

## 近红外特征光谱的羊肉 TVB-N 浓度预测模型

张旭<sup>1</sup>, 白雪冰<sup>1</sup>, 汪学沛<sup>2</sup>, 李新武<sup>2</sup>, 李志刚<sup>3</sup>, 张小栓<sup>2,4\*</sup>

1. 中国农业大学信息与电气工程学院, 北京 100083
2. 中国农业大学工学院, 北京 100083
3. 石河子大学信息科学与技术学院, 新疆 石河子 832003
4. 食品质量与安全北京实验室, 北京 100083

**摘要** 为提高生鲜羊肉储存期内(4, 8 和 20 °C 环境)挥发性盐基氮(TVB-N)的近红外光谱(NIR)检测的稳定性和准确性, 选取特征光谱和预测模型是关键步骤。以 121 个羊肉样品为实验对象, 采集生鲜羊肉 680~2 600 nm 波段的近红外光谱。以多元散射校正(MSC)、标准正态变换(SNV)等散射校正方法, Savitzky-Golay 卷积平滑(SGS)、移动平均平滑(MAS)等平滑处理方法, 以及归一化(Normalization)、中心化(Centering)、标准化(Autoscaling)等尺度缩放方法分别预处理光谱数据后建立偏最小二乘法(PLS)预测模型。比较发现 SGS 处理的光谱建模效果最好。利用蒙特卡罗采样(MCS)法及马氏距离法(MD)消除了羊肉光谱的 5 个异常数据。运用光谱-理化值共生距离(SPXY)算法划分总样本的 75%(87 个)为校正集样本, 剩余 29 个为验证集样本, 利用竞争性自适应重加权法(CARS)、无信息变量消除法(UVE)、改进的无信息变量消除法(IUVE)和连续投影算法(SPA)提取特征光谱得到的波长个数分别为 14, 713, 144 和 15。将全光谱和 4 种方法提取的特征波长作为输入变量建立预测模型, CARS 提取的波长所建立模型的性能优于 UVE、IUVE 和 SPA 提取的波长所建立模型的性能, 表明 CARS 方法可以有效简化输入变量并提高预测模型的性能。改进后得到的 IUVE 法相比于 UVE 法, 筛选出的波长数更少且模型性能有所提升。以提取的特征波长建立 PLS, 支持向量机(SVM)和最小二乘支持向量机(LS-SVM)预测模型, SVM 模型得到最优的校正集预测结果, 其中 CARS-SVM 预测模型的校正决定系数( $R_c^2$ )和校正均方根误差(RMSEC)分别为 0.939 1 和 1.426 7, 最优的验证集预测效果为 LS-SVM 预测模型得到, 其中 IUVE-LS-SVM 预测模型的验证决定系数( $R_v^2$ )和验证均方根误差(RMSEV)分别为 0.856 8 和 1.886 2。基于近红外特征光谱建立简化、优化的生鲜羊肉储存期 TVB-N 预测模型, 为实现快速无损检测生鲜羊肉中的 TVB-N 浓度提供技术支持。

**关键词** 近红外光谱; 挥发性盐基氮; 特征光谱; 偏最小二乘法; 支持向量机

**中图分类号:** O657.33 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)11-3377-08

### 引言

中国的羊肉产量及羊肉消费量均居世界首位。保证储存期内羊肉的新鲜度和防范质量安全问题愈发紧要。在羊肉的贮藏过程中, 在微生物和内外源酶的作用下, 羊肉中的脂肪和蛋白质分解产生有毒的氨( $\text{NH}_3$ )和胺类( $\text{R-NH}_2$ )<sup>[1]</sup>, 并与腐败产生的有机酸结合生成挥发性盐基氮(total volatile basic nitrogen, TVB-N), 因此 TVB-N 浓度是评估羊肉质量的关键参数。

食品安全国家标准中规定的 TVB-N 浓度的检测方法包括自动凯氏定氮仪法、半微量定氮法、微量扩散法。这些化学检测方法需要破坏样品, 操作过程复杂, 耗时费力, 且结果易受操作水平影响, 不能满足快速、非破坏的质量安全检测要求。

近红外光谱(near infrared spectroscopy, NIR)检测技术具有分析快速、操作简便、无破坏性等特点, 在肉品质检测领域已有大量研究。被应用于新鲜度分级<sup>[2]</sup>、掺假识别<sup>[3]</sup>、等级划分<sup>[4]</sup>、鲜冻肉鉴别等定性分析, 以及化学组成(包括胆固醇<sup>[5]</sup>、脂肪<sup>[6]</sup>、水分<sup>[7]</sup>)分析、感官品质(包括肉色<sup>[8]</sup>、

