

文章编号:1000-8551(2012)08-1154-05

# 基于最小二乘支持向量回归的鹅肉弹性的 可见-近红外光谱测定

赵进辉 袁海超 刘木华 涂冬成 吁芳

(江西农业大学工学院,江西 南昌 330045)

**摘要:**为简化鹅肉弹性的可见-近红外光谱模型和提高预测精度,采用联合区间偏最小二乘法(synergy interval partial least square algorithm, siPLS)结合遗传算法(Genetic algorithm, GA)提取可见-近红外光谱特征波长,用最小二乘支持向量回归(least square support vector for regression, LSSVR)建立鹅肉弹性的预测模型。试验以万能试验机获取恢复距离S作为鹅肉弹性实际值。在模型建立过程中,先利用sym8小波的2层分解对原始的可见-近红外光谱进行光谱预处理;然后用siPLS优选出4个特征光谱子区间(分别为第3、5、9、13子区间),在这4个特征光谱子区间内继续用GA优选出74个特征波长,并建立基于LSSVR的鹅肉弹性的预测模型。模型预测集的决定系数( $R^2$ )和预测均方根误差(root mean squared error of prediction, RMSEP)分别为0.9096和0.0588。试验结果表明,siPLS结合GA法能够有效提取光谱中的鹅肉弹性对应的特征波长,有利于提高LSSVR模型预测鹅肉弹性的精度。

**关键词:**可见-近红外光谱;弹性;最小二乘支持向量回归;联合区间偏最小二乘法;遗传算法

## DETERMINATION OF ELASTICITY OF GOOSE MEAT USING VISIBLE-NEAR INFRARED SPECTROSCOPY AND LSSVR

ZHAO Jin-hui YUAN Hai-chao LIU Mu-hua TU Dong-cheng YU Fang

(College of Engineering, Jiangxi Agricultural University, Nanchang, Jiangxi 330045)

**Abstract:** To improve and simplify the prediction model of elasticity of goose meat, the optimized characteristic spectral wavelengths were extracted from NIR spectra of goose meat combined with synergy interval PLS (siPLS) and genetic algorithm (GA), then prediction model of elasticity of goose meat was developed using least squares support vector regression (LSSVR). Recovery distances obtained by universal testing machine were used as actual value of elasticity of goose meat. Firstly, sym8 wavelet with two levels decomposition was used to complete the pretreatment of the original visible-near infrared spectroscopy. Secondly, 4 subintervals, i. e. No. 3, 5, 9 and 13 were selected by siPLS, and 74 characteristic wavelengths were selected in these spectral regions by GA. Finally, 74 characteristic wavelengths were used to build prediction model based on LSSVR. The determination coefficient ( $R^2$ ) and the root mean squared error of prediction (RMSEP) for LSSVR prediction model were 0.9096 and 0.0588, respectively. This work proved that siPLS-GA could determine characteristic spectral wavelengths and improve the prediction accuracy of LSSVR model.

**Key words:** visible-near infrared spectroscopy; elasticity; least squares support vector regression (LSSVR); synergy interval PLS (siPLS); genetic algorithm (GA)

收稿日期:2012-05-10 接受日期:2012-08-14

基金项目:国家高技术研究发展计划(863计划)项目(2008AA10Z209),国家自然科学基金项目(31101295)

作者简介:赵进辉(1978-),男,湖南华容人,博士,副教授,主要从事图像处理与农畜产品无损检测研究。E-mail:zjhxiaocao@sina.com

通讯作者:刘木华(1969-),男,江西赣州人,博士,教授,主要从事农畜产品无损检测研究。E-mail:suikelmh@sina.com

鹅肉是我国传统的、具有代表性且深受消费者喜爱的禽肉食品。随着人们生活水平的提高,人们对鹅肉品质要求也越来越高。弹性是评价鹅肉品质好坏的一个最直接、最重要的指标。弹性好则品质好,弹性差则品质差<sup>[1]</sup>。此外,鹅肉弹性大小还跟鹅肉新鲜度有关<sup>[2]</sup>。因此,鹅肉弹性大小的测定对判定鹅肉新鲜度及其品质有着重要意义。传统的鹅肉弹性测定方法有主观评价和仪器测量,主观评价方法由于受到人和外界因素的影响,误差较大。仪器测量方法虽然结果可靠,但取样过程复杂,比较费时、费力,达不到快速、在线检测要求<sup>[1]</sup>。因此,如何运用一种快速、准确、在线的检测方法获取鹅肉弹性已受到国内外学者的关注。

