

MIV 波长优选改善 VIS/NIR 光谱 TVB-N 模型性能研究

陈亦凡, 李芸婧, 彭苗苗, 杨春勇*, 侯金, 陈少平

智能无线通信湖北省重点实验室, 中南民族大学电子信息工程学院, 湖北 武汉 430074

摘要 挥发性盐基氮(TVB-N)是衡量肉品新鲜的重要理化指标, 利用可见/近红外(VIS/NIR)光谱对 TVB-N 含量进行定量检测具有重要意义。预测模型是 VIS/NIR 光谱检测 TVB-N 含量性能的关键要素, 使其兼顾准确性与稳健性可有效改善 TVB-N 的定量分析结果。以猪肉为例, 采集 51 组不同新鲜度样本的 VIS/NIR 光谱数据, 去除低信噪比区间 200~450 和 900~1 000 nm, 选取有效波段 450~900 nm 的光谱数据用于建模。随后利用主成分分析(PCA)对光谱信息降维, 构建一个反向传播神经网络(BPNN)模型。在此基础上, 提出用平均影响值(MIV)方法从有效波段中优选与肉质 TVB-N 含量强相关的特征波长, 最终基于 221 个优选波长, 构建一个 MIV-PCA-BPNN 预测模型。实验表明, 初步构建的 PCA-BPNN 非线性预测模型, 校正相关系数(R_c)和校正均方根误差(RMSEC)分别为 0.96 和 1.47 mg/100 g, 预测相关系数(R_p)和预测均方根误差(RMSEP)分别为 0.93 和 1.74 mg/100 g, 模型稳健性指标为 1.18, 优于经典的线性预测模型主成分分析回归和偏最小二乘回归, 证明 TVB-N 具有较强的非线性效应。最终构建的 MIV-PCA-BPNN 预测模型的 R_c 和 RMSEC 分别为 0.98 和 1.21 mg/100 g, R_p 和 RMSEP 分别为 0.96 和 1.12 mg/100 g, 模型稳健性指标为 1.08, 在所构建的预测模型中, RMSEC 和 RMSEP 最小, R_c 和 R_p 最大, 模型的准确性和稳健性最佳。另外, MIV 方法筛选出的特征波长集中在 7 个波峰附近, 皆分布于肉品中化学成分的吸收区内, 且与 TVB-N 中的含氢基团的特征吸收峰表现出高度一致性, 为利用 MIV 方法筛选波长变量提供了理论依据。研究结果显示, MIV 波长优选可有效改善预测模型的性能, 为利用神经网络剔除无关波长变量提供了新思路, 所构建的 MIV-PCA-BPNN 预测模型满足了肉质中 TVB-N 定量分析的需求。

关键词 VIS/NIR 光谱检测; 反向传播神经网络; 波长优选; 挥发性盐基氮

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)05-1413-07

引言

根据我国鲜冻畜肉的国家标准(GB/T 5009.44—2003), 挥发性盐基氮(TVB-N)是衡量肉品新鲜度的一项重要理化指标。传统标定新鲜度的方法费时、繁琐、破坏性大, 难以满足快速检测 TVB-N 含量的需求。近年来, 可见/近红外(VIS/NIR)光谱检测凭借快捷、简便和无损等优势, 成为检测领域的热点^[1-3], 也已应用在 TVB-N 含量的检测。

在光谱分析检测领域, 预测模型属于研究重点^[4]。关于预测模型, 目前大致分为线性预测模型和非线性预测模型两类。为了发展肉质中 TVB-N 含量快速检测法, 针对这两类模型, 国内外学者就光谱预测模型进行了深入研究。意大利

的 Alamprese 等^[5]利用偏最小二乘回归(partial least squares regression, PLSR)引入 12 个主成分, 建立了 942~2 667 nm 的 NIR 全谱线性预测模型, 模型预测相关系数(R_p)为 0.95, 预测均方根误差(root mean square error of prediction, RMSEP)为 6.47 mg/100 g, 初步满足了对牛肉 TVB-N 定量分析的需求。国内的蔡健荣等学者^[6]对 1 000~2 500 nm 的猪肉 NIR 光谱数据, 建立了基于波长筛选的联合偏最小二乘回归(siPLSR)预测模型, R_p 为 0.82, RMSEP 为 4.17 mg/100 g, 进一步提高了线性预测模型 PLSR 的准确性。此后, Huang 等^[7]使用主成分分析(principal component analysis, PCA)降维相同波段的猪肉光谱数据作为反向神经网络(back propagation neural network, BPNN)的输入, 建立了 NIR 全谱非线性预测模型, R_p 为 0.94, RMSEP 为 3.63 mg/100 g。