收稿日期: 2020-09-01, 修订日期: 2021-01-11

基金项目: 科学技术部重点研发计划项目(2017YFE0111200)资助

作者简介: 张旭, 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: zhxu@cau.edu.cn

\* 通讯作者 e-mail: zhxshuan@cau.edu.cn

系水力、嫩度<sup>[8]</sup>)评价等定量分析。

原始近红外光谱数据中虽然包含与特定成分相关的有效信息,但也受到噪声及散射等因素的干扰,这些干扰会降低光谱模型的预测性能。因此,应用适当的光谱预处理方法(散射校正、平滑处理、尺度缩放)和变量筛选方法可有效消除与被测指标无关的噪声、散射等干扰,提高光谱与被测指标间的相关性。将竞争性自适应重加权法(competitive adaptive reweighted sampling, CARS)<sup>[9]</sup>、无信息变量消除法(uninformative variable elimination, UVE)<sup>[9]</sup>和连续投影算法(successive projections algorithm, SPA)等方法应用于简化和优化近红外光谱预测模型已有报道。

在回归预测分析中,偏最小二乘(partial least squares, PLS)是被广泛应用的线性回归方法,当被测指标与光谱数据间存在非线性关系时,采用线性的回归方法无法实现光谱信息的充分提取,影响模型准确性。为解决光谱数据的非线性问题,非线性算法包括支持向量机(support vector machine, SVM)和偏最小二乘支持向量机(least squares-support vector machine, LS-SVM)已被用于近红外光谱预测模型的构建。为此,探讨近红外光谱的线性及非线性预测模型对稳定可靠的定量分析非常必要。

本研究以反映生鲜羊肉质量安全的 TVB-N 浓度为预测对象,采集样品 680~2 600 nm 的近红外光谱数据,经 2 种方法剔除离散程度大的异常样本后,对比多种预处理方法对预测模型性能的影响,以不同变量筛选方法优选特征波长,探讨线性及非线性建模方法的预测性能,建立优化的生鲜羊肉储存期 TVB-N 预测模型,为实现快速无损检测生鲜羊肉中的 TVB-N 浓度提供参考和技术支持。

## 1 实验部分

### 1.1 材料与仪器

在江苏省东台市华东山羊市场购买当天屠宰的绵羊,取背最长肌,切除其脂肪和肌膜,整形切成 3 cm×3 cm×1 cm (长×宽×高)的块状 150 个,置入无菌袋并以 4 °C 冷链运输车在 24 h 内运抵中国农业大学工学院实验室,将样品置于生化培养箱(LRH-250,上海一恒仪器公司)中,分组后在 0, 4, 8 和 20 °C 温度下储藏。低温试验时每隔 24 h 各取出 3 个样品进行测试,20 °C 下每隔 12 h 取 2 块样本进行测试,共测试 11 d,共获得 121 份样本。

### 1.2 光谱数据获取

采用 SpectraStar 2600 XT-R 型近红外光谱仪(美国 Unity Scientific 公司)采集羊肉样本光谱,扫描次数为 12 次,扫描范围为 680~2 600 nm,分辨率为 1 nm。

### 1.3 挥发性盐基氮(TVB-N)浓度测定

根据 GB 5009.228—2016《食品安全国家标准食品中挥发性盐基氮的测定》中的微量扩散法测定样品的 TVB-N 浓度。

### 1.4 数据处理与建模方法

通过 MATLAB R2018b 软件(美国 Mathworks 公司)完成数据处理和模型构建。

#### 1.4.1 异常值剔除方法

利用蒙特卡罗采样(Monte-Carlo sampling, MCS)法消除异常样本。异常剔除阈值设置为式(1)和式(2)所示,超过阈值之一的即为异常样本。

$$M_{\text{threshold}} = \mu_M + 3\sigma_M \quad (1)$$

$$S_{\text{threshold}} = \mu_S + 3\sigma_S \quad (2)$$

其中, $M_{\text{threshold}}$ 为各样本预测误差均值的阈值, $\mu_M$ 和 $\sigma_M$ 为各样本预测误差均值的均值和标准差; $S_{\text{threshold}}$ 为各样本预测误差标准差的阈值, $\mu_S$ 和 $\sigma_S$ 为各样本预测误差标准差的均值和标准差。

同时利用马氏距离法(Mahalanobis distance, MD)剔除 MD 过大的样本,阈值设置为

$$MD_{\text{threshold}} = \mu + 3\delta \quad (3)$$

式(3)中, $MD_{\text{threshold}}$ 为各样本 MD 的阈值, $\mu$ 和 $\delta$ 为各样本 MD 的均值和标准差。

#### 1.4.2 样本集划分方法

剔除异常样本之后,运用光谱-理化值共生距离(sample set partitioning based on joint x-y distance, SPXY)算法划分出 75% 的样本为校正集,其余为验证集样本。

#### 1.4.3 光谱预处理方法

根据处理效果预处理方法可分为散射校正、平滑处理、尺度缩放等。散射校正包括多元散射校正(multiple scattering correction, MSC)和标准正态变换(standard normal variate, SNV),用于消除样品颗粒尺寸差异和分布差异对漫反射光的影响。平滑处理可有效提高信噪比,常用的有 Savitzky-Golay 卷积平滑(S-G smoothing, SGS)、移动平均平滑(moving average smoothing, MAS)等方法。尺度缩放包括归一化(Normalization)、中心化(Centering)、标准化(Autoscaling)等,用来消除数据尺度差异的影响。分别采用这 7 种方法对样本进行光谱预处理,以找到最佳预处理方法。