可见-近红外光谱技术是一种快速、简便、高效、准确、无污染的方法,被广泛用于食品、农业、医药等多个领域<sup>[3~6]</sup>。目前,关于近红外光谱技术在弹性检测方面的应用很少见。罗阳等<sup>[1]</sup>探讨了利用近红外光谱分析技术测定鱼丸弹性的可能性研究。在鹅肉等肉类弹性的近红外光谱测定方面研究尚未见相关文献报道。

与人工神经网络(artificial neural network, ANN)相比,最小二乘支持向量回归(LSSVR)既具有解决非线性问题的显著优势、较好的控制过拟合现象、具有全局最优及更好的泛化能力等支持向量回归(support vector for regression, SVR)方法优点,又具有更好的抗噪声能力和更快运算速度的特点,使得其在模式识别、函数逼近、数据挖掘等领域得到广泛应用<sup>[7~11]</sup>。鉴于LSSVR的优点,本研究将LSSVR应用到鹅肉弹性的可见-近红外光谱测定中。先运用联合区间偏最小二乘法(siPLS)优选经小波去噪后的可见-近红外光谱特征子区间,然后用遗传算法(GA)在特征子区间内进行特征波长的提取,在此基础上建立基于LSSVR的鹅肉弹性的可见-近红外光谱测定模型,尝试建立一种鹅肉弹性的快速、无损检测方法。

## 1 材料与方法

### 1.1 试验材料与仪器设备

从江西农业大学菜市场随机挑选48只活体莲花白鹅,体重在3.5~5.5kg之间;QualitySpec Pro台式可见光/近红外光谱仪(Analytical Spectral Devices, Inc., USA),光谱测量范围为350~1800nm,其中在350~1000nm采样间隔为1.4nm,分辨率为3nm,在1000~1800nm采样间隔为2nm,分辨率为10nm;WDW-10C型华龙微机控制电子万能试验机(上海华龙测试仪器

有限公司)。

### 1.2 试验方法

(1)样本前处理。将活鹅宰杀烫毛、褪毛后于铁架上悬挂半个小时;然后用手术刀剔取出胸脯肉并对表面进行修整,同时对胸脯肉进行标号。(2)样本的可见-近红外反射光谱的采集。采集样本光谱前,用系统配置优化标准白板进行标定。采集样本光谱时,将样本置于密闭的暗箱中,用光谱仪自带附件高密度光源检测探头采集样本的可见-近红外反射光谱。每个样本在3个不同部位采集得到3个光谱数据,并取其平均值作为该样本的原始光谱。(3)鹅肉弹性的物理测定。先将万能试验机的主要参数设置如下:压头下行和返回速度均为150mm/min,测试过程中压头速度为10mm/min,下行位移为3mm。然后将修剪后的肉样置于固定的不锈钢托板上,用万能试验机进行鹅肉弹性测试,每个肉样取3个不同的部位进行测试,并以3个测试点的恢复距离S(在作用力消失后鹅肉靠自身弹力恢复的距离)的均值作为该肉样弹性的最终测量值。试验共获得48个样本,其中36个作为校正样本,12个作为预测样本,鹅肉弹性(恢复距离S)的测量结果如表1所示。

表1 鹅肉弹性(恢复距离)的测量结果  
Table 1 Measurement results of elasticity  
(recovery distance) of goose meat

项目 item	校正集 calibration set	预测集 prediction set
范围 range(mm)	0.4172 ~ 1.0107	0.482 ~ 0.9539
均值 mean value(mm)	0.6702	0.6821
标准方差 standard deviation(mm)	0.1409	0.1432
数量 number	36	12

### 1.3 数据分析

用Matlab软件分析可见-近红外反射光谱与鹅肉弹性之间的关系。先用小波分析进行光谱预处理,然后用siPLS进行特征子区间的选择和用GA法进行特征波长的提取,最后应用LSSVR建立鹅肉弹性的可见-近红外反射光谱预测模型。

