

鸡蛋新鲜度、pH值及黏度的高光谱检测模型

付丹丹¹, 王巧华^{1,2,*}

(1. 华中农业大学工学院, 湖北 武汉 430070; 2. 国家蛋品加工技术研发分中心, 湖北 武汉 430070)

摘要: 利用高光谱成像仪采集鸡蛋的高光谱透射图像, 并利用游标卡尺、pH计、黏度计测定鸡蛋的新鲜度、酸碱度与黏度, 用竞争性自适应重加权 (competitive adaptive reweighed sampling, CARS) 算法与连续投影算法 (successive projections algorithm, SPA) 选取特征波长分别建立鸡蛋品质与其高光谱特征的简单多元线性回归 (multiple linear regression, MLR) 模型, 并在CARS提取的特征波长基础上用SPA进行了二次波段提取, 建立相应的MLR模型, 对比一次波长提取与二次波长提取选择的特征波长建立的模型性能。结果表明, CARS与SPA建立的MLR模型的验证集相关系数均在0.9以上。二次特征波段建立MLR模型的验证集相关系数比一次特征波段提取建立的MLR相关系数高, 且均方根误差 (root mean square error, RMSE) 均有所减小, 选取的特征波段比单独使用CARS或SPA选取的波段要少。建立的鸡蛋新鲜度、酸碱度 (pH值)、黏度MLR模型的相关系数 R 分别为0.94、0.95、0.95, RMSE分别为6.36、0.17、149。即CARS、SPA提取特征波长可以优化模型, 二次特征波段提取能更进一步的优化模型, 提高了模型的稳定性, 该模型能完好无损预测鸡蛋品质。

关键词: 鸡蛋; 新鲜度; 酸碱度; 黏度; 高光谱; 连续投影算法

Predictive Models for the Detection of Egg Freshness, Acidity and Viscosity Using Hyper-Spectral Imaging

FU Dandan¹, WANG Qiaohua^{1,2,*}

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;

2. National Research and Development Center for Egg Processing, Wuhan 430070, China)

Abstract: In this study, hyper-spectral data of eggs were collected using a hyper-spectral imager, and their freshness, pH and viscosity were also measured with a vernier caliper, a pH meter and a viscometer, respectively. Characteristic wavelengths were selected using competitive adaptive reweighed sampling (CARS) and successive projections algorithm (SPA), respectively, for constructing a simple multiple linear regression model (MLR) between egg quality and hyper-spectral features. Moreover, another MLR model was also established based on secondary SPA extraction of the characteristic wavelengths selected by CARS. The performances of the three models developed were compared. The results showed that the validation set correlation coefficients of both the CARS model and SPA model were above 0.9, which were lower than that of the CARS-SPA model with a lower root mean square error (RMSE) and fewer characteristic wavelength bands. The correlation coefficients of the MLR model for egg freshness, pH and viscosity were 0.94, 0.95, and 0.95 respectively, and the RMSE were 6.36, 0.17, and 149, respectively. This study indicates that optimized MLP model could be obtained using CARS or SPA to extract characteristic wavelengths and be further optimized by their combined use for non-destructive prediction of egg quality with improved stability.

Key words: eggs; freshness; acidity; viscosity; hyper-spectral imaging; successive projection algorithm

DOI:10.7506/spkx1002-6630-201622026

中图分类号: S24

文献标志码: A

文章编号: 1002-6630 (2016) 22-0173-07

引文格式:

付丹丹, 王巧华. 鸡蛋新鲜度、pH值及黏度的高光谱检测模型[J]. 食品科学, 2016, 37(22): 173-179. DOI:10.7506/spkx1002-6630-201622026. <http://www.spkx.net.cn>

FU Dandan, WANG Qiaohua. Predictive models for the detection of egg freshness, acidity and viscosity using hyper-spectral imaging[J]. Food Science, 2016, 37(22): 173-179. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-201622026. <http://www.spkx.net.cn>

收稿日期: 2016-04-11

基金项目: 国家自然科学基金面上项目 (31371771); “十二五”国家科技支撑计划项目 (2015BAD19B05); 湖北省科技支撑计划项目 (2015BBA172); 公益性行业 (农业) 科研专项 (201303084)

作者简介: 付丹丹 (1991—), 女, 硕士研究生, 研究方向为智能化检测与控制。E-mail: fudandan@webmail.hzau.edu.cn

*通信作者: 王巧华 (1970—), 女, 教授, 博士, 研究方向为机电一体化、智能化检测与控制、机器视觉。

E-mail: wqh@mail.hzau.edu.cn

有关贮藏期间鸡蛋品质的评价与预测及其与贮藏时间和条件的关系一直是食品加工和保鲜领域的研究热点问题之一。而对于鸡蛋品质的评价有多个指标,例如鸡蛋哈夫单位(Haugh unit, HU)值、挥发性盐基氮值、酸碱度、黏度等,这些指标在鸡蛋贮藏期间都会有较显著的变化。目前这些指标的检测方法主要还是以生化方法为主,但是生化方法是破坏性的检测,且存在检测周期长、程序复杂等弊端,因此,要满足快速、无损的现代检测要求,就必须运用现代无损检测技术实现鸡蛋内部品质指标的无损检测。常见的鸡蛋内部品质无损检测的方法有机器视觉法^[1-2]、近红外检测法^[3-5]、高光谱检测法^[6-7]等。