#### 1.4.4 特征波长筛选方法

所采集羊肉光谱共 1 921 个波长,存在冗余和多重共线性信息,筛选特征波长取代全光谱可提高模型简洁性和计算效率。

CARS 选择波数的方法是基于回归系数的权重,权重值越大则代表该变量对模型建立的贡献越大,被选取的概率越大。

UVE 以输入变量及等量的随机噪声建立 PLS 模型得到回归系数矩阵,计算各变量的稳定性并筛选稳定性大的变量,阈值设为随机变量稳定性最大绝对值的 0.99 倍。UVE 选出的波长呈局部连续分布,波段之间仍存在严重的多重共线性问题,岭回归(ridge regression, RR)法是解决此类问题的有效方法,但是鲜有研究将岭回归用于光谱检测中的特征波长筛选,由此在 UVE 法的基础上引入岭回归法得到改进的无信息变量消除(improved uninformative variable elimination, IUVE)法,以进一步简化模型和提高预测精度。

SPA 以正交投影分析全部波长变量,并保留对 TVB-N 敏感的特征变量,使变量之间共线性达到最小,降低模型输入量。

#### 1.4.5 预测建模方法

PLS 在主成分分析的基础上,对光谱和理化值同时进行分解,保留对光谱贡献大的主成分,进而构建误差最小化的最佳线性回归模型。

SVM 模型利用核函数将低维输入映射到高维特征空间,并在高维特征空间进行线性回归,适于处理小样本、非线性以及高维数等问题。LS-SVM 是 SVM 的改进方法,进一步降低计算复杂性和提高计算速度。设定径向基函数(radial basis function, RBF)为 SVM 模型及 LS-SVM 模型的核函数,以具备交互验证的网格搜索(grid-search)法对 SVM 模型的关键参数( $C, \gamma$ ),以及 LS-SVM 模型的关键参数( $\gamma, \sigma^2$ )进行寻优。

选择校正决定系数  $R_c^2$ 、验证决定系数  $R_v^2$ 、校正均方根误差 RMSEC、验证均方根误差 RMSEV 为模型性能评估指标。

## 2 结果与讨论

### 2.1 剔除异常样本

图 1 为 121 个羊肉样本的原始近红外漫反射吸光度光谱。由于水分子中 O—H 键伸缩振动的二级倍频和一级倍频吸收,在 980, 1 440 和 1 940 nm 附近呈现出吸收峰,在 1 200 和 2 400 nm 附近则是与 C—H 键拉伸和伸缩振动相关的波峰<sup>[8]</sup>。样本的光谱曲线趋势相似,但不同波段的上升和下降的趋势不同,说明其内部化学成分存在差异。

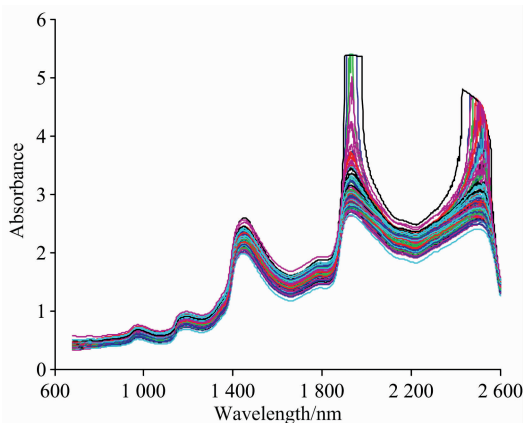


图 1 羊肉样本近红外光谱图

Fig. 1 The NIR spectra of mutton samples

采用 MCS 法识别异常样本,主成分个数设置为 11,预处理方法为 Centering,抽样次数 1000 次。异常样本检测结果见图 2,得到阈值  $M_{\text{threshold}}$  为 6.4,  $S_{\text{threshold}}$  为 2.198,第 47, 89, 94, 103 和 121 个样本被判定为异常样本。

图 3 是 121 个光谱样本到平均光谱的 MD 分布图,由 MD 的均值和标准偏差得到阈值为 5.596,超出此阈值的异常样本为 47 号,89 号,94 号,103 号,与 MCS 法的检测结果重合度较高。