## 2 试验结果与分析

### 2.1 基于小波的光谱预处理

图1给出1个鹅肉样本的原始光谱和相应的小波去噪光谱,为更好地区分原始光谱与小波去噪光谱,已将图中不同波长下的小波去噪光谱的相对反射率都增加了0.1。由图1可见,在350~430nm波长范围内存

在严重噪声,这些噪声会影响鹅肉弹性有效光谱信息的利用。因此,在提取光谱特征波长前有必要进行光谱预处理以消除噪声。根据小波变换在近红外光谱除噪和有效信息提取方面的优势<sup>[12~14]</sup>,本研究尝试用 sym8 小波消噪方法除噪,并对 2,3,4,5,6 共 5 层分解的小波除噪效果进行对比分析。本研究从总样本中随机选择 4 个样本,并统计不同层次分解的小波的信噪比(SNR)和均方根误差(RMSE)的平均值,以比较不同层次分解的小波除噪效果。从表 2 可以看出,以 2 层小波分解时的信噪比最大,均方根误差最小,说明 2 层小波分解时的消噪效果最好。同时从图 1 可以看出,与原始光谱相比,经过 2 层小波分解的小波除噪光谱曲线变得更加光滑。因此,在后面试验中选择 2 层小波分解。

表 2 不同层次分解的小波消噪效果

Table 2 De-noising effect for different levels decomposition of wavelet

分解层数 levels decomposition	信噪比 SNR	均方根误差 RMSE
2	105.429	0.0036
3	102.341	0.0042
4	99.0065	0.0049
5	83.4034	0.0108
6	83.1959	0.0109

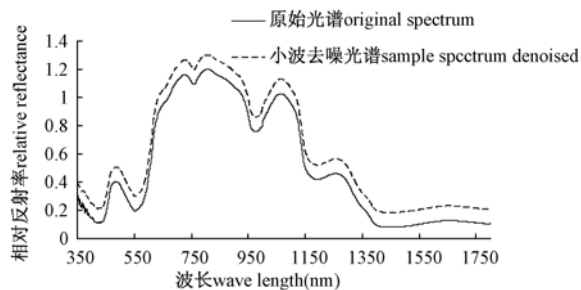


图 1 原始光谱和小波去噪后的光谱

Fig. 1 Original sample spectrum and sample spectrum denoised by wavelet

## 2.2 基于 siPLS 的特征光谱区间选择

鹅肉的可见-近红外光谱的谱峰较宽,在多个波长处存在吸收峰、光谱谱峰重叠现象严重且数据量大,这些局限性使得在建立鹅肉弹性预测模型前,有必要采用一定的化学计量学方法从中提取有用信息<sup>[15]</sup>。本研究采用 siPLS-GA 法提取与鹅肉弹性有关的有用光谱信息,即先用 siPLS 选择特征光谱子区间,然后在特征光谱子区间内用 GA 提取特征波长。试验将整个小波去噪光谱划分为  $n$  个子区间, $n$  的取值范围为 15 ~

30,通过交互验证法优化被划分的区间数与联合子区间,联合子区间数取值范围为 2 ~ 4,得到的最小交互验证均方根误差(RMSECV)值所对应模型的子区间组合即为特征光谱区间组合,特征光谱区间选择结果如表 3 所示。

表 3 基于 siPLS 的特征光谱区间选择结果

Table 3 Results of feature spectral regions selected by siPLS method

总区间数 total interval number	主成分因子 PCs	被选择的子区间 subinterval selected	交互验证均 方根误差 RMSECV
15	5	[ 2 3 8 12 ]	0.02846
16	5	[ 2 3 11 16 ]	0.02905
17	11	[ 3 5 9 13 ]	0.02279
18	5	[ 3 18 ]	0.02734
19	5	[ 3 13 ]	0.0269
20	5	[ 3 9 14 18 ]	0.02637
21	7	[ 3 11 15 ]	0.02507
22	7	[ 3 19 21 ]	0.02543
23	9	[ 3 10 12 19 ]	0.0246
24	11	[ 3 7 11 22 ]	0.02573
25	9	[ 4 7 13 19 ]	0.02304
26	8	[ 4 25 26 ]	0.02367
27	7	[ 4 8 12 19 ]	0.02469
28	7	[ 4 8 12 20 ]	0.02436
29	8	[ 4 15 20 ]	0.0251
30	6	[ 4 26 ]	0.0258

