

基于二次迭代 Monte Carlo 的羊肉硬度定量检测研究

白雪冰¹, 李鑫星¹, 张小栓², 罗海玲³, 傅泽田^{2*}

1. 中国农业大学信息与电气工程学院食品质量与安全北京实验室, 北京 100083
2. 中国农业大学工学院, 北京 100083
3. 中国农业大学动物科学技术学院动物营养国家重点实验室, 北京 100083

摘要 随着居民生活水平的提高和对健康饮食结构的重视, 羊肉作为一种高蛋白且低脂肪和胆固醇的畜肉, 需求量逐年上涨。根据国家统计局统计, 2012年—2019年我国畜肉产业中羊肉产量占比从6.27%上升到9.02%。研究提出了一种基于二次迭代 Monte Carlo (MC) 算法剔除异常样本的羊肉硬度定量检测 PLSR 模型。采用 GaiaSorter 高光谱分选仪的 Image- λ -V10E-H 相机采集羊肉样品 400~950 nm 的高光谱数据, Image- λ -N17E 相机采集羊肉样品 900~1 650 nm 的高光谱数据。首先, 对比分析了 S-G 平滑、二阶求导、多元散射校正 (MSC)、标准正态变换 (SNV) 等光谱预处理方法在消除噪声影响, 提高光谱分辨率等方面的能力, 选取最佳光谱预处理方法。然后, 在第一次 MC 抽样中, 计算所有样本预测误差均值和标准差的平均值, 以该平均值的 2.5~3 倍作为可疑样本阈值, 3 倍作为异常样本阈值; 剔除异常样本, 保留并标注可疑样本, 进行第二次 MC 抽样, 以样本预测误差均值和标准差的 3 倍值为阈值进行异常样本二次剔除; 对第一次 MC 抽样中标注可疑样本进行二次检测。最后, 对比分析了基于全波长建立的偏最小二乘回归 (PLSR) 模型和基于回归系数法 (RC) 提取的特征波长建立的 PLSR 模型。研究表明, 所提出的二次迭 MC 算法可以准确判别可疑样本是否为异常样本, 有效优化样本集, 为建模提供良好的数据基础。以 MSC 作为光谱预处理算法基于 400~950 和 900~1 650 nm 两段高光谱数据建立 PLSR 模型的 R_p^2 分别为 0.947 2 和 0.978 3, $RMSE_p$ 分别为 47.789 9 和 30.590 1 g, 优于其他三种光谱预处理算法。另外, 基于 900~1 650 nm 建立的 PLSR 模型明显优于基于 400~950 nm 波长样本集建立的模型。通过 RC 算法选取出羊肉硬度在 400~950 和 900~1 650 nm 波长范围的特征波长分别为 14 个 (410, 438, 450, 464, 539, 558, 612, 684, 701, 734, 778, 866, 884 和 935 nm) 和 10 个 (915, 949, 1 085, 1 156, 1 206, 1 262, 1 318, 1 384, 1 542 和 1 580 nm)。其中, 基于 900~1 650 nm 波长建立的 PLSR 模型的 R_p^2 为 0.985 0, $RMSE_p$ 为 24.397 0 g, 为羊肉硬度预测的最佳模型。结果表明, 所提出的融合二次迭代 MC 算法的 PLSR 模型可以有效预测羊肉冷藏过程中硬度特性变化趋势, 为羊肉品质无损检测相关研究提供参考。

关键词 可见-近红外光谱; 二次迭代 Monte Carlo; 偏最小二乘回归模型; 羊肉硬度

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)07-2057-07

引言

随着居民生活水平的提高和对健康饮食结构的重视, 羊肉作为一种高蛋白且低脂肪和胆固醇的畜肉, 需求量逐年上涨。根据国家统计局对 2012 年至 2019 年我国肉类参量的统计, 八年间猪肉在畜肉产量中的占比从 84.22% 下降到 78.65%, 而羊肉从 6.27% 上升到 9.02%。2019 年我国羊肉

产量达到 488 万吨。羊肉的感官特性主要包括颜色、硬度等, 能够很大程度上反映鲜肉的外观属性和质地特征, 直接影响消费者的购买决策^[1]。其中, 羊肉的硬度信息与肌肉纤维密度、脂肪含量与分布均匀性等密切相关, 可以反映出活羊养殖年龄、屠宰与贮藏时间等羊肉品质状况^[2]。羊肉的硬度可通过质地剖面分析 (TPA) 准确测量, 但该方法对样本的破坏性较强, 不适用于羊肉生产在线检测的需求^[3]。目前, 在羊肉产业中常由人工采用“指压法”测定, 存在较大的主观