将剔除异常值后的 116 个样本,采用 SPXY 算法划分校正集和验证集样本。羊肉 TVB-N 浓度的统计分析结果如表 1 所示。

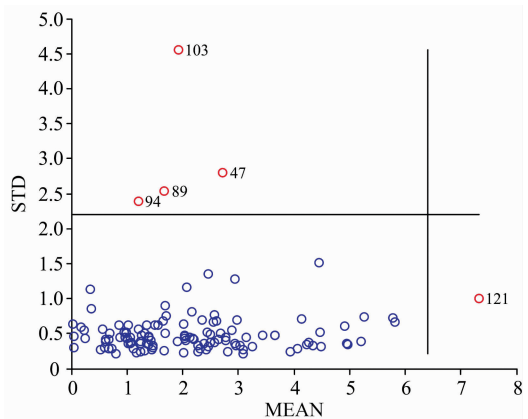


图 2 蒙特卡洛采样法异常值检测结果

Fig. 2 The outlier results detected by Monte Carlo sampling method

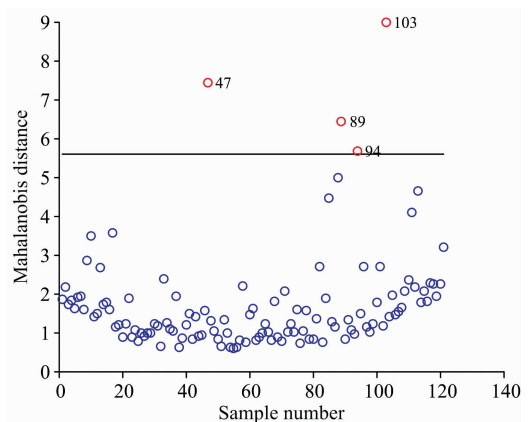


图 3 羊肉样本的马氏距离分布

Fig. 3 Distribution of Mahalanobis distance of mutton samples

表 1 羊肉样品 TVB-N 浓度的统计结果

Table 1 Statistical results of TVB-N concentration of mutton samples

样本集	样本数	TVB-N 含量/(mg · 100 g <sup>-1</sup> )			
		最大值	最小值	平均值	标准差
校正集	87	26.58	4.81	14.51	5.82
验证集	29	23.26	5.55	12.66	5.07

### 2.2 原始光谱及光谱预处理比较

分别使用原始光谱及经过 7 种方法预处理的光谱建立全波段的 PLS 预测模型,建模结果见表 2。与原始光谱所建立的 PLS 模型相比,由 SNV, MSC, Autoscaling 处理的光谱数据建立的模型性能均下降,经 Normalization, Centering, MAS 处理的光谱数据所建立模型的性能没有明显改善,而以 SGS 预处理的数据建模效果最好,确定 SGS 为最优预处理方法。

### 2.3 特征波长的选取

#### 2.3.1 应用 CARS 筛选特征波长

采用 CARS 提取特征波长, 设置蒙特卡洛采样次数为 50, 采用 7 折交叉验证计算。随着采样次数增加, 图 4(a) 曲线呈指数衰减, 在运行次数 1~5 次, 变量选择个数曲线快速下降, 对应粗选过程, 之后进入缓慢递减的细选过程。图 4(b) 为交互验证均方根误差的变化趋势图, 在运行次数 1~36, 交互验证均方根误差缓慢波动降低, 随后逐渐升高。从图 4(c) 回归系数曲线中的“\*”标出了交互验证误差的最低点, 在采样次数为 36 次时, 达到最小值 2.206。此时变量筛选个数为 14 个, 分别为 720, 725, 823, 834, 925, 1 162, 1 230, 1 278, 1 441, 1 473, 1 867, 1 981, 2 484 和 2 554 nm。

表 2 不同预处理方式的 PLS 预测模型比较  
Table 2 Comparison of PLS prediction models with different pretreatment methods

预处理方法	主成分	校正集		验证集	
		$R_c^2$	RMSEC	$R_v^2$	RMSEV
RAW	11	0.848 1	2.253 4	0.808 5	2.180 9
SNV	9	0.710 3	2.509 7	0.654 6	3.207 9
MSC	12	0.775 5	2.293 8	0.696 6	3.296 2
Normalization	10	0.839 3	2.317 7	0.798 5	2.237 3
Centering	9	0.840 6	2.308 6	0.814 8	2.144 8
Autoscaling	8	0.705 1	2.515 2	0.694 0	3.290 2
MAS	12	0.839 9	2.313 7	0.814 6	2.146 0
SGS	11	0.850 2	2.211 5	0.824 6	2.105 8

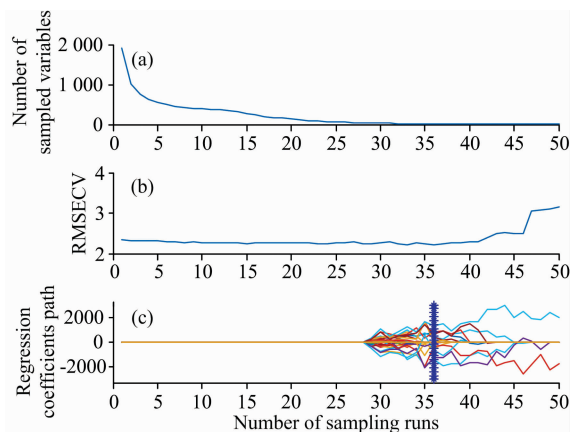


图 4 基于 CARS 的变量选择过程

Fig. 4 Characteristic wavelength selection process using CARS

2.3.2 应用 IUVE 筛选特征波长

运行 IUVE 算法计算 1 921 个光谱波长和等量随机噪声的稳定性, 阈值设为 ±11.76, 将超过阈值的输入变量进行岭回归分析, 以岭迹法确定岭回归参数 k 值为 0.2, 变量筛选的原则参考文献[10]。岭回归分析的结果如图 5 所示, 可以看出各回归系数的岭估计在 k=0.2 时基本稳定, 根据回归系数随 k 值的变化趋势结合误差选择变量。IUVE 与未改进的 UVE 的波长选择结果如图 6 所示, 蓝色的曲线为未改进的 UVE 法计算出的各变量稳定性曲线, 超过阈值的波长达到 703 个, 占总波长的 36.60%, 洋红色竖线为结合岭回

