSW, 厚度 3 mm。

1.3 方法

1.3.1 光谱采集

高光谱系统预热半小时后进行校准和参数设定^[13-14]。羊肉样本暴露于室温 0.5 h, 用纸巾擦去表面水分, 放置于高光谱电控位移平台, 进行光谱扫描。使用 ENVI4.8 软件提取样本感兴趣(1.5 cm \times 1.5 cm)。

1.3.2 嫩度测量

参照国标 NY/T 1180—2006《肉嫩度的测定剪切力测定法》。

1.3.3 IVISSA 算法

IVISSA 算法^[12]是从整体和个体两个方面对波段进行挑选。其中整体和个体分析采用波段的权重优化, 保证各个波段的间隔位置和宽度。设光谱数据为 X , 嫩度值为 Y , 样品个数为 n , 波长变量个数为 p 。其原理如下:

(1) 采用权重二进制矩阵采样法(WBMS)将 X 进行随机组合, 产生子数据集 $x(k \times p)$, 其中 $0 < k < n$, 权重列数为 0.5, 且 x 个数均为 0.5k, 保证了波长变量在运算时有相同的可能性;

(2) 将得到的子数据集 x 作为校正集模型进行回归分析;

(3) 分析比较 m 个子模型的 RMSECV 值和 RMSEP 值, 取 k 个模型中两值较低子模型, 求得每个波长在这些子模型中出现的次数 f_i , 从而重新定义该波长的权重, 为 $w_i =$

$$\frac{f_i}{k_{\text{best}}};$$

(4) 重复(2)和(3)过程来变换波长和权重值, 当所有子模型的 RMSECV 值和 RMSEP 值误差改变时, 权重结果为 1 的波长构成原始光谱数据的特征波长, 运行程序结束。

2 结果与讨论

2.1 原始光谱分析

由图 1 可知, 可见-近红外高光谱在 450~550 和 700~800 nm 之间存在两个明显的吸收峰, 在 700~800 nm 之间样本存在较强的反射率。近红外高光谱在 1 000~1 010 nm 处出现一个较为明显的吸收峰。两个波段区域的大部分光谱曲线具有相同的趋势, 但也存在极个别光谱曲线出现差异的现象, 其原因可能是存在异常样本或光谱扫描过程中噪音的影响。在 900~1 000 nm 波段中出现较为明显的噪音波段, 其原因可能是由于仪器噪音或者暗电流的影响。

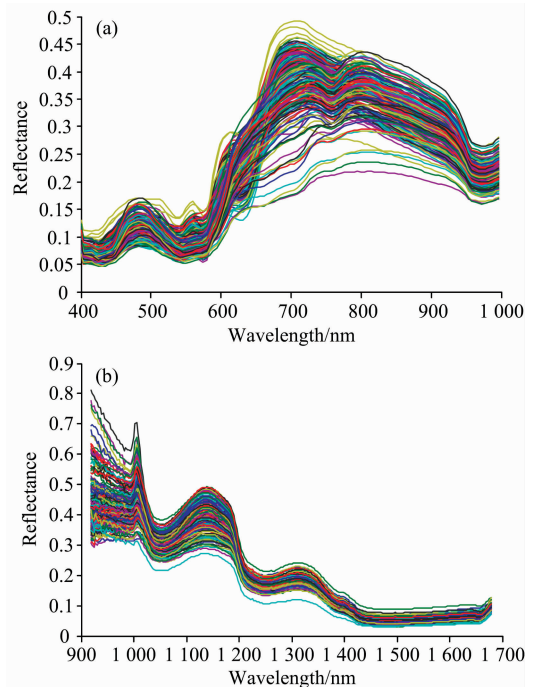


图 1 滩羊肉样本原始光谱曲线

(a): 400~1 000 nm; (b): 900~1 700 nm

Fig. 1 Original spectra of mutton samples

(a): 400~1 000 nm; (b): 900~1 700 nm

2.2 样本集划分

由于光谱信息采集过程中, 存在外界噪音等影响因素, 导致某些光谱信息出现异常, 故采用蒙特卡洛法剔除异常值。将剔除异常值的滩羊肉样本采用 SPXY(sample set partitioning based on joint x-y distances)法将按 3:1 划分为校正集样本和验证集样本。如表 1 所示: 400~1 000 nm 波段, 校正集为 152 个样本, 验证集为 50 个样本; 1 000~1 700 nm 波段, 校正集为 146 个样本, 验证集为 50 个样本。