收稿日期: 2020-07-19, 修订日期: 2020-12-06

基金项目: 国家肉羊产业技术体系 (CARS-38) 资助

作者简介: 白雪冰, 1991 年生, 中国农业大学信息与电气工程学院博士研究生 e-mail: 464161695@qq.com

* 通讯作者 e-mail: fzt@cau.edu.cn

性^[4]。

高光谱成像技术可以获得羊肉样本图像上每个像素的光谱信息,通过计算感兴趣区域(ROD)的平均光谱可以具有代表性地表述羊肉样本的硬度属性,为羊肉硬度的无损快速检测提供了技术支持。Cheng 等^[5-6]选出了 450, 530, 550, 616, 720, 955 和 980 nm 的 7 个最佳波长,建立了鱼片硬度的 LS-SVM 预测模型。Tao 等^[7]基于 400~1 100 nm 波段的猪肉高光谱散射图像,采用 Lorentzian 分布函数拟合散射剖面,建立了猪肉硬度和弹性的无损预测模型。Kamruzzaman 等^[8]基于 900~1 700 nm 高光谱成像系统建立了羊肉硬度和剪切力的预测模型,研究表明,900~1 700 nm 光谱数据在羊肉硬度的无损检测研究中更具潜力。刘贵珊^[9]等采集了羊肉 400~950 和 900~1 650 nm 两段高光谱数据,区间变量迭代空间收缩法提取了 123 个特征波长,提高了羊肉嫩度预测模型的准确性和稳定性。现有相关研究一般建立在相对理想的实验数据基础上,对于由仪器误差、操作误差等客观因素造成的异常样本的检测方法研究相对较少。

本工作旨在研究 400~950 和 900~1 650 nm 两段高光谱数据对羊肉贮藏过程中硬度变化趋势预测的可行性。首先,提出了二次迭代 Monte Carlo(MC)法对异常样本进行有效的检测、验证和剔除;然后,基于 PLSR 算法建立了羊肉硬度定量预测模型;最后,采用回归系数法(R_c)从 400~950 和 900~1 650 nm 两段高光谱数据中提取了羊肉硬度特征波长,并建立基于特征波长的 PLSR 模型。

1 实验部分

1.1 数据采集

实验样品选定为绵羊的背部最长肌,是一次性从北京某超市购置并通过恒温箱送至实验室。将羊肉里脊横切处理为厚度约 1cm 的肉片,分装入无菌袋中作为待测样本。共制备 140 个样本,对其中 5 个样本立即进行高光谱数据采集和硬度测定,作为样品的初始标定值。其余 135 个样本分为三组贮藏于 0, 4 和 8 °C 的恒温箱中。每隔 24h 从各温度组随机选取 5 个样本,进行高光谱数据采集和硬度测定。实验中选用 GaiaSorter 高光谱分选仪的 Image- λ -V10E-H 相机采集样品 400~1 000 nm 的高光谱数据,设置电控平台移动速度为 0.35 cm·s⁻¹,曝光时间调整为 19.3 ms; Image- λ -N17E 相机采集样品 900~1 700 nm 的高光谱数据,设置电控平台移动速度为 0.65 cm·s⁻¹,曝光时间调整为 25 ms; TA.XTPlus 物性测试仪测量样本的硬度值,设置探头测试速度为 1 mm·s⁻¹,样品压缩程度为 50%。

1.2 异常样本剔除

MC 抽样法对异常样本判定阈值具有较强的敏感性,阈值设定过大将无法检测出部分异常样本,而阈值设定过小则导致部分正常样本被误判^[10]。此外,当异常样本被随机分配到校正集次数过多时,会使正常样本的预测均值产生偏差,从而无法对该样本做出异常认定。本研究提出一种基于二次迭代理论的 MC 抽样改进方法,以提高异常样本检测的准确能力。具体步骤如下:

Step 1: 以所有样本为校正集建立主成分预测模型,得到最优主成分数;

Step 2: 采用 MC 方法从样本集中抽取 3/4 或以上的样本作为校正集,根据 Step 1 中所得的最优主成分数建立预测模型,并对剩余样本进行预测,获得剩余样本的预测误差;

Step 3: 将 Step 2 重复执行 N 次(N 需足够大),确定每个样本都作为剩余样本得到相应的预测误差;

Step 4: 基于上一步每个样本所得的多个预测误差,计算各样本的均值 μ 和标准差 σ ,计算方法见式(1)和式(2)

$$\mu_j = \frac{1}{k} \sum_{i=1}^k \epsilon_i \quad (1)$$

$$\sigma_j = \sqrt{\frac{1}{k-1} \sum_{i=1}^k (\epsilon_i - \mu_j)^2} \quad (2)$$