目前已有利用高光谱技术检测普通鸡蛋的新鲜度、种蛋、血斑等^[8-11]的报道,均为常规指标检测,没有深入综合地对鸡蛋内部多个品质指标进行研究。有研究^[12-16]已经发现鸡蛋的可见-近红外光谱与鸡蛋HU值、蛋白含量、pH值、贮藏时间有很好的相关性。对于高光谱联合检测鸡蛋内部理化指标如酸碱度、黏度等鲜见报道。鸡蛋在贮藏期间有多个品质变化,各品质指标之间存在一定的相关性,本实验利用高光谱结合鸡蛋新鲜度对鸡蛋内部的pH值与黏度进行研究,找出能够表征鸡蛋品质的最优波长组合,建立鸡蛋内部多个理化品质指标与高光谱的关系模型。

1 材料与方法

1.1 材料

1日龄的褐壳鸡蛋200枚 湖北武汉九峰养鸡场。将所有鸡蛋放置于(22±1)℃、相对湿度65%恒温培养箱中贮藏。

1.2 仪器与设备

SF2000电子数显卡尺 桂林广陆数字测控股份有限公司; JA2002电子天平 上海浦春计量仪器有限公司; MP511 pH计 上海三信仪表厂; NDJ-5S数显黏度计 上海衡平仪器仪表厂; Zolix HyperSIS-VNIR-CL高光谱成像系统 美国海洋光学公司。

1.3 方法

1.3.1 高光谱相机校正

高光谱成像系统主要是由高光谱成像仪、计算机、透射光源箱、高光谱采集软件(SpectraSENS)等组成,其中高光谱成像仪是由光源、开关、CCD相机和丝杆式移动平台等部分组成。设置光谱波长范围为300~1100 nm,相机曝光时间为0.15 s,丝杆移动平台的速率为1.7 mm/s,移动距离为80 mm。

CCD相机在拍摄鸡蛋的透射图像时,各波段光源的强度形成了不均匀的分布,获取的图像存在一定的噪声,故必须对高光谱图像进行黑白校正^[10]。在采集鸡蛋的

透射图像之前,在光源正上方放置标准的白色校正板,得到全白标定图像W,然后盖住镜头,得到全黑标定图像D,扫描鸡蛋采集的原始高光谱透射图像I,按公式

(1)计算校正后的图像R:

$$R = \frac{I - D}{W - D} \quad (1)$$

1.3.2 鸡蛋新鲜度、pH值及黏度的测定

实验步骤:1)贮藏第0、10、20、30、40天时,随机挑选出40枚鸡蛋,分别进行测定。2)将每次随机挑选出来的鸡蛋大头朝上,依次竖直放在自制的高光谱仪中的透射光源箱的孔口,在保证透射光源箱无漏光条件下将丝杆移动平台移动到合适的位置,通过相机及图像采集系统采集每一枚鸡蛋的高光谱透射图像。3)将采集完高光谱数据的鸡蛋打破,利用常规方法,用电子天平测鸡蛋质量,用数显游标卡尺测鸡蛋蛋白高度。4)利用分蛋器将蛋清分离,校准pH计,测定每个鸡蛋蛋清的pH值,每个鸡蛋测3次。5)将每枚鸡蛋蛋清分别倒入50 mL离心管中,测量的实际黏度占当前转子和转速条件下测量黏度最大值的百分比在20%~90%之间时测定的黏度为有效数值,根据这一原则,从1、2、3、4号转子与6、12、30、60 r/min转速中选定合适的转子与转速测定待测蛋清的黏度,在室温条件下,待黏度计测定15 min后,记录10组数据。6)记录所有数据,根据鸡蛋质量与蛋白高度计算鸡蛋的HU值,用其来表示鸡蛋的新鲜度,用测定的pH值、黏度的平均值作为鸡蛋最终的酸碱度值与黏度。

1.3.3 光谱预处理

本实验选用微分处理和平滑处理2种方法。光谱微分是通过对透射光谱进行数学模拟,计算不同阶数的微分值,提取不同的光谱参数。

一阶微分处理公式为:

$$\rho'_{(\lambda_i)} = \frac{\rho_{(\lambda_{i+1})} - \rho_{(\lambda_{i-1})}}{2\Delta\lambda} \quad (2)$$

二阶微分处理公式为:

$$\rho''_{(\lambda_i)} = \frac{\rho'_{(\lambda_{i+1})} - \rho'_{(\lambda_{i-1})}}{2\Delta\lambda} \quad (3)$$