收稿日期: 2019-04-17, 修订日期: 2019-08-28

基金项目: 湖北省自然科学基金科技支撑计划项目(2015BCE048), 国家科技支撑计划课题(2015BAD29B01), 湖北省自然科学基金重点项目(2014CFA051), 中央高校基本科研业务费专项资金本科培育项目(CZP17026)资助

作者简介: 陈亦凡, 1999 年生, 中南民族大学电子信息工程学院本科生 e-mail: anguoc@gmail.com

* 通讯联系人 e-mail: cyyang@mail.scuec.edu.cn

相对于线性预测模型, NIR 全谱非线性预测模型准确性更高。对比相关典型文献, 大多数全谱线性预测模型可用于肉质 TVB-N 含量的定量分析, 尽管其结合波长筛选一定程度上提升了模型性能, 但总体性能与全谱非线性预测模型 BPNN 相比仍有差距。然而, 全谱非线性预测模型 BPNN 的全谱信息中含有大量无关信息变量, 严重影响模型性能, 建模过程使用波长筛选仍然必不可少^[8-9]。另据可查阅文献, 波长筛选改良 BPNN 预测模型在 VIS/NIR 光谱检测领域鲜有研究。虽然, Dombi^[10]等在 1995 年提出平均影响值(mean impact value, MIV)方法对剔除无关变量有显著优势, 但是这一方法主要用于质谱分析、图像处理和生物医学, 鲜见涉及光谱分析领域的应用^[11-13]。

为了构建基于波长筛选的猪肉 TVB-N 含量的 BPNN 预测模型, 首先利用 PCA 对猪肉光谱数据进行降维, 构建 TVB-N 含量的 PCA-BPNN 预测模型; 再进一步引入 MIV 方法对 VIS/NIR 全谱波长变量进行强相关特征优选, 获得改良预测模型 MIV-PCA-BPNN。最后, 从准确性和稳健性两个方面评估 MIV-PCA-BPNN 预测模型。

1 实验部分

1.1 猪肉样品的制备

猪肉样本购于周边农贸市场, 取自不同猪个体的背最长肌部位。参考文献[5]的做法, 将样本切割成 30 mm×30 mm×20 mm 尺寸块, 进行编号排序, 共得 51 个样本。将样本用密封袋封装冷藏于 4 °C 环境。测量时, 每间隔 24 h 按编号大小依次取出 5~7 个样品, 静置于 25 °C 室温环境约 20 min, 再分别对其进行光谱数据采集和 TVB-N 含量测定。

1.2 光谱数据的采集

光谱采集系统包括海洋光学的光谱仪 QE65pro, 卤钨光源 HL-2000, 载物台和计算机等部件; 光谱仪采集范围为 200~1 000 nm, 光学分辨率为 0.14~7.7 nm, 光源波长范围为 360~2 400 nm; 通过计算机和光谱仪配套采集软件 SpectraSuit, 以反射模式实施检测, 光谱采集系统如图 1 所示。在检测过程中, 保证光源探头垂直向下并距离样品 1.5 cm, 在视场覆盖样品时应避开其边缘和反光区。

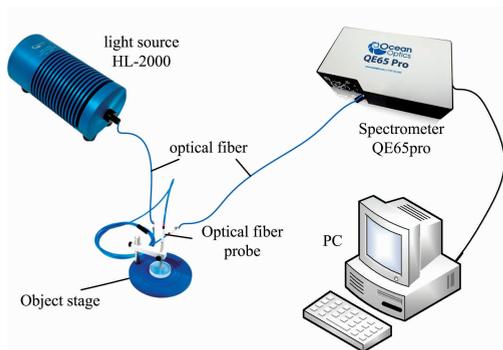


图 1 近红外光谱检测系统示意图

Fig. 1 Diagram of spectral acquisition system

1 000 nm 区间的光谱信噪比较低, 严重干扰模型的建立。因此, 选取 450~900 nm 的数据用于后续建模。51 个样品的反射光谱如图 2 所示。

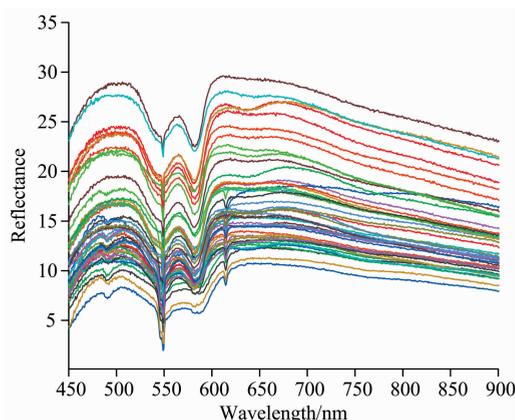


图 2 51 个猪肉样本的原始反射光谱

Fig. 2 Raw reflectance spectra of 51 pork samples

1.3 TVB-N 含量的测定

国家标准(GB/T 5009.44—2003)规定对肉质 TVB-N 含量的检测采用半微量定氮法, 据此方法测定结果如图 3 所示。从图中可看出, TVB-N 含量总体上随着时间推移而不断增加, 从第 5 日起 TVB-N 含量加速增加。该结果表明, TVB-N 含量是肉品腐败程度的反映, 存放第 5 日后, 肉品表面的微生物污染逐步向深层蔓延, 蛋白质分解加剧, 导致肉品理化性质改变。