2.3 预处理方法优选

如表 2 所示, 同种预处理方法所构建的 1 000~1 700 nm 波段的 PLSR 模型效果均优于 400~1 000 nm 波段(De-trending 预处理方法除外), 这是由于 1 000~1 700 nm 波段所采集光谱反映的是样本的物理特性, 更适用于嫩度模

型的构建。在 1 000~1 700 nm 波段所构建的 PLSR 模型中，原始光谱的相关系数 R_c 值为 0.83， R_p 值为 0.79，表明原始光谱所建立的滩羊肉嫩度模型更加稳定和精确。预处理后所

建立的滩羊肉嫩度模型效果变差的原因可能是原始光谱经过预处理后其中某些光谱特征信息被覆盖或者隐藏。

表 1 不同波段校正集与预测集滩羊肉的嫩度(单位: g)

Table 1 Mutton tenderness of calibration and prediction in different bands(unit: g)

Band /nm	Quantity	Calibration Set				Quantity	Prediction Set			
		Maximum	Minimum	Average Value	Standard Deviation		Maximum	Minimum	Average Value	Standard Deviation
400~1 000	152	12 205.69	549.71	4 872.04	2 629.22	50	9 473.68	1 154.26	4 650.41	1 619.30
1 000~1 700	146	10 578.62	941.32	4 731.51	2 350.09	50	9 473.68	1 306.89	4 690.13	1 584.63

表 2 不同预处理方法所建立的 PLSR 模型效果对比

Table 2 Comparison of PLSR models established by different pretreatment methods

Band/nm	Pretreatment Method	Calibration Set			Prediction Set	
		R_c	RMSEC	RMSECV	R_p	RMSEP
400~1 000	Original Spectra	0.74	1 079.61	1 316.64	0.72	1 911.34
	MSC	0.73	1 093.27	1 288.98	0.73	1 893.33
	De-trending	0.74	1 081.06	1 344.75	0.73	1 950.31
	SNV	0.75	1 068.35	1 324.32	0.74	1 871.64
	Baseline	0.76	1 048.62	1 323.23	0.74	1 892.56
	Normalize	0.73	1 095.05	1 321.50	0.72	1 904.65
	Savitzky-Golay	0.73	1 108.09	1 323.98	0.72	1 965.37
1 000~1 700	Original Spectra	0.83	874.94	1 287.84	0.79	1 465.97
	MSC	0.82	908.60	1 341.70	0.77	1 513.08
	De-trending	0.60	1 267.08	1 374.09	0.67	1 743.06
	SNV	0.81	917.37	1 338.56	0.76	1 550.99
	Baseline	0.82	891.29	1 347.08	0.79	1 456.17
	Normalize	0.82	884.82	1 281.13	0.79	1 485.75
	Savitzky-Golay	0.75	1 039.58	1 288.26	0.81	1 417.51

2.4 特征波长提取

如图 2 所示，在 1 000~1 700 nm 波段内对原始光谱数据进行降维处理。优选过程中，RMSECV 值逐渐减小，在第 15 个波段 RMSECV 最小为 1 535.40，随后小幅度增大。因此选用前 15 个波段建立 PLSR 模型，其优选波段占 228 个总波段的 7%。