归分析的 IUVE 法最终选择出的 144 个有效变量, 占总波长的比值降至 7.50%, 因此改进后的 UVE 可有效地消除各波长变量间的共线性。

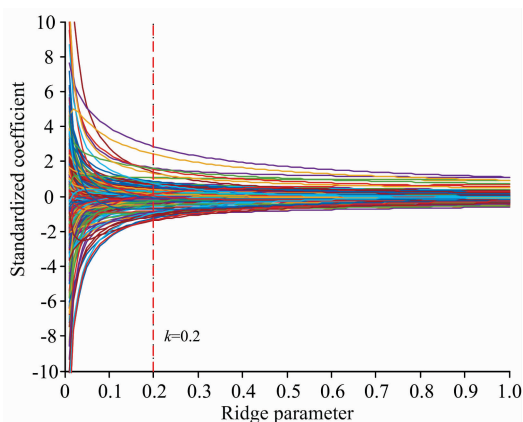


图 5 岭回归分析岭迹图

Fig. 5 Ridge trace of ridge regression analysis

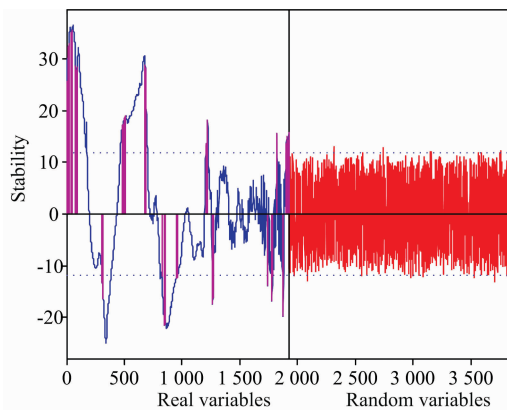


图 6 基于 IUVE 的特征波长筛选

Fig. 6 Characteristic wavelength selection using IUVE

2.3.3 应用 SPA 筛选特征波长

在变量个数 1~28 的范围内优选波长, SPA 算法以 RMSE 的大小为依据确定特征波长数量。随着特征波长数量的增加, RMSE 的变化过程如图 7 所示。当波长数量由 1 增加到 7 时, RMSE 迅速下降, 表明此类波长变量为与羊肉 TVB-N 相关的重要波长变量。当波长数量由 7 个增加到 15 个时 RMSE 呈波动式下降, 此类波长为有用信息变量。随着波长数量由 15 个继续增加, RMSE 继续缓慢下降。因此以 15 个特征波长作为输入的特征变量。图 8 为 15 个特征波长在全光谱中的位置分布, 分别为 680, 798, 1 067, 1 266, 1 497, 1 498, 1 901, 1 920, 1 936, 2 009, 2 263, 2 386, 2 391, 2 575 和 2 583 nm。

2.4 模型建立

2.4.1 PLSR 模型

根据 CARS, UVE 及 SPA 的变量选择结果, 应用 PLS 方法分别建立羊肉 TVB-N 预测模型, 表 3 为各模型预测结果。由表 3 可知, CARS-PLS 模型的性能优于全光谱 PLS 模型, 校正集  $R_c^2$  由 0.840 2 上升到 0.890 0, 验证集  $R_v^2$  由