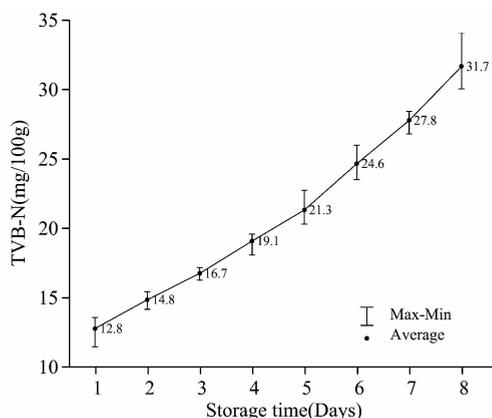


图 3 TVB-N 含量随天数变化情况(4 °C 冷藏环境)

Fig. 3 The TVB-N content of pork varies with days (4 °C in a refrigerated environment)

为了建立 TVB-N 含量的预测模型, 从 51 个样本中随机选取 10 个样本作为验证集, 其余的 41 个样本作为校正集, TVB-N 值统计结果如表 1 所示。

1.4 全谱 PCA-BPNN 预测模型的构建

构建的 PCA-BPNN 工作流程如图 4 所示。图中的 $X_1, X_2, X_3, \dots, X_n$ 是经 PCA 降维的光谱数据, 也是 BPNN 模型的输入节点; ω_{ij} 和 ω_{jk} 代表光谱数据在神经元之间传递的权重, 体现 BPNN 的网络权重; Y_1 表示 TVB-N 预测值, 是

据文献[14], 猪肉中 TVB-N 在波段 200~450 和 900~

BPNN 的输出节点。

表 1 校正集和验证集 TVB-N 值(mg/100 g)统计结果

Table 1 Statistics of pork TVB-N value (mg/100 g) in calibration and prediction sets

样本集	样本个数	平均值	最大值	最小值	标准差
校正集	41	20.14	34.10	11.45	5.68
验证集	10	18.50	28.42	13.56	4.29

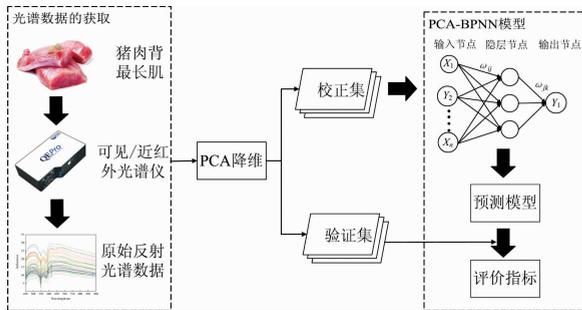


图 4 PCA-BPNN 流程图

Fig. 4 Flow chart of PCA-BPNN

每一个样本所携带的光谱信息含有近 600 个波长变量,如直接作为 BPNN 的输入节点,必将延长其收敛时间。为此,先用 PCA 对标准化后的样本光谱数据进行降维处理。选定占原始光谱信息量 99.99% 的 11 个主成分作为 BPNN 的输入节点。通过以校正相关系数(R_c)、预测相关系数、校正均方根误差(root mean square error of calibration, RMSEC)和预测均方根误差为评价指标的训练与优化,确定 BPNN 参数如表 2 所示。

表 2 BPNN 参数设定

Table 2 BPNN parameters setting

参数名称	参数设置
隐含层传递函数	Sigmoid
输出层传递函数	Purelin
训练函数	Levenberg-Marquardt
隐含层数	2
隐含层节点数	[11, 5]
学习速率	0.1
迭代次数	1 000
目标误差	0.001

1.5 MIV 波长筛选

MIV 在神经网络应用中是评价变量对结果影响较为有效的指标之一。根据 MIV 数值的相对大小,可以精确地定位到与 TVB-N 含量强相关的波长变量。MIV 波长筛选原理与流程如图 5 所示,操作步骤为:

- (1) 训练出 R_c 和 R_p 均大于 0.9 的 PCA-BPNN 模型。
- (2) 先将训练集 T 的第 i 个波长的反射强度数值分别增减 10%, 得到新数据集 X_i 和 Y_i , 随后将 X_i 和 Y_i 经 PCA 降维后作为仿真样本。