式(1)和式(2)中, k 为第 j 个样本作为剩余样本得到预测误差的次数, ϵ_i 为该样本在第 i 次作为剩余样本所得的预测误差。

Step 5: 分别计算各样本作为剩余样本所得的预测误差均值和标准差;再求的所有样本预测误差均值和标准差的平均值,以该平均值的 2.5~3 倍作为可疑样本阈值,3 倍作为异常样本阈值。剔除异常样本,保留并标记可疑样本,对剩余样本进行第二次 MC 抽样法。

Step 6: 对剔除异常样本后的样本集进行第二次 MC 抽样,再次计算所有样本预测误差均值和标准差的平均值,以该平均值 3 倍作为异常样本阈值。

Step 7: 对 Step 5 标记的可疑样本进行逐一剔除判定,确定异常样本集。

1.3 羊肉硬度预测模型

光谱预处理可以有效消除原始光谱中噪声、基线漂移等干扰,是基于光谱数据建立预测模型中不可或缺的步骤。光谱预处理方法可分为两类:基线校正^[11]和散射校正^[12]。基线校正主要以 S-G 多项式的一阶求导和二阶求导为代表,散射校正主要以多元散射校正(MSC)、标准正态变换(SNV)等算法为代表。本研究选取 S-G 平滑、二阶求导、MSC、SNV 等四种光谱预处理方法。

PLSR 是一种有效克服变量间存在多重线性关系的线性回归分析方法。PLSR 的主要原理是采用主成分分析样本的光谱数据矩阵和测量值矩阵,探索测量值与光谱信息之间的关联性,建立基于光谱信息预测测量值的最佳关联函数^[13]。本研究采用 PLSR 算法建立羊肉硬度的定量预测模型。

1.4 模型评估

通过对比分析模型的决定系数(R^2)、均方根误差(RMSE)和剩余预测偏差(RPD)评估模型的性能。 R^2 反映了样本测量值与预测值的相关性,取值范围为[0, 1],计算公式见式(3)

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

式(3)中, y_i 为第 i 个样本的测量值, \bar{y} 为所有样本测量值的平均值, \hat{y}_i 为第 i 个样本的预测值。 R^2 越大,说明模型的

预测能力越好。

RMSE 可以用于评估样本测量值和预测值之间的偏差, 计算方法如式(4)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

RPD 可以通过衡量预测集的验证结果, 对模型性能进行评价, 计算方法如式(5)

$$RPD = \frac{SD}{RMSE_p} \quad (5)$$

式(5)中, SD 为预测集样本测量值的标准偏差。当 RPD 小于 1.5, 说明模型的预测的性能较差; 当 RPD 在 1.5~2 之间, 说明模型可对待测量值的高低; 当 RPD 在 2~2.5 之间, 说明模型可作近似定量计算; 当模型的 PRD 大于 2.5, 说明模型预测能力较高。

2 结果与讨论

2.1 羊肉硬度变化与光谱数据分析

羊肉硬度随时间和温度变化的平均值与标准偏差如图 1 所示。在 0, 4 和 8 °C 环境下贮藏期间, 羊肉的硬度呈现明显的下降趋势。羊肉样本硬度初始值为: 硬度 1 291.128 g, 随着自溶现象的加深, 不同贮藏温度的样本硬度持续下降至初始值的 65.09%, 47.90% 和 41.43%。冷藏过程中羊肉硬度下降的原因是微生物和酶的作用下肌细胞中的蛋白纤维网架体系被破坏, 导致细胞承受外力、保持内部结构的能力

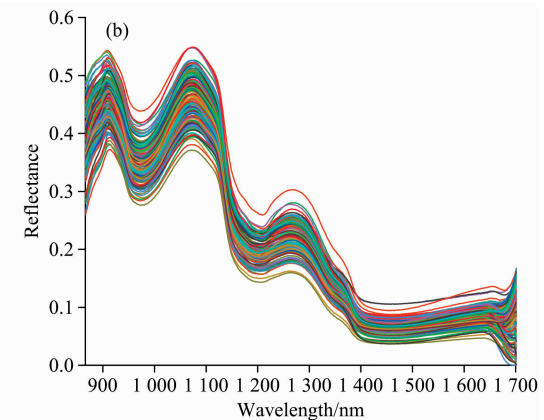
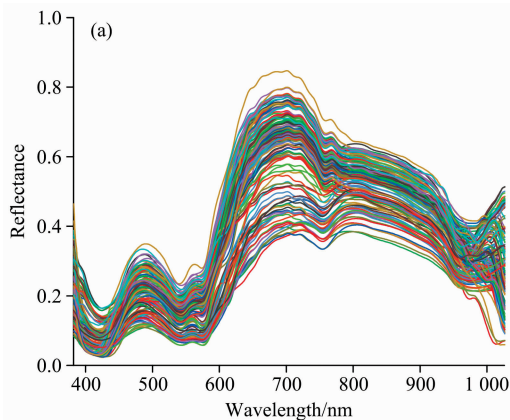