式中: λ_i 为波长/nm; $\rho'_{(\lambda_i)}$ 为 λ_i 处的一阶微分光谱; $\rho_{(\lambda_{i+1})}$ 为 λ_{i+1} 波长处的光谱; $\rho_{(\lambda_{i-1})}$ 为 λ_{i-1} 波长处的光谱; $\rho''_{(\lambda_i)}$ 为 λ_i 处的二阶微分光谱; $\Delta\lambda$ 为相邻两波段间的波长间隔/nm。

平滑处理是为了消除外界干扰产生的噪声,常用的平滑处理方法包括移动平均法、Savitzky-Golay法等。本实验选用Savitzky-Golay(S-G)平滑法进行预处理,其可以有效地消除基线漂移、倾斜等噪声。SG平滑法中的平滑点数的选择尤为重要,平滑点数过少,容易产生新的计算误差,导致模型精度下降;平滑点数过多,包

含样品信息的光谱数据被磨光丢失,同样也会导致模型精度下降,因此,必须选择合适的平滑点数,保证模型精度^[17]。

1.3.4 特征波段提取

在对全波段进行特征波段提取时,主要用到2种方法。第1种是竞争性自适应重加权(competitive adaptive reweighed sampling, CARS)算法。CARS变量选取法遵循的是“适者生存”的原则,也就是将输入的每一个变量都看成是一个独立的个体,通过自适应加权采样技术筛选出模型中回归系数绝对值较大的波长,去掉权重较小的波长并且通过交互验证优选出模型交互验证均方根误差(root mean square error of cross validation, RMSECV)最小时所对应的波长组合^[18-20]。因此,对于高光谱多维数据来说,用这种变量选择方法可以较好的简化模型。第2种是连续投影算法(successive projections algorithm, SPA)。SPA是一种前向循环的选择方法,从一个波长开始,每一次迭代时加入新的波长,直到达到指定的波长数 N 。每次循环,计算它在未选入波长的投影,投影向量 Px_j 按公式(4)计算:

$$Px_j = x_j - (x_j^T - x_k^T) x_k (x_k^T - x_k^T)^{-1} \quad (4)$$

式中: x_j 为光谱矩阵中任一列向量; x_k 表示第 n 步($n=1, 2, 3 \cdots N$)迭代时得到的波长 $k_{(n-1)}$ 中的列向量。

将投影向量最大的波长引入波长组合,直到循环 N 次,每一个新选入的波长,都与前一个线性关系最小^[21-24]。

1.3.5 多元线性回归(multiple linear regression, MLR)模型

当因变量同时受多个重要的因素影响时,且自变量与因变量之间呈现密切的线性关系,此时进行的回归分析即为MLR分析,设 Y 为因变量, $X_1, X_2 \cdots X_i$ 为自变量,则其MLR模型为:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_i X_i \quad (5)$$

式中: b_0 为常数项; $b_1, b_2 \cdots b_i$ 为回归系数。

2 结果与分析

2.1 鸡蛋HU值、pH值及黏度的相关性分析

依次将鸡蛋的HU值、pH值、黏度作为控制变量输入对剩下的2个指标进行偏相关性分析可以发现,当HU值作为控制变量时,鸡蛋pH值与黏度之间的偏相关系数为-0.428,当pH值作为控制变量时,鸡蛋HU值与黏度之间的偏相关系数为0.486,当黏度作为控制变量时,鸡蛋的HU值与pH值之间的偏相关系数为-0.484。相关系数的绝对值均下降,但仍然均高于0.3,呈一定的弱相关。

表1 鸡蛋HU值、pH值、黏度的Pearson的相关性检验

Table 1 Pearson's correlation test of HU, pH, and viscosity of eggs

指标	类目	HU	pH	黏度
HU	Pearson相关系数	1	-0.875**	0.876**
	显著性(双侧)		0	0
	样本数	200	200	200
pH	Pearson相关系数	-0.875**	1	-0.867**
	显著性(双侧)	0		0
	样本数	200	200	200
黏度	Pearson相关系数	0.876**	-0.867**	1
	显著性(双侧)	0	0	
	样本数	200	200	200

注: **在0.01水平(双侧)上显著相关。

鸡蛋的新鲜度、酸碱度与黏度随着贮藏时间的延长都有较为显著的变化,利用SPSS 13.0对鸡蛋的这3个指标进行Pearson相关性分析,由表1可知,鸡蛋HU值与pH值的Pearson相关系数为-0.875,鸡蛋黏度与pH值的Pearson相关系数为-0.867,鸡蛋HU值与黏度的Pearson相关系数为0.876,相关系数的绝对值均大于0.8,表明鸡蛋HU值、pH值、黏度任意两者之间均呈强相关。显著性(双侧)均为0,表示其均具有统计学差异。综合Pearson相关分析及偏相关分析结果可得,鸡蛋的HU值、pH值、黏度这三者之间均存在强相关性,且鸡蛋贮藏期间的品质变化都是相互影响的。