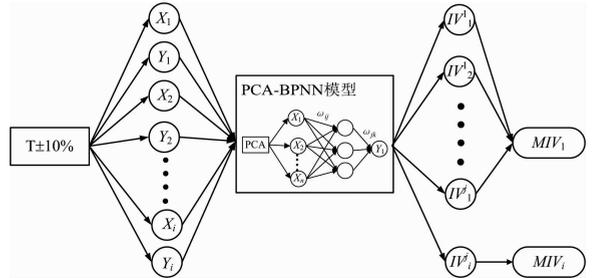


图 5 MIV 方法波长筛选流程图

Fig. 5 Flow chart of wavelength selection in MIV

(3) 利用已训练的 PCA-BPNN 模型进行预测, 得到两组预测结果。将预测结果代入式(1), 得到样本变动后对输出产生的影响变化值 IV_i^j (impact value, IV), 随后对各样本所得 IV 的绝对值求均值, 即为各波长的平均影响值 MIV_i , 如式(2)。

$$IV_i^j = \text{net}(X_i^j) - \text{net}(Y_i^j) \quad (1)$$

$$MIV_i = \frac{1}{n} \sum_{j=1}^n |IV_i^j| \quad (2)$$

式(1)中 $\text{net}()$ 表示 BPNN 的非线性变换, X_i^j 和 Y_i^j 分别为第 j 组样本在第 i 个波长的反射强度数值增加和减少 10% 的结果, 所得 IV_i^j 的符号代表作用方向; 式(2)中 n 为样本组数, 所得 MIV_i 表示第 i 个波长对网络影响权值, 随后按照权值大小进行波长筛选。

1.6 MIV-PCA-BPNN 预测模型的构建

利用 MIV 方法获得优选波长, 最终建立 MIV-PCA-BPNN 预测模型, 以此提高预测模型的准确性和稳健性。建模流程如图 6 所示。

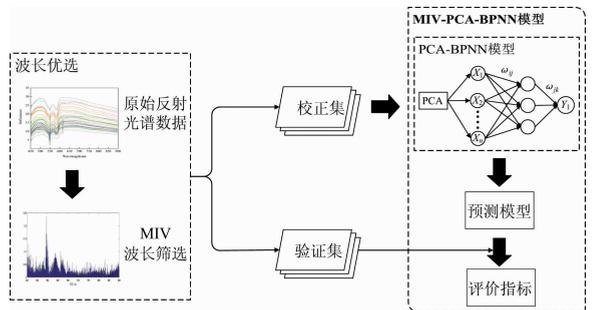


图 6 MIV-PCA-BPNN 流程图

Fig. 6 Flow chart of MIV-PCA-BPNN

MIV-PCA-BPNN 工作流程包括波长优选与建模两个步骤。波长优选把原始光谱数据中的无关波长变量剔除, 将优选的特征波长作为 PCA-BPNN 模型的输入以构建 MIV-PCA-BPNN 预测模型。建模完成后, 用验证集对模型准确性和稳健性进行验证, 得到其评价指标。在上述流程中, 利用 MIV 去除光谱数据中的冗余信息, 提高模型准确性和稳健性, 而 PCA 则用于简化网络规模, 缩短模型收敛时间。

1.7 模型性能评价

预测模型以 R_c , R_p , RMSEC 和 RMSEP 作为评价性能的指标。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i, \text{actual}} - y_{i, \text{predicted}})^2} \quad (3)$$

$$R = \frac{\sum_{i=1}^n (y_{i, \text{predicted}} - \bar{y}_{\text{predicted}})(y_{i, \text{actual}} - \bar{y}_{\text{actual}})}{\sqrt{\sum_{i=1}^n (y_{i, \text{predicted}} - \bar{y}_{\text{predicted}})^2} \sqrt{\sum_{i=1}^n (y_{i, \text{actual}} - \bar{y}_{\text{actual}})^2}} \quad (4)$$

式(3)与式(4)中 n 同式(2)的定义, 为样本组数, $y_{i, \text{actual}}$ 表示第 i 组样品的 TVB-N 测定值, $y_{i, \text{predicted}}$ 表示所建模型的第 i 组样品 TVB-N 预测值。

评价预测模型性能时, 先考虑其准确性, 随后再评估其稳健性。通常, R_c 数值越接近 1, RMSEC 的数值越低, 预测模型的拟合精度越高; R_p 数值越接近 1, RMSEP 的数值越低, 预测值与测定值的误差越小, 预测模型的预测精度越

高。模型的准确性主要取决于拟合精度和预测精度。衡量预测模型稳健性的指标是 RMSEP/RMSEC, 其数值越低, 模型的稳健性越好。通常, 一个具有优良稳健性的模型 RMSEP/RMSEC ≤ 1.2 ^[15]。