图 2 羊肉样本原始光谱

(a): 400~1 020 nm; (b): 860~1 700 nm

Fig. 2 Original spectrum of mutton sample

(a): 400~1 020 nm; (b): 860~1 700 nm

2.2 羊肉硬度异常样本检测结果

采用二次迭代 MC 抽样法对羊肉硬度的异常样本检测结果如图 3 所示。抽样次数为 5 000 次, 校正集和预测集的划分比例为 3 : 1, 红色虚线和黑色虚线分别表示可疑样本检测阈值和异常样本检测阈值。

图 3(a)和(c)分别为基于 400~950 nm 光谱数据的羊肉硬度异常样本检测中第一次 MC 抽样和第二次 MC 抽样的结果。由图 3(a)可知, 在第一次 MC 抽样中, 样本 3, 4, 83 和

下降^[14]。

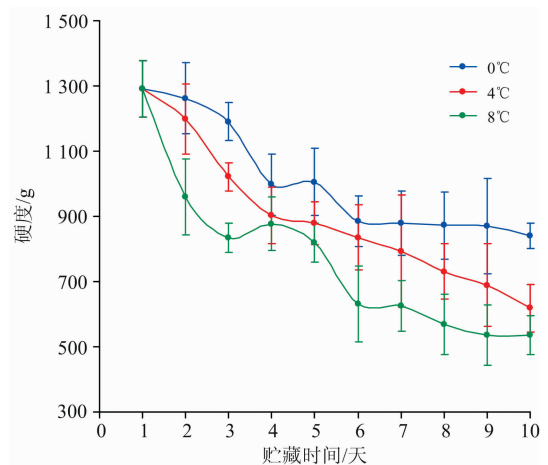


图 1 不同冷藏温度下羊肉硬度的变化情况

Fig. 1 The development trends of muttons' hardness under different storage temperatures

图 2(a,b)分别为羊肉样本在 380~1 020 和 860~1 700 nm 范围的反射光谱。从图中可知, 所有羊肉样本的反射光谱变化趋势基本相同, 在每个波长处的反射强度存在一定差异; 另外, 靠近两端的光谱信号由于干扰噪声的影响产生了明显的基线漂移。为避免光谱曲线抖动、基线漂移等干扰噪声对数据分析和预测模型的影响, 选取 400~950 和 900~1 650 nm 范围的反射光谱信息作为后续研究的基础数据。

109 分布于 2.5~3 倍阈值线之间, 为可疑样本, 样本 2, 35 和 74 分布于与 3 倍阈值线之外, 为异常样本。由表 1 可知, 在剔除 3 个异常样本, 并保留样本 3, 4, 83 和 109 的情况下, 剩余样本集 PLSR 的交叉验证参数为 $R_{cv}^2 = 0.9159$, $RMSE_{cv} = 61.6570$ 。将样本预测误差均值和标准差的平均值的 3 倍设为异常样本阈值, 对剩余样本进行第二次 MC 抽样异常值检测, 如图 3(c)所示。样本 3, 4 和 104 预测误差的均值超出了 3 倍阈值线, 被确定为异常样本。在进一步剔除

样本 3, 4 和 104 后, 剩余样本集 PLSR 的交叉验证参数为 $R_{CV}^2=0.936 2$, $RMSE_{CV}=53.697 4$ 。最后二次检测可疑样本 83 和 109 对模型的影响。剔除样本 83 后, 剩余样本集 PLSR 的交叉验证参数为 $R_{CV}^2=0.940 0$ 和 $RMSE_{CV}=52.090 5$, 模型性能有所提升, 表明样本 83 是异常样本; 剔除样本 109

后, 剩余样本集 PLSR 的交叉验证参数为 $R_{CV}^2=0.933 7$, $RMSE_{CV}=54.739 3$, 模型性能略微降低, 表明样本 109 是正常样本。综上所述, 采用二次迭代 MC 抽样法检测出样本 2, 3, 4, 35, 74, 83 和 104 等 7 个异常样本。