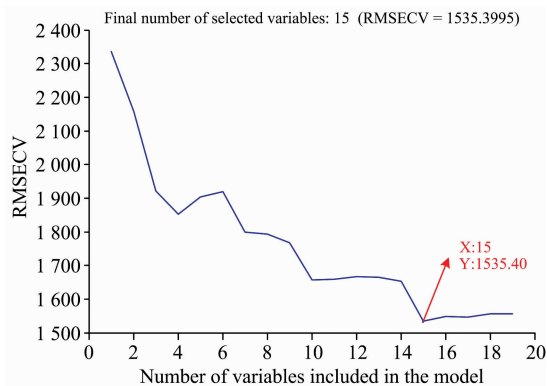


图 2 SPA 算法提取 PLSR 模型中的特征波长
Fig. 2 Characteristic wavelength in PLSR model by SPA algorithm extraction

如图 3 所示，采用 CARS 法提取特征波长，设定其运行次数为 1 000。图 3(a)为筛选特征变量数的过程，分为为两部分，即第一阶段快速减少(粗选)而第二阶段非常缓慢(精选)。图 3(b)为 RMSECV 的变化趋势图，当 RMSECV 值最小为 1 164.31 时，挑选出 16 个特征波长，占 228 个总波段

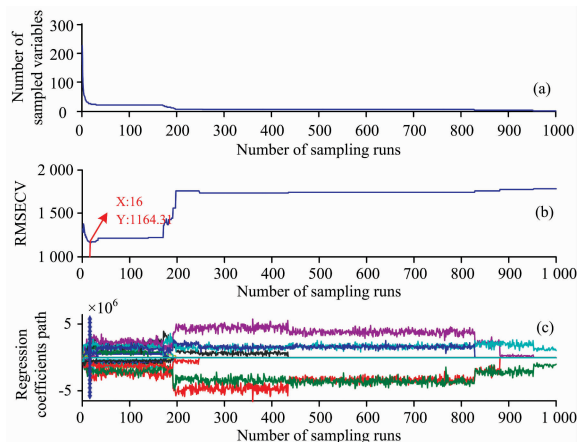


图 3 CARS 算法在 PLSR 模型中波长选择过程
Fig. 3 Wavelength selection process in PLSR model by CARS algorithm

的 6%。图 3(c)中每条线代表回归系数的变化趋势，* 号表示 RMSECV 最小的位置。

VCPA 法在 1 000~1 700 nm 原始光谱基础上所提取的特征波长共 13 个，占总波段数的 5%。其挑选过程中 RMSECV 值最小时为 1 296.24；IVISSA 算法在 1 000~1 700 nm 原始光谱基础上所挑选的特征波长个数为 123 个，占总波段数的 54%，其 RMSECV 最小值为 1 211.19。IVISSA 算法不仅降低了运算次数，而且还提高了模型性能。

2.5 PLSR 模型构建

图 4 显示在 900~1 700 nm 波段所构建的 PLSR 滩羊嫩度预测模型结果，其中 OS-IVISSA-PLSR 模型效果最优， $R_c = 0.85$ ， $RMSEC = 850.86$ ， $RMSECV = 1 193.42$ ， $R_p = 0.79$ ， $RMSEP = 1 497.11$ 。相对于王婉娇^[15]使用 NIR 高光谱成像技术检测冷鲜羊肉嫩度的预测相关系数(R_p)提高了 0.017。