2 结果与讨论

对 450~900 nm 的光谱数据进行 MIV 波长筛选, 波长分布结果如图 7 所示。从图中可以看出, 在全光谱波段不同波长变量的 MIV 数值有较大差异, 为了减少无关信息的影响, 将 MIV 数值小于平均值 0.034 且与 TVB-N 含量相关性较弱的 373 个波长变量全部剔除, 从而得到优选的 221 个波长变量。

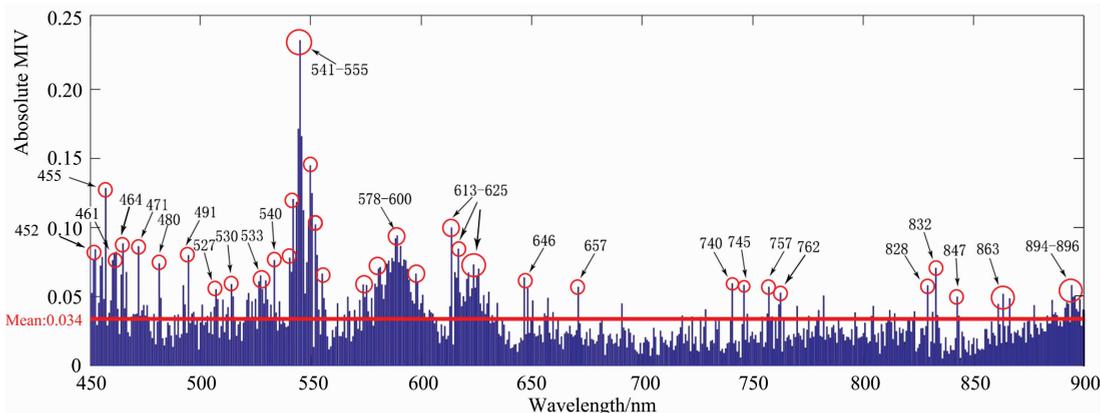


图 7 MIV 筛选出特征波长的分布

Fig. 7 Distribution of characteristic wavelength selected by MIV

由图 7 可见, MIV 数值较大的波段集中分布在 7 个波峰附近。在 VIS 波段(450~780 nm), 第一个波峰出现在 471 nm, 即以波长 471 nm 为中心, 左右两侧 20 nm 内有 452, 455, 461, 480 和 491 nm 特征波长, 这些波长为肉品中高铁肌红蛋白的吸收峰区。第二个波峰出现在 544 nm, 即以波长 544 nm 为中心, 左右两侧各 20 nm 内有 527, 530, 533, 540 和 541~555 nm 特征波长, 这些波段为肉品中氧合肌红蛋白的吸收峰区。第三和第四个波峰出现在 588 和 613 nm, 这两个波峰两侧各有 578~600 和 613~625 nm 特征波段, 分别为高铁肌红蛋白和硫肌红蛋白的吸收区。第五个波峰出现在 740 nm, 以波长 740 nm 为中心, 左右 20 nm 内有 745, 757 和 762 nm 特征波长, 这些波长为肉品 O—H 键的第三倍频的吸收峰区^[16]。在 NIR 波段(780~950 nm), 第六个波峰出现在 832 nm, 以 832 nm 为中心波长, 左右 20 nm 内有 828 和 847 nm 特征波长, 这三处特征波长为氨中 N—H 键第三倍频的吸收峰。最后一个波峰出现在 894 nm, 其两侧的特征波长 863, 895 和 896 nm 为肉品 C—H 键第三倍频吸收峰^[17]。

TVB-N 是猪肉存储过程中由于蛋白质分解而产生氨以及胺类等碱性含氮物质^[14, 19], 这些碱性含氮物质中的含氮基团(C—H, O—H, S—H, N—H 等化学键)的特征吸收峰

与 MIV 方法所筛选出来的特征波长表现出高度一致性, 由此为利用 MIV 方法筛选光谱的波长变量提供了理论依据。

基于所构建的 MIV-PCA-BPNN 预测模型获得的样本预测值与测定值之间的散点图如图 8(a) 所示。与图 8(b) 中的基于全谱的 TVB-N 预测模型 PCA-BPNN 的结果相比较, 图 8(a) 中校正集和验证集的数据点明显汇聚于斜线附近, 且误差大的极端数据点较少。可见波长优选获得了更好的模型拟合和预测效果, 在准确性和稳健性方面有较大提升。

MIV-PCA-BPNN 模型的变量数较少, 但主成分数较多。究其原因, MIV 波长筛选剔除了大部分非线性和不相关的波长变量, 使得变量之间相关性得到了增强, PCA 降维光谱数据时, 所提取的主成分包含了大量的与 TVB-N 含量相关的信息, 因此主成分数增加, 结果与文献[9]结论一致。