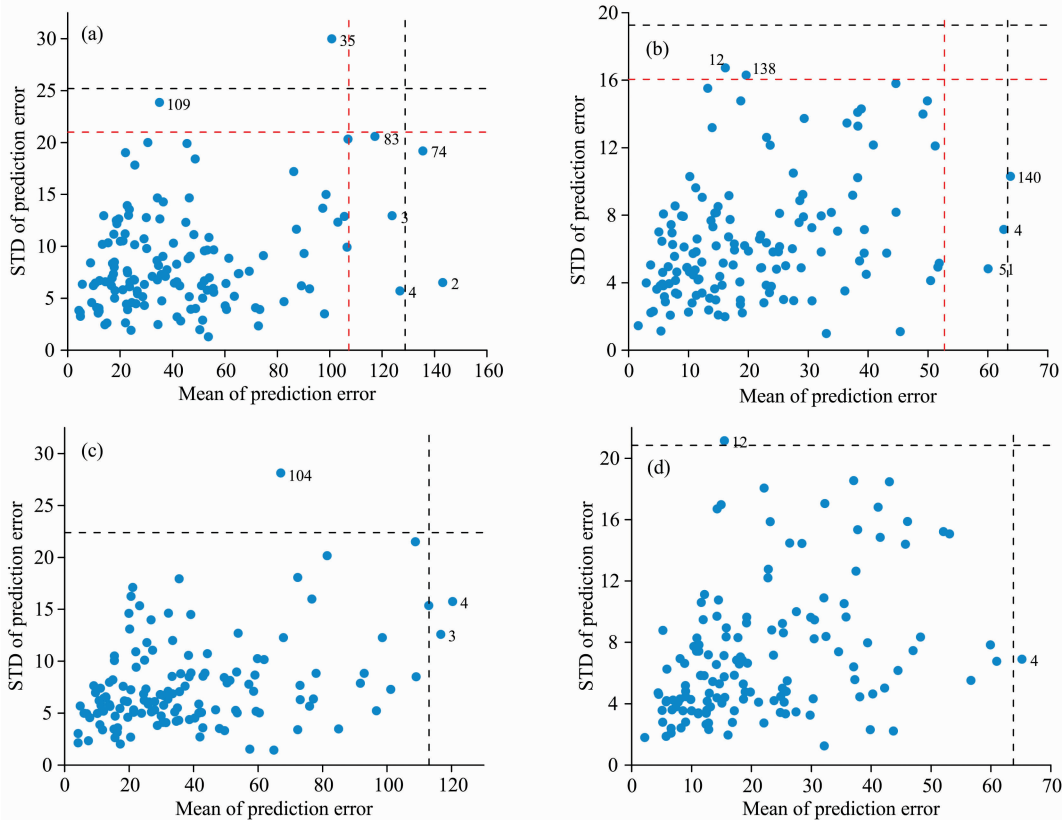


图 3 羊肉硬度异常样本二次迭代 MC 抽样检测结果

(a): 400~950 nm 1st MC; (b): 900~1 650 nm 1st MC; (c): 400~950 nm 2nd MC; (d): 900~1 650 nm 2nd MC

Fig. 3 Mutton hardness outliers' detection result by 2 iteration MC sampling

(a): 400~950 nm 1st MC; (b): 900~1 650 nm 1st MC; (c): 400~950 nm 2nd MC; (d): 900~1 650 nm 2nd MC

表 1 基于二次迭代 MC 抽样的羊肉硬度异常样本检测结果

Table 1 Detection results of mutton hardness outliers based on 2 iteration MC sampling

		400~950 nm			900~1 650 nm		
		样本编号	R_{CV}^2	$RMSE_{CV}$	样本编号	R_{CV}^2	$RMSE_{CV}$
1 st MC	异常样本	2, 35, 74	0.915 9	61.657 0	140	0.957 9	40.613 1
	可疑样本	3, 4, 83, 109	—	—	4, 5, 12, 138	—	—
2 nd MC		3, 4, 104	0.936 2	53.697 4	4, 12	0.964 7	37.191 3
样本二次检测		剔除 83	0.940 0	52.090 5	剔除 138	0.976 9	37.442 8
		剔除 109	0.933 7	54.739 3	剔除 5	0.982 8	37.040 3

图 3(b)和(d)分别为基于 900~1 650 nm 光谱数据的羊肉硬度异常样本检测中第一次 MC 抽样和第二次 MC 抽样的结果。同理分析, 采用二次迭代 MC 抽样法检测出样本 4, 5, 12 和 140 等 4 个异常样本。

2.3 最佳预处理方法选取

最佳预处理方法通过对比分析上述四种方法对 PLSR 模型的影响进而选出, 如表 2 所示。相比基于原始光谱建立的

模型, 采用二阶求导、MSC 和 SNV 预处理后的模型性能都有所提升, 其中经过 MSN 预处理的预测模型 R_p^2 为 0.978 3, $RMSEP$ 为 30.590 1 g, 经 SNV 预处理的预测模型 R_p^2 为 0.972 2, $RMSEP$ 为 30.716 7 g, 对模型性能的提升比较明显。但采用 MSC 预处理的模型无论 R_c^2 还是 R_p^2 都略优于采用 SNV 预处理的模型, 因此选定 MSC 为最佳预处理方法并应用于后续数据分析和模型构建中。