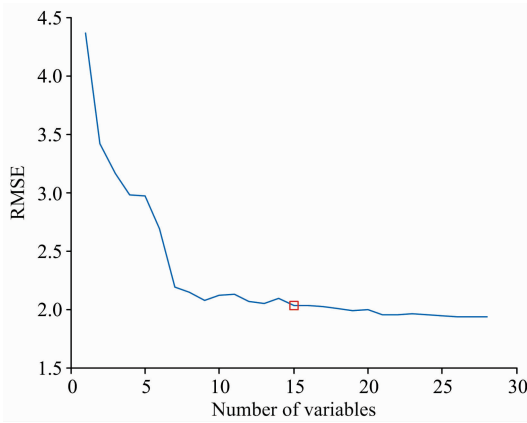


图 7 基于 SPA 的波长个数选择  
Fig. 7 Wavelength selection based on SPA

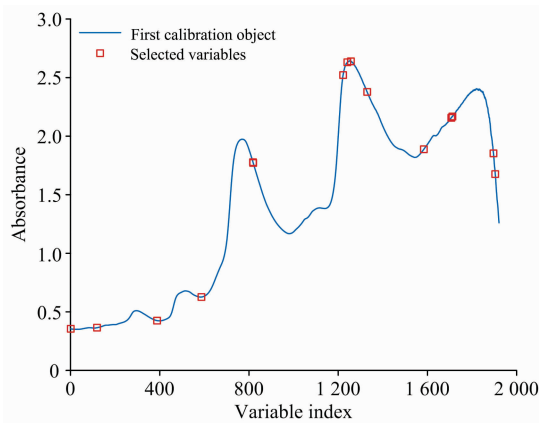


图 8 基于 SPA 的特征波长选择  
Fig. 8 Characteristic wavelength selection based on SPA

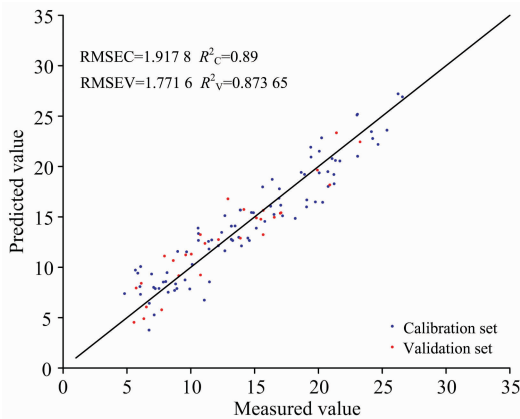


图 9 CARS-PLS 模型对羊肉 TVB-N 浓度的预测结果  
Fig. 9 Prediction results of TVB-N concentration in mutton by CARS-PLS model

0.814 6 上升到 0.873 6。图 9 为 CARS-PLS 模型对羊肉 TVB-N 浓度的预测结果。相比于全光谱 PLS 模型性能, IUVE-PLS 及 UVE-PLS 模型的性能与之接近, SPA-PLS 模型的性能比其略低, 但建模所用的波长变量数仅占全光谱的

0.78%, 数量大幅降低。IUVE-PLS 相比于改进前的 UVE-PLS 性能略有提升, 同时筛选的波长个数进一步降低。根据孙通等<sup>[11]</sup>的研究结果可以看出, 各筛选方法中 UVE 选择的特征波长数量显著高于 SPA 和 CARS, 本研究与其一致, 表明了 UVE 方法适合于粗选特征波长。同时, 各筛选方法所选波长的建模效果由高及低依次为 CARS, UVE 和 SPA, 本研究与其相符, 表明了 CARS 方法可有效地提高模型预测精度。

表 3 不同波长提取方法的 PLS 预测模型比较  
Table 3 Comparison of PLS models with different wavelength selection methods

特征波长提取方法	变量个数	主成分	校正集		验证集	
			$R_c^2$	RMSEC	$R_v^2$	RMSEV
Full spectrum	1 921	9	0.840 2	2.311 5	0.814 6	2.145 8
CARS	14	9	0.890 0	1.917 8	0.873 6	1.771 6
UVE	703	9	0.834 4	2.353 0	0.816 7	2.133 5
IUVE	144	9	0.844 8	2.278 0	0.823 7	2.092 3
SPA	15	10	0.826 1	2.411 4	0.798 4	2.237 7

### 2.4.2 SVM 模型

以全光谱和 3 种不同波长筛选方法获得的特征波长建立 SVM 预测模型, 建模预测结果见表 4。对比分析表 4, 由 CARS 筛选的 14 个变量输入 SVM 模型取得了最好的校正集预测结果,  $R_c^2$  和 RMSEC 分别为 0.939 1 和 1.426 7。图 10 为 CARS-SVM 模型对羊肉 TVB-N 浓度的预测结果。在验证集的预测结果中, 全波长 SVM 模型则取得了最高的  $R_v^2$  为 0.826 1, 以及最低的 RMSEV 为 2.078 2。相比于 PLS 预测模型, SVM 模型的预测效果均有所上升。

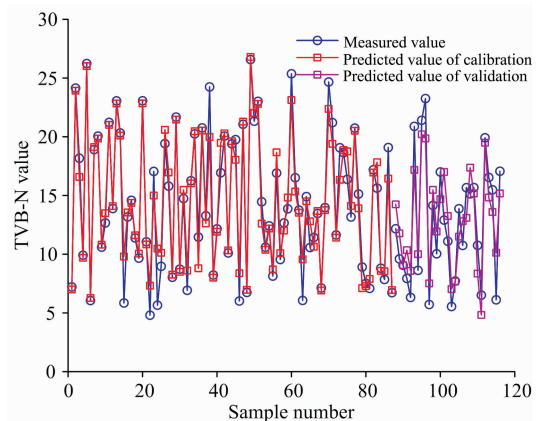


图 10 CARS-SVM 模型对羊肉 TVB-N 浓度的预测结果  
Fig. 10 Prediction results of TVB-N concentration in mutton by CARS-SVM model

### 2.4.3 LS-SVM 模型

利用变量筛选得到的特征波长建立 LS-SVM 非线性预测模型, 预测结果如表 5 所示。分析表 5 可知, CARS-LS-SVM 模型的  $R_c^2$  为 0.923 5, RMSEC 为 1.599 0, 较其他模