为了进一步比较非线性预测模型与线性预测模型的性能, 参考文献[18]的做法, 组合多种常见的算法对样本光谱信息进行预处理, 其中包括: 标准正态变量变换(SNV)、多元散射校正(MSC)、Savitzky-Golay(S-G)平滑, 窗口参数为 3~21(间隔为 2), 拟合次数为 1~5, 求导阶数为一阶或二阶(1D, 2D)、均值中心化、标准化等。通过反复调整上述算法的参数与输入的主成分数, 构建了优化的 PLSR 和 PCR 模型, 结果如图 9 所示。

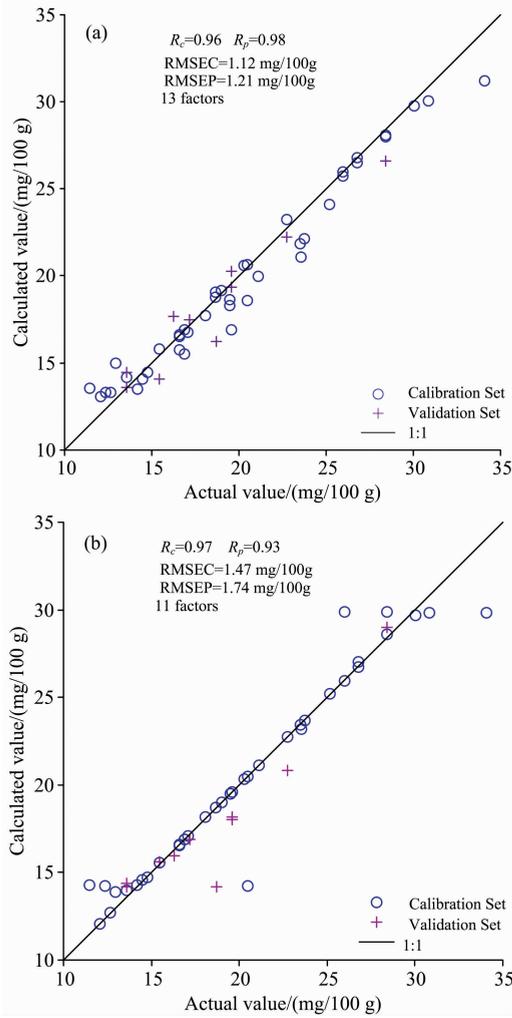


图 8 MIV-PCA-BPNN (a) 和 PCA-BPNN 模型的 (b) 预测值和测定值散点图

Fig. 8 Scatter plots of predicted and actual values (a): MIV-PCA-BPNN; (b): PCA-BPNN

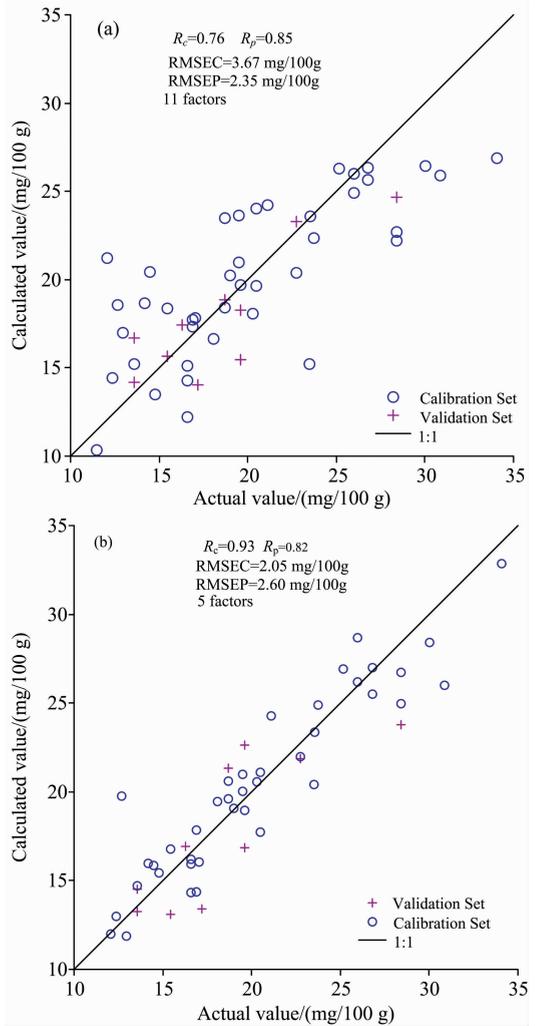


图 9 PCR (a) 和 PLSR (b) 模型的预测值和测定值散点图

Fig. 9 Scatter plots of predicted and actual values (a): PCR; (b): PLSR

表 3 对应不同预测模型的 TVB-N 含量分析结果

Table 3 Analysis results of TVB-N content for different prediction models

建模方式	预处理方法	波长 点数	主成 分数	R_c	RMSEC/ (mg/100 g)	R_p	RMSEP/ (mg/100 g)	RMSEP/RMSEC
PCR	MSC+S-G(11, 2)+2D	594	10	0.76	3.67	0.85	2.35	0.64
PLSR	SNV+S-G(9, 2)+1D	594	5	0.93	2.05	0.82	2.60	1.27
PCA-BPNN	标准化	594	11	0.96	1.47	0.93	1.74	1.18
MIV-PCA-BPNN	标准化	221	13	0.98	1.12	0.96	1.21	1.08