从表 3 可以看出,当整个光谱被划分为 17 个子区间(前 6 个区间各包含 86 个波长,剩余区间各包含 85 个波长),主成分因子数为 11,联合子区间为 [ 3 5 9 13 ] 时,建模的 RMSECV 值最小,因此本研究将子区间 [ 3 5 9 13 ] 作为后续研究的特征光谱区间。图 2 给出了基于 siPLS 的最佳特征光谱子区间组合,图中黑色条带区表示被选中的 4 个特征光谱子区间,各子区间对应的波长范围分别为:522 ~ 607nm,694 ~ 779nm,1036 ~ 1120nm 和 1376 ~ 1460nm,共 342 个波长。

## 2.3 基于 GA 的特征波长的提取

基于 siPLS 的特征光谱区间选择结果是将波长数从 1451 个压缩到了 342 个,波长压缩率为 76.43%,由此剔除了大量与鹅肉弹性无关的波长,但 342 个波长中仍可能包含一些冗余信息,因此有必要进一步缩小有用波长的范围。本研究采用 GA 算法结合交互验证均方根误差(RMSECV)在 [ 3 5 9 13 ] 区间内进一步提取特征波长。GA 算法的相关参数为:初始群体为 30,变异概率  $P_m$  为 0.01,交叉概率  $P_c$  为 0.5,以 RMSECV 值构造 GA 算法的适应度函数,当迭代 100 次时本次由大到小重新排序并依次作为 PLS 模型的输入。选

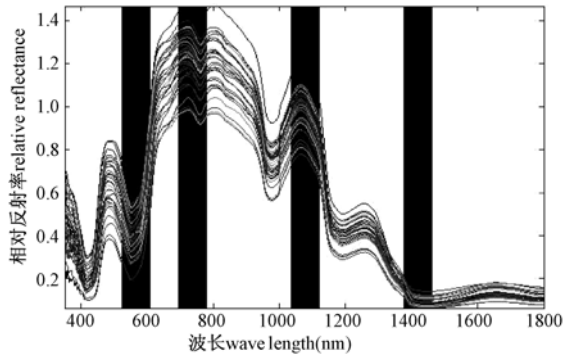


图2 基于 siPLS 选择的最佳特征光谱区间

Fig. 2 Characteristic spectral regions selected by siPLS

代终止后,输入波长个数为 74(即 342 个波长按频率大小重新排序后的前 74 个波长)时,对应的 RMSECV 最小,其大小为 0.0208。因此,将这 74 个波长作为后续建模的最佳输入特征向量。

#### 2.4 基于 LSSVR 模型的预测结果

以 GA 优选出的 74 个特征波长作为 LSSVR 模型的输入建立预测模型,预测结果如图 3 所示。LSSVR 预测模型以径向基函数作为核函数,并采用交叉验证方法对正则化参数  $\gamma$  和核函数宽度  $\sigma^2$  进行参数优化。从图 3 和表 4 可以看出,预测集的决定系数( $R^2$ )和预测均方根误差(RMSEP)分别为 0.9096 和 0.0588,与全光谱相比,不仅入选的波长数从 1451 个减少到了 74 个,使模型得到简化,而且还使预测模型的精度和速度得到提高,使之满足鹅肉弹性的快速测定要求。

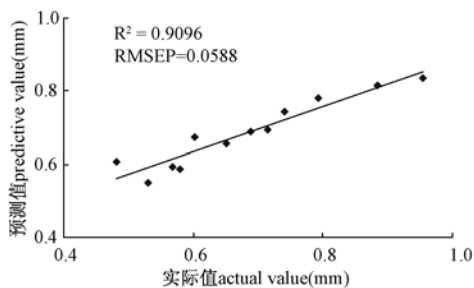


图3 鹅肉弹性预测值和实际值之间的关系

Fig. 3 Relation between actual value and predictive value of elasticity of goose meat