型最优, IUVE-LS-SVM 模型的  $R^2_V$  和 RMSEV 较其他模型最优, 分别为 0.856 8 和 1.886 2。图 11 为 CARS-LS-SVM 模型对羊肉 TVB-N 浓度的预测结果。对比分析表 5 和表 4 可知, 与 SVM 模型相比, LS-SVM 模型对于校正集样本的预测性能较低, 对于预测验证集样本的  $R^2_V$  更高, RMSEV 更低。在 Thomas<sup>[12]</sup> 和 Balabin<sup>[13]</sup> 等的研究结果中 SVM 与 LS-SVM 模型的校正与验证效果各有优劣, 结合本研究结果, 表

表 4 不同波长提取方法的 SVM 预测模型比较

Table 4 Comparison of SVM models with different wavelength selection methods

特征波长 提取方法	参数 (C)	参数 ( $\gamma$ )	校正集		验证集	
			$R^2_C$	RMSEC	$R^2_V$	RMSEV
Full spectrum	16.000 0	0.003 9	0.926 8	1.565 0	0.826 1	2.078 2
CARS	48.502 9	1.000 0	0.939 1	1.426 7	0.801 6	2.219 7
UVE	9.189 6	0.006 8	0.900 5	1.824 2	0.823 3	2.095 2
IUVE	1.741 1	0.011 8	0.916 1	1.674 9	0.824 1	2.090 1
SPA	256.000 0	0.189 5	0.881 3	1.992 1	0.791 9	2.273 7

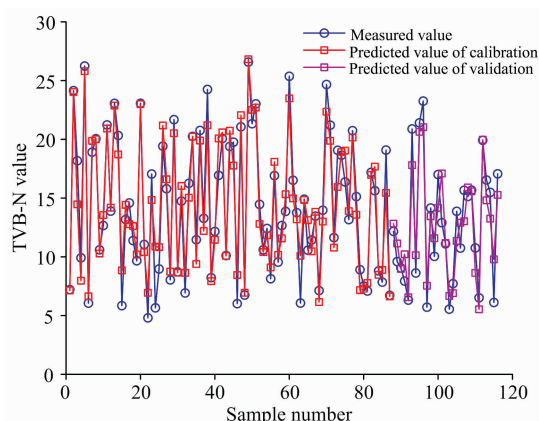


图 11 CARS-LS-SVM 模型对羊肉 TVB-N 浓度的预测结果

Fig. 11 Prediction results of TVB-N concentration in mutton by CARS-LS-SVM model

明这两种方法的建模效果取决于建模参数和建模样本。

表 5 不同波长提取方法的 LS-SVM 预测模型比较

Table 5 Comparison of LS-SVM models with different wavelength selection methods

特征波长 提取方法	参数 ( $\gamma$ )	参数 ( $\sigma^2$ )	校正集		验证集	
			$R^2_C$	RMSEC	$R^2_V$	RMSEV
Full spectrum	93.497 6	3.641.32 8	0.914 5	1.690 7	0.844 7	1.964 3
CARS	12.031 3	5.627 6	0.923 5	1.599 0	0.850 1	1.929 4
UVE	26.185 0	1.118.810	0.901 3	1.817 0	0.855 1	1.896 9
IUVE	94.239 7	1.716.1	0.911 2	1.723 4	0.856 8	1.886 2
SPA	131.846 9	20.751 5	0.879 3	2.008 8	0.824 3	2.089 2

### 3 结 论

利用近红外光谱对羊肉 TVB-N 浓度进行预测, 主要结论如下:

(1) 以 MCS 法和 MD 法剔除了羊肉的光谱数据的 5 个异常值, 且 2 种方法的检测结果重合度和可信度较高。

(2) 以原始光谱和 6 种不同方法预处理的光谱建立了 PLS 预测模型, SGS 处理的光谱建模效果最好, 平滑处理、尺度缩放方法的建模效果整体上好于散射校正方法。

(3) 利用 CARS, UVE, IUVE, SPA 提取特征光谱得到的波长个数分别为 14, 703, 144, 15, 占全光谱 1921 个波长的 0.73%, 36.60%, 7.50%, 0.78%。对比全光谱和各方法提取的特征波长所建立的预测模型, CARS 提取的波长建立的模型性能最优。对比 UVE 法, IUVE 可消除波长间的共线性和提高模型性能。

(4) 对提取的特征波长建立了储存期生鲜羊肉 TVB-N 的 PLS, SVM 和 LS-SVM 预测模型, 最好的校正集预测结果由 SVM 模型取得, 最好的验证集预测效果由 LS-SVM 模型得到, 这两种方法的建模效果与模型参数和建模样本密切相关。