如表 3 所示, PCR 预测模型使用了 10 个主成分, 与表中的其他建模方式相比其 RMSEP/RMSEC 值最低, 模型稳健性最强, 但由图 9(a) 可以看出, 其样本预测值和测定值数据点分布散乱, 没有出现明显汇聚。PCR 预测模型的性能指标 $R_c=0.76$, $RMSEC=3.67$ mg/100 g, $RMSEP-RMSEC < 0$, 对校正集拟合误差较大, 出现严重的欠拟合, 模型准确性表现最差, 因此不能满足对 TVB-N 的定量分析。同为线

性预测模型的 PLSR 模型性能相较于 PCR 有较大提升。虽然 PLSR 预测模型使用较少主成分降低了模型的拟合误差, 拟合精度相较于 PCR 模型也有了较大提升, 但其预测精度低于前者, 且 $RMSEP/RMSEC=1.27$, 略大于要求的 1.20, 因此模型稳健性不足。

从表 3 还可以看出, 在构建的 PLSR, PCR 和 BPNN 三种基于全谱的预测模型中, PCA-BPNN 的 RMSEC 和 RM-

SEP 最小, R_c 和 R_p 最大, 模型准确性最高。因此, 在本研究显示, 非线性预测模型比线性预测模型具备更好准确性和稳健性。究其原因, 光谱信息与待测组分 TVB-N 含量之间不仅存在线性关系, 而且存在非线性关系, 本工作所检测的 TVB-N 正好表现出较强的非线性效应。

另由表 3 可知, 使用 MIV 波长优选的 221 个波长变量所建立的 MIV-PCA-BPNN 预测模型, 使用了 13 个主成分, 与基于全谱的 PCA-BPNN 预测模型相比, R_c 从 0.96 变为 0.98, R_p 从 0.93 上升到 0.96, 并且其 RMSEC 和 RMSEP 为最小, 分别为 1.12 mg/100 g 和 1.21 mg/100 g, 在所有构建的预测模型中, 准确性和稳健性最佳。

3 结 论

以猪肉背最长肌为研究对象, 依据选定的 450~900 nm

VIS/NIR 光谱数据, 构建了 PLSR, PCR 和 BPNN 三种 TVB-N 含量预测模型, 开展了 TVB-N 含量的定量分析研究。结果表明, 非线性预测模型 BPNN 优于线性模型 PCR 和 PLSR, 其中改进的 MIV-PCA-BPNN 预测模型性能最佳。MIV-PCA-BPNN 性能指标 R_c 和 RMSEC 分别为 0.98 和 1.21 mg/100 g, R_p 和 RMSEP 分别为 0.96 和 1.12 mg/100 g, RMSEP/RMSEC=1.08, 相比全谱 BPNN 预测模型, 其准确性和稳健性有较大提升。经 MIV 数值分析, 获得了与 TVB-N 吸收峰一致的优选特征波长, 表明 MIV 方法可有效筛选光谱波长变量。本工作改进了 TVB-N 预测模型, 为利用神经网络剔除无关波长变量提供了新思路。