2.2 光谱数据预处理

在利用高光谱成像仪获取鸡蛋的透射图像的基础上,选择像素值为20 000的整蛋为感兴趣区域,提取感兴趣区域的平均光谱作为原始光谱。由于原始的光谱信息在两端存在较大的噪声信号,这些噪声在后续的光谱特征参数提取与建模时会造成较大的干扰,必须去除,因此本实验选取500~950 nm的波段进行研究。

在贮藏的50 d中,随着贮藏时间的延长,鸡蛋质量逐渐减少,且浓厚蛋白逐渐稀化, HU值逐渐减小,且随着浓蛋白逐渐转化为稀蛋白,蛋清的黏度减小明显,且蛋清酸碱度逐渐上升,鸡蛋的透射率逐渐降低,如图1所示。

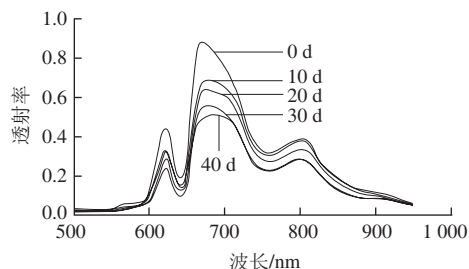


图1 不同贮藏时期的高光谱500~950 nm波段原始曲线

Fig. 1 Original hyper-spectral curves in the range of 500~950 nm for eggs during different storage periods

由于光谱在500~950 nm波段区间仍然存在较为明显的噪声信号,为了减小甚至消除其对后续研究及模型预测准确性的影响,提取特征参数建模之前需要对原始的光谱数据进行预处理。本研究中的光谱数据预处理均在MATLAB 2012(a)中完成,预处理方法主要包括样本集的划分、微分校正、平滑处理等。

2.2.1 训练集与预测集的分选

利用Kennard-Stone(KS)法对建模样本进行挑选。KS法是将光谱差异较大的样品选入训练集,而其余相近的样品选入预测集^[23]。如表2所示,通过KS法选择总体样本约1/3用来预测,剩下的样本用来建模。由于对比的理化指标实验历时长,本实验采集数据的时间间隔为10 d,且pH值的变化范围很小;黏度计测定的是蛋清搅拌均匀前的黏度,因此其测定的黏度变化范围很大,这主要是因为贮藏期间蛋清浓厚蛋白逐渐稀化,实验第1天,蛋清中绝大部分是浓蛋白,而在实验第40天,蛋清中几乎只剩稀蛋白。

表2 利用KS算法划分出的样品集的划分统计
Table 2 Division and statistics of sample sets by KS algorithm

样本数量	指标	最大值	最小值	均值	标准偏差
训练集 (130个)	HU	101.473	39.876	66.61	15.43
	pH	9.7	8.08	9.219	0.401
	黏度/(Pa·s)	1 364	40	494.9	424.39
预测集 (70个)	HU	99.028	37.244	72.38	15.86
	pH	9.61	8.17	9.06	0.45
	黏度/(Pa·s)	1 364	42.3	594.77	478.72

2.2.2 微分处理

微分处理不仅可以提供比原光谱更清晰的光谱轮廓变换,而且可以消除光谱基线漂移或者平滑背景干扰的影响。本研究对原始光谱分别进行了一阶微分与二阶微分处理,从数学意义上讲,一阶微分可以反映光谱的变化率,二阶微分可以突出光谱的弯曲点。对原始光谱及不同微分处理后高光谱数据分别建立鸡蛋HU值、pH值、黏度MLR分析模型,如表3所示。经二阶微分预处理后的全波段MLR模型的验证集相关系数高于原始全波段光谱MLR模型验证集的相关系数。

表3 采用不同预处理方法的全波段MLR模型验证集的相关系数
Table 3 The correlation coefficients with different pre-treatment methods of full band by MLR

指标	原始	一阶	二阶	一阶+S-G 平滑	二阶+S-G 平滑
HU	0.14	0.16	0.22	0.39	0.46
pH	0.46	0.35	0.56	0.63	0.41
黏度	0.31	0.32	0.52	0.69	0.57

如表3所示,综合考虑这3个指标模型的相关系数可以看出,在一阶微分的基础上进行S-G平滑处理,能够同时较好的提高鸡蛋HU值、pH值、黏度的模型的相关系

数,说明这种组合预处理的效果最好,因此本研究后续的数据处理是基于高光谱一阶光滑处后的数据上进行的。

尽管不同预处理的全波段MLR模型的相关系数均不高,但是用偏最小二乘(partial least squares, PLS)模型分别建立各指标的全波段模型,验证集的相关系数均在0.7左右,表明鸡蛋HU值、pH值、黏度与高光谱的光谱信息之间存在较好的相关性。基于特征波长建立的鸡蛋HU值、pH值、黏度PLS模型的相关数均没有MLR模型的相关系数高,因此最终统一选用全波段的MLR模型与特征波段的MLR模型进行对比研究。表3表明,在建立全波段的MLR模型时,无信息波长干扰较强,故必须通过特征波长的提取,剔除无信息的干扰波长,选出信息量大的最优波长组合,提高模型预测的准确性。