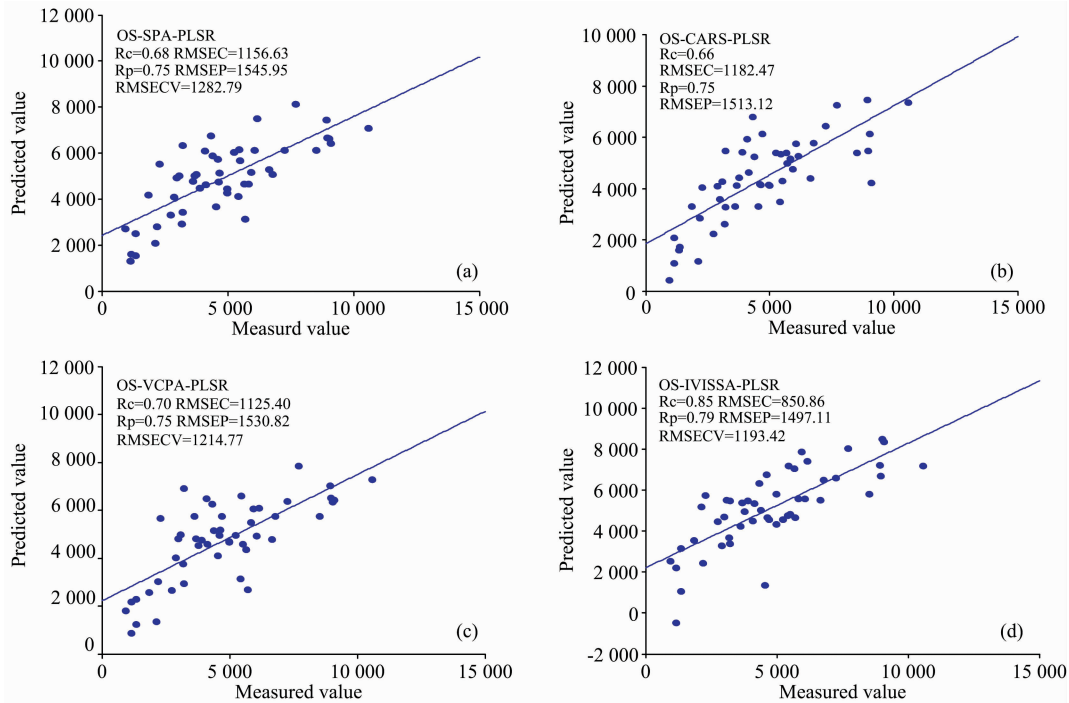


图 4 滩羊肉样本嫩度预测模型

(a): OS-SPA-PLSR; (b): OS-CARS-PLSR; (c): OS-VCPA-PLSR; (d): OS-IVISSA-PLSR

Fig. 4 Prediction Models of Tenderness in Tan mutton

(a): OS-SPA-PLSR; (b): OS-CARS-PLSR; (c): OS-VCPA-PLSR; (d): OS-IVISSA-PLSR

表 3 不同特征波长提取方法所建立的 PLSR 模型效果对比

Table 3 Performance comparison of PLSR models based on different wavelength extraction methods

Modeling method	method of extraction	Number of wavelengths	Calibration Set			Prediction Set		Wavelengths /nm
			R_c	RMSEC	RMSECV	R_p	RMSEP	
PLSR	SPA	15	0.68	1 156.63	1 282.79	0.75	1 545.95	1 002, 1 016, 1 031, 1 076, 1 160, 1 213, 1 225, 1 264, 1 312, 1 386, 1 416, 1 496, 1 630, 1 672, 1 678
	CARS	16	0.66	1 182.47	1 318.19	0.75	1 513.12	1 031, 1 037, 1 046, 1 097, 1 118, 1 121, 1 189, 1 207, 1 273, 1 294, 1 320, 1 326, 1 380, 1 386, 1 392, 1 428
	VCPA	13	0.70	1 125.40	1 214.77	0.75	1 530.82	1 052, 1 097, 1 100, 1 160, 1 201, 1 207, 1 228, 1 258, 1 386, 1 389, 1 398, 1 496, 1 633
	IVISSA	123	0.85	850.86	1 193.42	0.79	1 497.11	1 002~1 076, 1 091~1 097, 1 121~1 130, 1 160, 1 183~1 198, 1 216~1 228, 1 252~1 314, 1 329~1 350, 1 365~1 401, 1 449~1 478, 1 496, 1 526~1 553, 1 574~1 595, 1 621~1 633, 1 660~1 678

如表 3 为 SPA, CARS, VCPA 和 IVISSA 四种算法选取特征波长建立的 PLSR 模型。经过对比, 最优模型为 IVISSA-PLSR。分析原因如下: (1) 通过 IVISSA 提取出的特征波长信息与嫩度值拟合度高, 排除了样本中无关光谱信息, 使得预测性能更加优越; (2) 影响羊肉嫩度本身的因素有很多, 如蛋白质含量、脂肪含量、肉质纹理和水分等, 而通过前三种特征波长算法所提取的特征波段可能排除了某一因素下的光谱信息, 导致预测结果偏低。