References

- [1] Qiao L, Tang X, Dong J. Food Chemistry, 2017, 237: 1179.
- [2] Qu F, Ren D, He Y, et al. Meat Science, 2018, 146: 59.
- [3] Fan Y, Liao Y, Cheng F. International Journal of Food Properties, 2018, 21(1): 1180.
- [4] SONG Xiang-zhong, TANG Guo, ZHANG Lu-da, et al(宋相中, 唐果, 张录达, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(4): 1048.
- [5] Alamprese C, Casale M, Sinelli N, et al. LWT—Food Science and Technology, 2013, 53(1): 225.
- [6] CAI Jian-rong, WAN Xin-min, CHEN Quan-sheng(蔡健荣, 万新民, 陈全胜). Acta Optica Sinica(光学学报), 2009, (10): 2808.
- [7] Huang L, Zhao J, Chen Q, et al. Food Chemistry, 2014, 145: 228.
- [8] Li H D, Liang Y Z, Long X X, et al. Chemometrics & Intelligent Laboratory Systems, 2013, 122(5): 23.
- [9] YANG Mei-hua, ZHAO Xiao-min, FANG Qian, et al(杨梅花, 赵小敏, 方倩, 等). Scientia Agricultura Sinica(中国农业科学), 2014, (12): 2374.
- [10] Dombi G W, Nandi P, Saxe J M, et al. Journal of Trauma and Acute Care Surgery, 1995, 39(5): 915.
- [11] Tan X, Ji Z, Zhang Y. Technology & Health Care, 2018, 26(6): 1.
- [12] Zhang Z, Yang J. International Journal of Oil, Gas and Coal Technology, 2016, 11(3): 279.
- [13] Fan H, Xuan J, Du X, et al. RSC Advances, 2018, 8(69): 39602.
- [14] WEI Wen-song, PENG Yan-kun, ZHENG Xiao-chun, et al(魏文松, 彭彦昆, 郑晓春, 等). Acta Optica Sinica(光学学报), 2017, 37(11): 382.
- [15] YAN Yan-lu, CHEN Bin, ZHU Da-zhou, et al(严衍禄, 陈斌, 朱大洲, 等). Near Infrared Spectroscopy—Principles, Technologies and Applications(近红外光谱分析的原理、技术与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2013. 123.
- [16] Wu J, Peng Y, Li Y, et al. Journal of Food Engineering, 2012, 109(2): 267.
- [17] Cheng J, Sun D, Qu J, et al. Journal of Food Engineering, 2016, 182: 9.
- [18] ZOU Hao, TIAN Han-you, LIU Fei, et al(邹昊, 田寒友, 刘飞, 等). Food Science(食品科学), 2016, 37(22): 180.

Improvements of VIS-NIR Spectroscopy Model in the Prediction of TVB-N Using MIV Wavelength Selection

CHEN Yi-fan, LI Yun-jing, PENG Miao-miao, YANG Chun-yong*, HOU Jin, CHEN Shao-ping

Hubei Key Laboratory of Intelligent Wireless Communications, College of Electronics and Information Engineering, South-Central University for Nationalities, Wuhan 430074, China

Abstract Volatile Basic Nitrogen (TVB-N) is an important physicochemical property for the detection of meat freshness. Using visible/near-infrared (VIS/NIR) spectroscopy to analyze TVB-N content is of great importance quantitatively. The prediction model is the key factor for detection TVB-N content in visible or near infrared spectroscopy. Thus, an accurate and robust prediction model can improve the quantitative analysis results of TVB-N. Firstly, we collected 51 representative pork samples with different freshness, and determine the effective band from 450 to 900 nm after removing low signal-to-noise ratio band from 200 to 450 nm and from 900 to 1 000 nm. Then we use principal component analysis (PCA) to reduce spectral data in order to construct a back propagation neural network (BPNN) model. On this basis, we use the mean impact value (MIV) method to select characteristic wavelengths which strongly related to the content of Total Volatile Basic Nitrogen (TVB-N) in edible meat, and finally construct a MIV-PCA-BPNN prediction model based on 221 selected wavelengths. Experimental results show that the related coefficient of calibration (R_C), the related coefficient of prediction (R_P), the root means square error of calibration (RMSEC), the root mean square error of prediction (RMSEP) and the robustness index of the PCA-BPNN model are 0.96, 0.93, 1.47 mg/100 g, 1.74 mg/100 g and 1.18, respectively. The PCA-BPNN nonlinear prediction model is better than the classical linear prediction model principal component regression and partial least squares regression prediction model, which proves that TVB-N has strong nonlinear effects. The R_C , R_P , RMSEC, RMSEP and the robustness index of the MIV-PCA-BPNN model are 0.98, 0.96, 1.12 mg/100 g, 1.21 mg/100 g and 1.08, respectively, it is RMSEC and RMSEP are the smallest, while R_C , R_P are the largest. Therefore, MIV-PCA-BPNN is the most accurate and robust model in all constructed prediction model. In addition, the characteristic wavelengths selected by the MIV method are concentrated near 7 peaks, which are distributed in the absorption regions of chemical composition in meat. The characteristic wavelengths are consistent with the absorption peaks of H Contained Groups in TVB-N, which provides a theoretical basis for selecting wavelengths by the MIV method. It is found that the MIV wavelength selection is effective to improve the performance of the prediction model, which offers new thought for using the neural network to eliminate irrelevant wavelength variables. The MIV-PCA-BPNN prediction model could be used for the quantitative analysis of TVB-N in meat.

Keywords Visible/near-infrared spectroscopy; Back propagation neural network (BPNN); Wavelength selection; Total volatile basic nitrogen (TVB-N)

(Received Apr. 17, 2019; accepted Aug. 28, 2019)

* Corresponding author