2.2.3 平滑处理

平滑处理可以有效滤去杂波,使光谱曲线更平滑。其滤波效果与窗口宽度及光谱分辨率相关,经过反复尝试,确定采用窗口宽度为8的S-G法分别对一阶微分及二阶微分光谱进行平滑。将500~950 nm波段范围内的波长从0~450 nm依次进行编号,由图2可知,平滑处理有效的消除了杂波的干扰,光谱曲线较处理前更平滑。

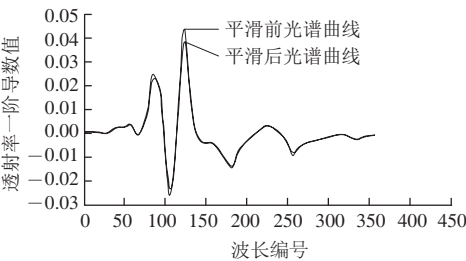
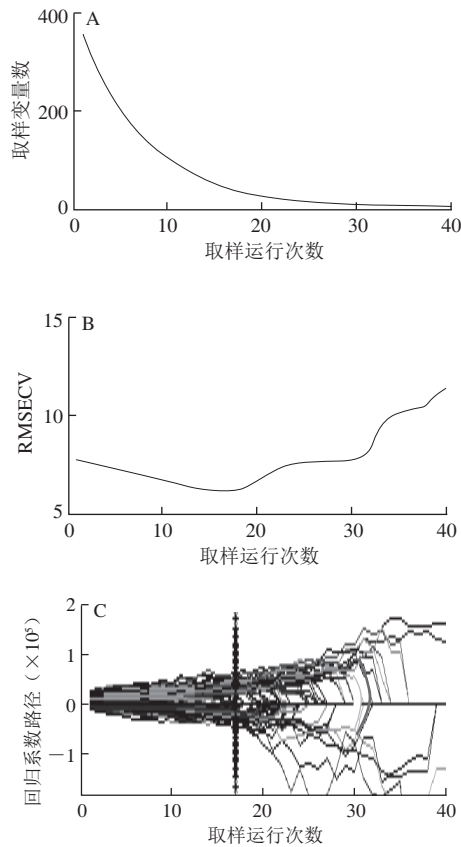


图2 平滑处理前后的光谱一阶曲线
Fig. 2 Smoothed and unsmoothed first derivative hyper-spectral curves

2.3 CARS变量选取法提取特征波长

对训练集的高光谱数据进行预处理后,采用CARS变量选取分别对鸡蛋的HU值、pH值、黏度的波长进行筛选,以HU的特征波长选取过程为例说明CARS的特征波长集的选取过程,如图3所示。由图3A可知,选取变量的变化趋势,变量数与运行次数之间呈指数关系递减^[17]。由图3B可知,RMSECV的变化趋势,RMSECV值随着取样运行次数先减小后增大,RMSECV值逐渐减小,表明光谱数据中部分无用信息被剔除,RMSECV值逐渐增大,表明光谱数据中有一些重要信息被剔除。由图3C可知,对应的RMSECV为最小值时,各变量的回归系数随着运行次数增加的变化趋势,如图3C中竖直线位置所示,当RMSECV达到最小值时,共取样运行17次。



A. 取样变量数; B. RMSECV; C. 回归系数路径。

图3 鸡蛋pH值高光谱特征波长选取图

Fig. 3 Hyper-spectral characteristic wavelength selected for detection of egg pH

经过反复的参数尝试,在RMSECV均为最小值时,最终选取的鸡蛋HU值、pH值、黏度具体的波段如表4所示。基于CARS变量选取法选择的HU值预测的最优波长组合中有48个波长,pH值预测的最优波长组合中有15个波长,黏度预测的最优波长组合中有6个波长。