3 结 论

利用可见-近红外高光谱(400~1 000 nm)和近红外高光谱(900~1 700 nm)系统采集了冷鲜滩羊肉原始光谱, 经过优选预处理方法和最佳波段后, 建立基于 SPA, CARS, VC-

PA 和 IVISSA 算法下的 PLSR 嫩度预测模型。结论如下:

(1) 近红外高光谱的原始光谱模型的预测效果优于可见-近红外高光谱。

(2) 在 900~1 700 nm 波段下, 采用不同预处理方法建模, 原始光谱模型效果最优, 其 $R_c=0.83$, $R_p=0.79$, $RMSEC=874.94$, $RMSEP=1 465.9$ 。

(3) 900~1 700 nm 波段的原始光谱经 SPA, CARS, VCPA 和 IVISSA 四种方法所挑选出 15, 16, 13 和 123 个特征波长, 占总波长的 7%, 6%, 5% 和 54%。

(4) 与其他的变量选择方法相比, IVISSA-PLSR 模型的 $R_c=0.85$, $R_p=0.79$, $RMSEC=850.86$, $RMSEP=1 497.11$, $RMSECV=1 193.42$, 预测精准度更高, 模型稳定性更好。结果表明, IVISSA 是一种有效简化特征波长的提取方法, 可提高肉品嫩度模型的预测精度和鲁棒性。

References

- [1] WANG Song-lei, WU Long-guo, MA Tian-li, et al(王松磊, 吴龙国, 马天利, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(9): 2937.
- [2] SONG Hong-xia, HUANG Feng, DING Zhen-jiang, et al(孙红霞, 黄峰, 丁振江, 等). Food Science(食品科学), 2018, 39(1): 84.
- [3] Sun H, Peng Y, Zheng X, et al. Journal of Food Engineering, 2019, 248: 1.
- [4] ZHAO Juan, PENG Yan-kun(赵娟, 彭彦昆). Journal of Agricultural Engineering(农业工程学报), 2015, 31(7): 279.
- [5] Lee Hoonsoo, Kim M S, Lee W H, et al. Sensors and Actuators B: Chemical, 2018, 259: 532.
- [6] Achata E M, Inguglia E S, Esquerre C A, et al. Journal of Food Engineering, 2019, 246: 134.
- [7] Qu F, Ren D, He Y, et al. Meat Science, 2018, 146: 59.
- [8] Lee H, Kim M S, Lee W, et al. Sensors and Actuators B: Chemical, 2018, 259: 532.
- [9] Zhang D, Xu Y, Huang W, et al. Infrared Physics & Technology, 2019, 98: 297.
- [10] CHENG Li-juan, LIU Gui-shan, HE Jian-guo, et al(程丽娟, 刘贵珊, 何建国, 等). Food Science(食品科学), 2019, 40(10): 285.
- [11] ZHAO Huan, HUAN Ke-wei, SHI Xiao-guang, et al(赵环, 宦克为, 石晓光, 等). Analytical Chemistry(分析化学), 2018, 46(1): 136.
- [12] Song X, Huang Y, Yan H, et al. Analytica Chimica Acta, 2016, 948: 19.
- [13] WU Long-guo, HE Jian-guo, LIU Gui-shan, et al(吴龙国, 何建国, 刘贵珊, 等). Photoelectron Laser(光电子激光), 2014, 25(1): 135.
- [14] Ma J, Sun D, Pu H, et al. Journal of Food Engineering, 2019, 240: 207.
- [15] WANG Wan-jiao, WANG Song-lei, HE Xiao-guang, et al(王婉娇, 王松磊, 贺晓光, 等). Science and Technology of Food Industry(食品工业科技), 2015, 36(20): 77.