表 2 光谱预处理结果
Table 2 Results of spectral data preprocessing

预处理	400~950 nm				900~1 650 nm			
	R_c^2	RMSE _c	R_p^2	RMSE _p	R_c^2	RMSE _c	R_p^2	RMSE _p
原始光谱	0.945 7	49.565 8	0.934 1	55.716 2	0.966 8	35.784 5	0.964 6	37.590 1
S-G 平滑	0.945 6	49.587 0	0.934 5	55.495 1	0.966 0	36.179 6	0.962 7	34.612 1
二阶求导	0.947 4	48.758 8	0.933 0	56.088 7	0.970 2	34.105 4	0.968 5	31.803 2
MSC	0.959 3	41.595 1	0.947 2	47.789 9	0.979 2	29.660 7	0.978 3	30.590 1
SNV	0.951 9	46.619 1	0.932 6	56.387 2	0.975 7	31.389 2	0.972 2	30.716 7

2.4 PLSR 模型建立与特征波长提取

基于 400~950 和 900~1 650 nm 两组羊肉高光谱数据, 采用留一交叉验证法建立 PLSR 模型, 结果如表 3 所示。基于 400~950 nm 波长建立的羊肉硬度 PLSR 预测模型的性能较好(RPD>2.5), 模型的 $R_p^2=0.947 2$, $RMSE_p=47.789 9$ g。基于 900~1 650 nm 波长建立的羊肉硬度 PLSR 预测模型的 R_p^2 为 0.978 3, $RMSE_p$ 为 30.590 1 g, 明显优于基于 400~950 nm 波长的样本集建立的模型。

高光谱数据的高维度导致预测模型的复杂度较高, 限制了其在实际应用中的时效性^[15]。基于最佳特征波长建立的简化预测模型, 可以在保证模型的预测精度不受较大影响的

同时最大限度的降低高光谱数据的复杂度。本研究采用 RC 方法提取 400~950 和 900~1 650 nm 波长范围内羊肉硬度的特征波长, 提取结果分别如图 4(a,b)所示。通过 RC 算法选取出羊肉硬度在 400~950 和 900~1 650 nm 波长范围的特征波长分别为 14 个和 10 个。基于两组特征波长分别建立羊肉硬度的 PLSR 模型, 结果如表 3 所示。与基于全波长建立的羊肉硬度模型相比, 基于特征波长建立的 PLSR 模型 R_p^2 分别为 0.951 9 和 0.985 0, $RMSE_p$ 分别为 46.602 7 g 和 24.397 0 g, 预测能力更高, 预测结果分别如图 5(a,b)所示。其中, 基于 900~1 650 nm 波长建立的 PLSR 模型为羊肉硬度预测的最佳模型。

表 3 羊肉硬度 PLSR 模型的预测结果
Table 3 Mutton hardness prediction results of PLSR models

	模型	LVs	校正集		交叉验证集		预测集		
			R_c^2	RMSE _c	R_{cv}^2	RMSE _{cv}	R_p^2	RMSE _p	RPD
400~950 nm	全波长	8	0.959 3	41.595 1	0.950 0	46.357 3	0.947 2	47.789 9	4.350
	RC	7	0.966 0	38.039 7	0.958 2	42.780 8	0.951 9	46.602 7	4.425
900~1 650 nm	全波长	8	0.979 2	29.660 7	0.978 5	29.855 3	0.978 3	30.590 1	6.717
	RC	6	0.987 3	22.418 6	0.986 6	23.480 9	0.985 0	24.397 0	7.524