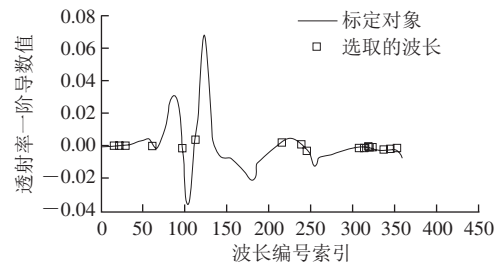
表4 基于CARS变量选取法选择的最优波长组合

Table 4 Optimal wavelength combinations based on CARS method

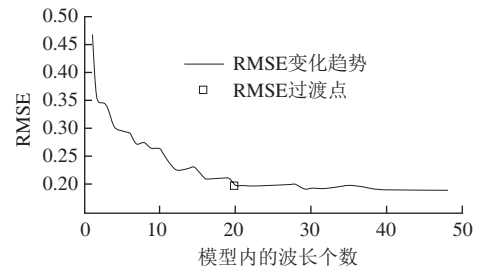
指标	波长/nm
HU (48个)	599.73, 600.98, 600.98, 616.06, 618.57, 637.41, 638.67, 656.25, 673.84, 675.09, 685.14, 685.14, 687.66, 688.91, 690.17, 703.98, 705.24, 706.5, 707.75, 709.01, 726.6, 727.85, 729.11, 739.16, 740.41, 741.67, 742.93, 744.18, 754.23, 756.74, 758, 881.1, 882.36, 883.61, 884.87, 886.13, 887.38, 888.64, 889.89, 891.15, 892.41, 893.66, 908.74, 909.99, 912.5, 920.04, 921.3, 922.55
pH (15个)	599.73, 627.36, 653.74, 672.58, 688.91, 690.17, 706.5, 726.6, 727.85, 758, 891.15, 829.41, 904.97, 906.22, 907.48
黏度 (6个)	686.4, 688.91, 758, 759.26, 799.45, 881.1

2.4 SPA提取特征波长

将KS法挑选出来的70个高光谱数据样本个体作为测试集输入,用于检验模型的预测性能,将其余的130个数据样本作为训练集输入,用于选择最优特征波长和建模。本实验对SPA优选出的波长序列分别与鸡蛋的HU值、pH值、黏度建立一个MLR模型。



A. SPA选择的最优特征波长



B. 模型RMSE的变化

图4 SPA提取特征波长

Fig. 4 SPA extraction of characteristic wavelengths

如图4所示,以鸡蛋pH值为例简要说明SPA的选取过程。根据SPA的原理,利用均方根误差(root mean square error, RMSE)来决定所建模型的优劣,选出RMSE减小过程中由快变慢的过渡值^[25]。在过渡值前,由于冗余波长的不断剔除,RMSE值随着变量的增加下降较快,在过渡值后,由于冗余的波长已经基本剔除,因此RMSE随着变量的增加变化稳定。选取RMSE过渡值对应的变量数作为最优的波长组合。由表5可知,基于SPA变量选取法选择的HU值预测的最优波长组合中有19个波长,pH值预测的最优波长组合中有20个波长,黏度预测的最优波长组合中有20个波长。

表5 基于SPA变量选取法选择的最优波长组合

Table 5 Optimal wavelength combinations based on SPA

指标	波长/nm
HU (19个)	505.51, 510.54, 533.15, 534.41, 578.37, 584.65, 585.91, 642.43, 760.51, 795.68, 799.45, 800.71, 815.78, 889.89, 922.55, 938.88, 942.65, 945.16, 946.42
pH (20个)	510.54, 511.8, 528.13, 533.15, 534.41, 578.37, 622.34, 642.43, 771.82, 800.71, 809.5, 886.13, 889.89, 892.41, 894.92, 902.46, 904.97, 925.07, 938.88, 943.91
黏度 (20个)	505.51, 513.05, 516.82, 525.61, 584.65, 585.91, 619.82, 622.34, 641.18, 642.43, 712.78, 756.74, 760.51, 806.99, 889.89, 897.43, 908.74, 938.88, 942.65, 946.42

2.5 二次特征波长选取

用CARS与SPA选取的特征波长组合,尽管能够有效削减光谱波数,但是选取的波长数仍然相对较多,因此本研究继续在CARS提取的最优波长组合的基础上,

利用SPA进行二次波长选取, 最终选取的特征波长。由表6可知, 基于CARS选取的特征波长组合的SPA二次特征波长选取法选择的HU值预测的最优波长组合中有13个波长, pH值预测的最优波长组合中有10个波长, 黏度预测的最优波长组合中有6个波长。对比可见, 二次特征波长提取可以在原先最优波长组合上更进一步的削减光谱波数, 使最终建立的模型结构得到进一步的简化与优化。

表6 基于CARS选取的特征波长组合的SPA二次特征波长选取组合
Table 6 Optimal wavelength combinations based on CARS methods and SPA

指标	波长/nm
HU	618.57、638.67、656.25、673.84、688.91、703.98、726.6、758.0、886.13、891.15、892.41、908.74、922.55
pH	599.73、627.36、653.74、672.58、690.17、706.5、758.0、829.41、904.97、907.48
黏度	686.4、688.91、758.0、759.26、799.45、881.1

2.6 鸡蛋HU值、pH值、黏度预测模型

表7 基于不同特征波长提取方法建立的MLR模型效果
Table 7 Comparison of MLR models based on different characteristic wavelength extraction methods

指标	CARS			SPA			CARS-SPA		
	RMSE	R	N	RMSE	R	N	RMSE	R	N
HU	7.45	0.91	48	7.30	0.92	19	6.36	0.94	13
pH	0.18	0.94	15	0.20	0.92	20	0.17	0.95	10
黏度	149	0.95	6	151	0.96	20	149	0.95	6