Hyperspectral Model Optimization for Tenderness of Chilled Tan-Sheep Mutton Based on IVISSA

LIU Gui-shan, ZHANG Chong, FAN Nai-yun, CHENG Li-juan, YU Jiang-yong, YUAN Rui-rui
College of Agriculture, Ningxia University, Yinchuan 750021, China

Abstract Hyperspectral imaging can obtain both the image information and spectral information of the detected object, and make qualitative and quantitative analysis with internal components. The research on meat quality by hyperspectral imaging technology focuses on water, total viable count, color, pH, total volatile basic nitrogen etc. There are few studies on meat tenderness based on interval variable iterative space contraction method. In this paper, tenderness values of chilled Tan-sheep were detected non-destructively by visible-near-infrared and near-infrared bands (400~1 000, 900~1 700 nm) combined with chemometric methods to obtain the best modeling bands. Firstly, hyperspectral images of lamb sample were collected and extracted the spectral values of the region of interest; lamb tenderness was measured using a TA-XTplus texture analyzer; Secondly, the original spectral da-

ta between 400~1 000 and 900~1 700 nm were preprocessed by multiple scattering correction (MSC), de-trending, standard normal variate (SNV), baseline, normalize, Savitzky-Golay; optimal bands were selected by successive projection algorithm (SPA), competitive adaptive reweighted sampling (CARS), variables combination population analysis (VCPA) and interval variable iterative space shrinkage approach (IVISSA). Finally, optimal bands were selected and partial least squares regression (PLSR) model of chilled mutton was established. The results showed that: (1) the prediction performance of original spectral model in the region of 900~1 700 nm was better than that of 400~1 000 nm. (2) original spectral mode of the tenderness of chilled Tan-sheep had the best performance $R_c=0.83$, $R_p=0.79$, $RMSEC=874.94$, $RMSEP=1465.97$ in the near-infrared region using different pretreatment methods. (3) the original spectrum of 900~1 700 nm was selected by SPA, CARS, VCPA and IVISSA with 15, 16, 13 and 123 characteristic wavelengths, accounting for 7%, 6%, 5% and 54% of the total wavelength. (4) the prediction model of chilled Tan-sheep tenderness was the best in combination with hyperspectral technique and OS-IVISSA-PLSR with $R_c=0.85$, $RMSEC=850.86$, $RMSECV=1193.42$, $R_p=0.79$, $RMSEP=1497.11$. It was indicated that IVISSA algorithm could greatly reduce the number of model operations and ensure the predictability and stability of the model. It is feasible to adopt OS-IVISSA-PLSR method to analyze the tenderness of chilled Tan-sheep mutton.

Keywords Chilled Tan-sheep mutton; Tenderness; Hyperspectral imaging technology; Interval variable iterative space shrinkage approach; Partial least squares regression

(Received Jul. 16, 2019; accepted Nov. 10, 2019)

《光谱学与光谱分析》对来稿英文摘要的要求

来稿英文摘要不符合下列要求者, 本刊要求作者重写, 这可能要推迟论文发表的时间。

1. 请用符合语法的英文, 要求言简意明、确切地论述文章的主要内容, **突出创新之处**。
2. 应拥有与论文同等量的主要信息, 包括四个要素, 即研究目的、方法、结果、结论。其中后两个要素最重要。有时一个句子即可包含前两个要素, 例如“用某种改进的 ICP-AES 测量了鱼池水样的痕量铅”。但有些情况下, 英文摘要可包括研究工作的主要对象和范围, 以及具有情报价值的其他重要信息。在结果部分最好有定量数据, 如检测限、相对标准偏差等; 结论部分最好指出方法或结果的优点和意义。
3. 句型力求简单, 尽量采用被动式, 建议经专业英语翻译机构润色, 与中文摘要相对应。用 A4 复印纸单面打印。
4. 摘要不应有引言中出现的内容, 换言之, 摘要中必须写进的内容应尽量避免在引言中出现。摘要也不要对论文内容作解释和评论, 不得简单重复题名中已有的信息; 不用非公知公用的符号和术语; 不用引文, 除非该论文证实或否定了他人已发表的论文。缩略语、略称、代号, 除相邻专业的读者也能清楚地理解外, 在首次出现时必须加以说明, 例如用括号写出全称。