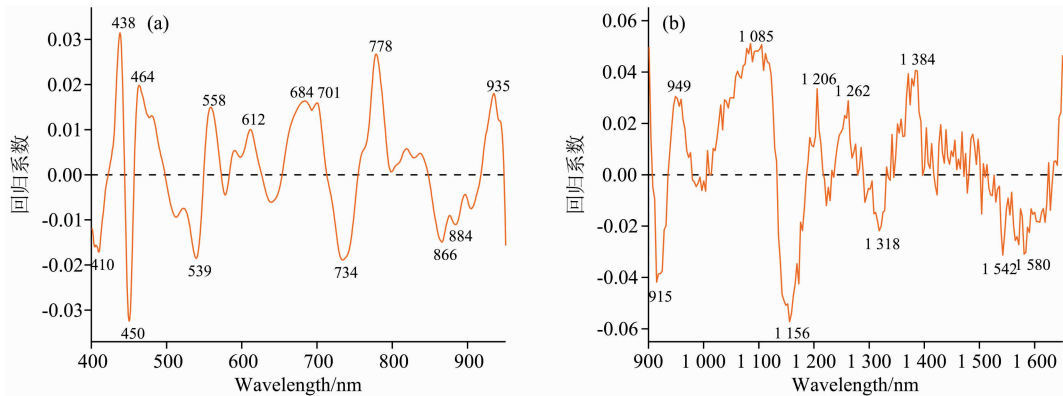


图 4 采用 RC 法提取羊肉硬度的特征波长
(a): 400~900 nm; (b): 900~1 650 nm

Fig. 4 Optimal wavelengths of mutton's hardness by RC

(a): 400~900 nm; (b): 900~1 650 nm

3 结 论

探讨了采用 400~950 和 900~1 650 nm 两段高光谱数

据预测羊肉冷藏期间硬度变化的可行性, 具体结论如下:

(1)本研究提出的二次迭代 MC 方法, 将羊肉样本划分为正常样本、可疑样本和异常样本。通过第一次 MC 抽样剔除异常样本, 然后通过调整异常样本阈值进行第二次 MC 抽

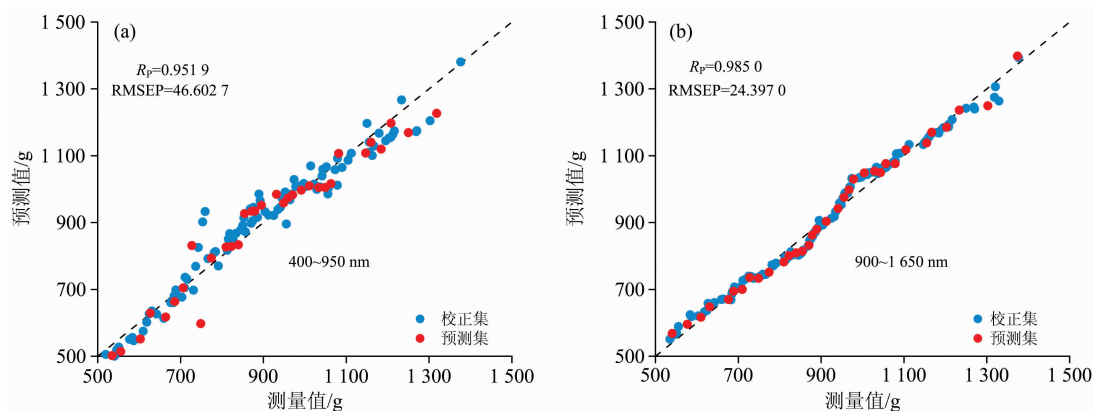


图 5 羊肉硬度 PLSR 模型预测结果

(a): 400~950 nm 时的预测结果; (b): 900~1650 nm 时的预测结果

Fig. 5 Mutton's hardness prediction results of PLSR

(a): 400~950 nm prediction results; (b): 900~1650 nm prediction results

样,进一步剔除异常样本,最后对可疑样本进行二次判定。实验结果证明该方法能够准确的辨别异常样本,有助于提高建模的准确性和稳定性。

(2)研究发现, MSC 预处理方法能有效消除噪声的影响,提高模型的性能。基于 400~950 和 900~1650 nm 两段高光谱数据,采用留一交叉验证法分别建立了羊肉硬度 PLSR 模型,模型的 R_p^2 为 0.947 2 和 0.978 3, $RMSEP_p$ 为

47.789 9 g 和 30.590 1 g。

(3)采用 RC 方法分别从 400~950 和 900~1650 nm 两段高光谱数据内提取出羊肉硬度特征波长为 14 个和 10 个。根据对比分析,基于特征波长建立的 PLSR 模型的预测能力更高。其中,基于 900~1650 nm 波长建立的 PLSR 模型的 R_p^2 为 0.985 0, $RMSEP_p$ 为 24.397 0 g,为羊肉硬度预测的最佳模型。