如表7所示, 对CARS、SPA及CARS-SPA选取的最优波段组合分别建立MLR模型, 基于CARS建立的鸡蛋HU值、pH值、黏度MLR模型预测值与真实值的相关系数均大于0.9, 此结论与Wang Qiaohua等^[8]对鸡蛋新鲜度的研究结论相同, 且基于SPA建立的鸡蛋HU值、pH值、黏度MLR模型预测值与真实值的相关系数也均大于0.9, 此预测效果较Zhang Wei等^[6]对鸡蛋新鲜度的预测效果好。由此可以看出, 特征波段的提取能够有效消除冗余的波段, 模型的复杂度大大降低, 且大大提高模型稳定性与准确性。

基于CARS-SPA的二次波段提取建立的鸡蛋HU值、pH值、黏度MLR模型预测值与真实值的相关系数均在0.95左右, 均高于一次波段提取的模型相关系数, 其相关系数高于Wang Qiaohua^[8]、Zhang Wei^[6]等建立的模型的相关系数。且RMSE值均较CARS-MLR、SPA-MLR有所降低, 由此可见二次波段提取比一次波段提取效果好, 其能够更进一步的有效降低模型的复杂度, 提高模型的稳定性与准确性, 优化模型。

3 结论

本研究以比白壳鸡蛋更难以透光的褐壳鸡蛋为研究对象, 研究基于高光谱技术与鸡蛋新鲜度有关的内部理

化指标的无损检测方法, 建立鸡蛋HU值、pH值、黏度的MLR模型。用KS法挑选出70个样本作为模型的预测集, 其余的130个样本作为模型的训练集。对不同的预处理方法分别建立模型预测, 发现高光谱一阶微分平滑数据的预测效果较好。鸡蛋的新鲜度、酸碱度、黏度之间存在较强的相关性, 且彼此之间互相影响。基于高光谱一阶微分的全波段MLR模型验证集的鸡蛋HU值、pH值、黏度的相关系数分别为0.39、0.63、0.69, 基于CARS特征参数提取的MLR模型验证集的相关系数分别为0.91、0.94、0.95, 基于SPA特征参数提取的MLR模型验证集的相关系数分别为0.92、0.92、0.96。因此, 特征参数提取可以有效的降低模型复杂度, 提高模型准确性。基于CARS提取的鸡蛋HU值、pH值、黏度的最佳波段组合数分别为48、15、6个, 基于SPA提取的鸡蛋HU值、pH值、黏度的最佳波段组合数分别为19、20、20个, 基于CARS-SPA二次波段提取的最佳波段组合数分别为13、10、6个。二次波段提取可以进一步削减贡献较小的波长, 进一步简化模型结构。与全波段建模相比, 有效的特征参数提取均可以较好的提高预测模型的准确性。不论是用CARS算法, 还是用SPA提取特征波长, 都可以大大的提高预测模型的准确性。通过特征波长的提取, 能够剔除冗余的无或少信息量的波段, 选出具有较高信息量的波长来表征鸡蛋的HU值、pH值及黏度, 简化模型。在CARS提取的特征波长组合的基础上采用SPA再次进行特征波长提取, 并建立MLR模型, 建立的鸡蛋HU值、pH值、黏度MLR模型验证集的相关系数分别为0.94、0.95、0.95。且二次特征波段提取可以进一步削减贡献较小的波长, 选出有用的光谱表征信息, 去除更多的无关信息, 继而优化模型。

参考文献:

- [1] 王巧华, 文友先, 林雪冬, 等. 鸡蛋透射光图像形态特征与其新鲜度的相关关系[J]. 农业工程学报, 2008, 24(3): 179-184. DOI:10.3321/j.issn.1002-6819.2008.03.036.
- [2] 王栓巧, 郁志宏, 韩宝, 等. 基于机器视觉鸡蛋品质检测机构的研究[J]. 机械设计与制造, 2009(12): 84-86. DOI:10.3969/j.issn.1001-3997.2009.12.033.
- [3] 刘燕德, 彭彦颖, 孙旭东. 鸡蛋蛋白pH可见/近红外光谱在线检测信息变量提取研究[J]. 江西农业大学学报, 2010, 32(5): 1075-1080. DOI:10.3969/j.issn.1000-2286.2010.05.031.
- [4] 刘燕德, 周延睿, 彭彦颖. 基于近红外漫反射光谱检测鸡蛋品质[J]. 光学精密工程, 2013, 21(1): 40-46. DOI:10.3788/OPE.20132101.0040.
- [5] 侯卓成, 杨宁, 李俊英, 等. 傅里叶变换近红外反射用于鸡蛋品质的研究[J]. 光谱学与光谱分析, 2009, 29(8): 2063-2067. DOI:10.3964/j.issn.1000-0593(2009)08-2063-04.
- [6] ZHANG Wei, PAN Leiqing, TU Sicong, et al. Non-destructive internal quality assessment of eggs using a synthesis of hyperspectral imaging and multivariate analysis[J]. Journal of Food Engineering, 2015, 157: 41-48. DOI:10.1016/j.jfoodeng.2015.02.013.