### References

- [1] Li Y, Tang X, Shen Z, et al. Food Chemistry, 2019, 287: 126.
- [2] JIANG Pei-hong, ZHANG Yu-hua, CHEN Dong-jie, et al(姜沛宏, 张玉华, 陈东杰, 等). Food Science(食品科学), 2016, 37(6): 161.
- [3] BAI Jing, LI Jia-peng, ZOU Hao, et al(白京, 李家鹏, 邹昊, 等). Food Science(食品科学), 2019, 40(2): 287.
- [4] PENG Yan-kun, YANG Qing-hua, WANG Wen-xiu(彭彦昆, 杨清华, 王文秀). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2018, 49(3): 347.
- [5] WANG Hui, TIAN Han-you, ZHANG Shun-liang, et al(王辉, 田寒友, 张顺亮, 等). Food Science(食品科学), 2016, 37(20): 134.
- [6] Dixit Y, Casado-Gavalda M P, Cama-Moncuñil R, et al. Journal of Near Infrared Spectroscopy, 2016, 24(4): 353.
- [7] PENG Yan-kun, YANG Qing-hua, WANG Wen-xiu(彭彦昆, 杨清华, 王文秀). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2018, 49(03): 347.
- [8] WANG Wen-xiu, PENG Yan-kun, SUN Hong-wei, et al(王文秀, 彭彦昆, 孙宏伟, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(23): 290.

- [9] Li Z, Wang J, Xiong Y, et al. *Vibrational Spectroscopy*, 2016, 84: 24.
- [10] ZHANG Man, LIU Xu-hua, HE Xiong-kui, et al(张曼, 刘旭华, 何雄奎, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2010, 30(5): 1214.
- [11] SUN Tong, JIANG Shui-quan(孙通, 江水泉). *Food & Machinery(食品与机械)*, 2016, 32(3): 69.
- [12] Thomas S, Pillai G N, Pal K. *Geomatics Natural Hazards & Risk*, 2017, 8(2): 177.
- [13] Balabin R M, Lomakina E I. *Analyst*, 2011, 136(8): 1703.

## Prediction Model of TVB-N Concentration in Mutton Based on Near Infrared Characteristic Spectra

ZHANG Xu<sup>1</sup>, BAI Xue-bing<sup>1</sup>, WANG Xue-pei<sup>2</sup>, LI Xin-wu<sup>2</sup>, LI Zhi-gang<sup>3</sup>, ZHANG Xiao-shuan<sup>2, 4\*</sup>

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

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

3. College of Information Science and Technology, Shihezi University, Shihezi 832003, China

4. Beijing Laboratory of Food Quality and Safety, China Agricultural University, Beijing 100083, China

**Abstract** In order to improve the stability and accuracy of near-infrared spectroscopy (NIR) detection of total volatile basic nitrogen (TVB-N) in fresh mutton during storage (at 4 °C, 8 °C, 20 °C), the selection of characteristic spectra and prediction models is the key step of NIR spectroscopy research. The 121 mutton samples were taken as experimental objects, the NIR spectra between 680 and 2 600 nm of fresh mutton samples were collected. The scattering correction methods, including multi scattering correction (MSC), standard normal transformation (SNV), and smoothing methods including Savitzky Golay convolution smoothing (SGS), moving average smoothing (MAS), and scaling methods including normalization, centring and auto scaling, were adopted to pretreat NIR spectra, and then PLS prediction models were built, by comparison, it is found that the spectra treated with SGS got the best modeling effect. Monte Carlo sampling (MCS) method and Mahalanobis distance method (MD) were used to eliminate 5 abnormal data of mutton spectra. The sample-set partitioning based on joint  $x$ - $y$  distance (SPXY) algorithm was used to split 75% (87 samples) of the total samples as calibration set samples and the remaining 29 were validation set samples. The competitive adaptive reweighted sampling (CARS) algorithm, uninformative variable elimination (UVE) algorithm, improved uninformative variable elimination (IUVE) algorithm, successive projections algorithm (SPA) were employed to select characteristic wavelengths, and wavelength numbers were 14, 703, 144 and 15, respectively. The full spectra and the characteristic wavelengths selected by the four methods were taken as input variables to build prediction models, the results show that the performance of the model built with the wavelengths selected by CARS is better than the model built with the wavelengths selected by UVE, IUVE and SPA, and it shows that CARS method can effectively simplify the input variables and improve the performance of the prediction model. Compared with the UVE algorithm, the IUVE algorithm can select fewer wavelengths and improve the model's performance. The PLS models, support vector machine (SVM) models and least squares support vector machine (LS-SVM) models were established with the selected characteristic wavelengths. The optimal prediction results of the calibration set are obtained by SVM models, in which the calibration determination coefficient ( $R_c^2$ ) and root mean square error of calibration (RMSEC) of the CARS-SVM prediction model were 0.939 1 and 1.426 7, respectively. LS-SVM prediction model achieves the optimal prediction results of validation set, and the validation determination coefficient ( $R_v^2$ ) and the root mean square error of validation (RMSEV) of IUVE-LS-SVM prediction model were 0.856 8 and 1.886 2, respectively. The simplified and optimized TVB-N prediction models for fresh mutton during the storage period are established based on NIR characteristic spectra, which provides reference and technical support for rapid and non-destructive detection of TVB-N concentration in fresh mutton.