References

- [1] Xiong Z, Sun D, Zeng X, et al. Journal of Food Engineering, 2014, 132: 1.
- [2] Pu H, Kamruzzaman M, Sun D. Trends in Food Science & Technology, 2015, 45(1): 86.
- [3] Sanchez B G, Bowker B C, Zhuang H. J. Poult Sci., 2016, 95(10): 2472.
- [4] Cheng J H, Sun D W, Han Z, et al. Comprehensive Reviews in Food Science & Food Safety, 2013, 1(13): 52.
- [5] Cheng J, Qu J, Sun D, et al. Food Research International, 2014, 56: 190.
- [6] Cheng J H, Sun D W, Pu H, et al. Food & Bioprocess Technology, 2014, 11(7): 3109.
- [7] Tao F, Peng Y, Li Y, et al. Meat Science, 2012, 90(3): 851.
- [8] Kamruzzaman M, ElMasry G, Sun D, et al. Food Chemistry, 2013, 141(1): 389.
- [9] LIU Gui-shan, ZHANG Chong, FAN Nai-jun, et al(刘贵珊, 张 翀, 樊奈均, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2020, 40(8): 2558.
- [10] Jun L, Yutong F, Weijian L, et al. IEEE Transactions on Signal Processing, 2020, 68(5): 3022.
- [11] Lobanova E G, Lobanov S V. Analytica Chimica Acta, 2019, 1050: 32.
- [12] Reis M M, Van Beers R, Al-Sarayreh M, et al. Meat Science, 2018, 144: 100.
- [13] Xinhua J, Heru X, Lina Z, et al. Computers and Electronics in Agriculture, 2018, 155: 371.
- [14] Estelles-Lopez L, Ropodi A, Pavlidis D, et al. Food Res. Int., 2017, 99(1): 206.
- [15] Xiong Z, Sun D, Xie A, et al. Food Chemistry, 2015, 175: 417.

Quantitative Detection of Mutton Hardness Based on Twice Iterative Monte Carlo Method

BAI Xue-bing¹, LI Xin-xing¹, ZHANG Xiao-shuan², LUO Hai-ling³, FU Ze-tian^{2*}

1. Beijing Laboratory of Food Quality and Safety, College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. College of Engineering, China Agricultural University, Beijing 100083, China

3. State Key Laboratory of Animal Nutrition, College of Animal Science and Technology, China Agricultural University, Beijing 100083, China

Abstract Mutton, as a kind of meat with high protein content and low fat and cholesterol content, is becoming more and more popular with consumers. The demand for mutton is on the rise. According to the National Bureau of Statistics, China's mutton production rose from 6.27% to 9.02% from 2012 to 2019. This study proposed a quantitative detection PLSR model of mutton hardness based on the twice iterative Monte Carlo (MC) method. In this study, the Image- λ -V10E-H camera of the GaiaSorter hyperspectral sorter was used to collect the hyperspectral data of mutton samples at 400~950 nm, and the Image- λ -N17E camera was used to collect the hyperspectral data of mutton samples at 900~1 650 nm. Firstly, the study compared and analyzed four spectral pretreatment methods (S-G smoothing, 2 derivations, MSC and SNV) in eliminating interference factors, such as noise and baseline drift. Then, in the first MC sampling, the samples were divided into normal samples, suspicious samples and abnormal samples according to the 2.5 and 3 times the means of the prediction error means and standard deviations of each sample. The second MC sampling was performed based on rejecting abnormal samples, retaining and labeling suspicious samples. The new abnormal samples were eliminated by 3 times of the means of the prediction error means and standard deviations of each sample. Finally, the PLSR model based on the full wavelengths and the characteristic wavelengths extracted by the regression coefficient method (RC) were established and analyzed. The experiment results show that the twice iterative Monte Carlo method proposed in the study could abnormal samples, optimize the sample set, and provide a good foundation for modeling. With MSC as the spectral preprocessing algorithm, the PLSR model based on 400~950 and 900~1 650 nm hyperspectral data was superior to the other three spectral preprocessing algorithms $R_p^2 = 0.947\ 2$ and $0.978\ 3$, $RMSE_p = 47.789\ 9\ g$ and $30.590\ 1\ g$. And, the accuracy and stability of the PLSR model based on 900~1 650 nm were significantly better than that based on 400~950 nm. 14 characteristic wavelengths (410, 438, 450, 464, 539, 558, 612, 684, 701, 734, 778, 866, 884, 935 nm) and 10 characteristic wavelengths (915, 949, 1 085, 1 156, 1 206, 1 262, 1 318, 1 384, 1 542 and 1 580 nm) of mutton hardness were selected by RC algorithm from 900~1 650 and 400~950 nm. The PLSR model based on 900~1 650 nm was the optimal model for predicting the hardness of mutton with $R_p^2 = 0.985\ 0$ and $RMSE_p = 24.397\ 0\ g$. In conclusion, the PLSR model based on the twice iterative MC algorithm can effectively predict the changing trend of mutton hardness during cold storage and provide a reference for related research on non-destructive detection of mutton quality.

Keywords Visible-near infrared spectroscopy; Twice iterative Monte Carlo; PLSR; Mutton hardness

(Received Jul. 19, 2020; accepted Dec. 6, 2020)

* Corresponding author