- [7] 张伟, 潘磊庆, 林红英. 基于高光谱图像检测禽蛋新鲜度的研究[J]. 南京晓庄学院学报, 2015, 11(6): 46-51. DOI:10.3969/j.issn.1009-7902.2015.06.011.
- [8] WANG Qiaohua, ZHOU Kai, WANG Caiyun, et al. Egg freshness detection based on hyperspectral image technology[J]. Advance Journal of Food Science and Technology, 2015, 7(8): 652-657.
- [9] 潘磊庆, 张伟, 于敏莉, 等. 基于高光谱图像的鸡种蛋孵化早期胚胎性别鉴别[J]. 农业工程学报, 2016, 32(1): 181-187. DOI:10.11975/j.issn.1002-6819.2016.01.025.
- [10] 祝志慧, 刘婷, 马美湖. 基于高光谱信息融合和相关向量机的种蛋无损检测[J]. 农业工程学报, 2015, 31(15): 285-292. DOI:10.11975/j.issn.1002-6819.2015.15.039.
- [11] 张伟, 潘磊庆, 屠康. 利用高光谱透射图像检测鸡种蛋早期孵化[J]. 农业工程学报, 2012, 28(21): 149-155. DOI:10.3969/j.issn.1002-6819.2012.21.021.
- [12] NICOLAS A N, MICHAEL N, SHIV P, et al. Prediction of egg freshness and albumen quality using visible/near infrared spectroscopy[J]. Food Bioprocess Technologies, 2011, 4: 731-736. DOI:10.1007/s11947-009-0265-0.
- [13] LIN H, ZHAO J W, SUN L, et al. Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis[J]. Innovative Food Science and Emerging Technologies, 2011, 12: 182-186. DOI:10.1016/j.ifset.2011.01.008.
- [14] KAROUI R, KEMPS B, BAMELIS F, et al. Methods to evaluate egg freshness in research and industry: a review[J]. European Food Research and Technology, 2006, 222: 727-732. DOI:10.1007/s00217-005-0145-4.
- [15] KEMPS B J, de KETELAERE B, BAMELIS F R, et al. Albumen freshness assessment by combining visible near-infrared transmission and low-resolution proton nuclear magnetic resonance spectroscopy[J]. Poultry Science, 2007, 86(4): 757-759. DOI:10.1093/ps/86.4.752.
- [16] IQBAL A, SUN D W, ALLEN P. Prediction of moisture, color and pH in cooked, presliced turkey hams by NIR hyperspectral imaging system[J]. Journal of Food Engineering, 2013, 117: 42-51. DOI:10.1007/s11947-014-1327-5.
- [17] 陈华舟, 潘涛, 陈洁梅, 等. 多元散射校正与Savitzky-Golay平滑模式的组合优选应用于土壤有机质的近红外光谱分析[J]. 计算机与应用化学, 2011, 28(5): 518-522. DOI:10.3969/j.issn.1001-4160.2011.05.002.
- [18] 李洪波, 郭志明, 黄文倩, 等. 应用CARS和SPA算法对草莓SSC含量NIR光谱预测模型中变量及样本筛选[J]. 光谱学与光谱分析, 2015, 35(2): 372-378. DOI:10.3964/j.issn.1000-0593(2015)02-0372-07.
- [19] 刘燕德, 施宇, 蔡丽君, 等. 基于CARS算法的脐橙可溶性固形物近红外在线检测[J]. 农业机械学报, 2013, 44(9): 138-145. DOI:10.6041/j.issn.1000-1298.2013.09.025.
- [20] 吴静姝, 徐云. 基于CARS-PLS的食用油脂脂肪酸近红外定量分析模型优化[J]. 农业机械学报, 2011, 42(11): 162-167. DOI:10.3969/j.issn.1000-1298.2011.10.032.
- [21] 高洪智, 卢启鹏, 丁海泉, 等. 基于连续投影算法的土壤总氮近红外特征波长的选取[J]. 光谱学与光谱分析, 2009, 29(11): 2951-2954. DOI:10.3964/j.issn.1000-0593(2009)11-2951-04.
- [22] 洪涯, 洪添胜, 代芬, 等. 连续投影算法在砂糖橘总酸无损检测中的应用[J]. 农业工程学报, 2010, 26(增刊2): 380-383. DOI:10.3969/j.issn.1002-6819.2010.z2.072.
- [23] 吴迪, 金春华, 何勇. 基于连续投影算法的光谱主成分组合优化方法研究[J]. 光谱学与光谱分析, 2009, 29(10): 2734-2737. DOI:10.3964/j.issn.1000-0593(2009)10-2734-04.
- [24] 吴迪, 宁纪锋, 刘旭, 等. 基于高光谱成像技术和连续投影算法检测葡萄果皮花色苷含量[J]. 食品科学, 2014, 35(8): 57-62. DOI:10.7506/spkx1002-6630-201408010.
- [25] 廖宜涛, 樊玉霞, 成芳, 等. 连续投影算法在猪肉pH值无损检测中的应用[J]. 农业工程学报, 2010, 26(1): 379-383. DOI:10.3969/j.issn.1002-6819.2010.z1.067.