#### 2.5 不同模型的预测结果比较

为比较本文研究方法建模效果,共建立了全光谱的 LSSVR 模型、本研究优选的 74 个特征波长作为输入的 LSSVR 和 BP 模型 3 种模型,并对它们的预测结果进行了比较,这 3 种模型的预测结果如表 4 所示。从表 4 可以看出,在全光谱范围内建立的 LSSVR 模型的预测效果最差,这是因为全光谱范围内含有大量与鹅肉弹性无关的冗余信息、噪声等,这些信息不可避免

的影响到模型的预测精度。而以 siPLS-GA 优选波长建立的 LSSVR 模型预测效果最好,这表明 siPLS 和 GA 分别能从光谱中有效挑选出鹅肉弹性对应的特征光谱区间和特征波长,且用 LSSVR 模型预测鹅肉弹性是可行的。

表 4 不同模型的预测结果比较

Table 4 Results and comparison of different models

光谱范围 spectral range(nm)	模型 model	决定系 数 $R^2$	预测均方根 误差 RMSEP
全光谱 full spectrum	LSSVR	0.6236	0.1038
siPLS-GA 优选波长 wavelength selected by siPLS-GA	LSSVR	0.9096	0.0588
siPLS-GA 优选波长 wavelength selected by siPLS-GA	BP	0.8945	0.0636

### 3 结论

本研究探讨了利用可见-近红外光谱结合 siPLS、GA 和 LSSVR 等化学计量学方法来快速测定鹅肉弹性。为有效获取可见-近红外光谱中的与鹅肉弹性相关的信息,本研究先用 siPLS 优选了 4 个特征光谱子区间(分别为第 3、5、9、13 子区间)。在这 4 个特征光谱子区间内继续用 GA 优选出 74 个特征波长,并建立了 LSSVR 预测模型,预测集的决定系数和预测均方根误差分别为 0.9096 和 0.0588。研究结果表明,siPLS-GA 法能够有效提取光谱中的鹅肉弹性对应的特征波长,且用 LSSVR 模型预测鹅肉弹性是可行的。

#### 参考文献:

- [1] 罗阳,王锡昌,邓德文. 近红外光谱分析检测鱼丸弹性的可行性研究[J]. 食品科学,2008, 29(8): 530-533
- [2] 蒋予箭,周雁. 肉类弹性测定方法的研究[J]. 食品科学,2002, 23(4): 99-102
- [3] 吴军,白琪林,苏胜宝,陈绍江,孟庆翔,严衍禄. 近红外反射光谱法分析玉米秸秆纤维素含量的研究[J]. 分析化学,2005, 33(10): 1421-1423
- [4] 申艳,张晓平,梁爱珍,时秀焕,范如芹,杨学明. 近红外光谱法在土壤有机质研究中的应用[J]. 核农学报,2010, 24(1): 199-207
- [5] 李勇,魏益民,王锋. 影响近红外光谱分析结果准确性的因素[J]. 核农学报,2005, 19(3): 236-240
- [6] 赵进辉,刘木华,吁芳,沈洁,涂东成. 鸭肉中谷氨酸含量的可见-近红外光谱测定研究[J]. 核农学报,2011, 25(3): 529-533
- [7] 吴荣晖,邵学广. 近红外光谱用于植物样品中水溶性氯离子含量的测定[J]. 光谱学与光谱分析,2006, 26(4): 617-619
- [8] 姜云春,邱静,刘冠军. 基于最小二乘支持向量机建模的电液伺服系统故障检测方法[J]. 机床与液压,2007, 35(1): 229-231

- [ 9 ] 郭 辉,刘贺平,王 玲. 基于最小二乘支持向量机对偶优化问题的核偏最小二乘[J]. 北京科技大学学报,2006,28(8):790 - 793
- [10] 黎 锐,李存军,徐新刚,王纪华,杨小冬,黄文江,潘瑜春. 基于支持向量回归(SVR)和多时相遥感数据的冬小麦估产[J]. 农业工程学报,2009,25(7):114 - 117
- [11] 白茂金,陈 刚,刘 青,张作鹏,张雪君. 一种基于 LS-SVR 的电网在线暂态稳定性预测新方法[J]. 电力系统保护与控制, 2008,36(19):9 - 14
- [12] 韩小平,左月明,李灵芝. 水培番茄施氮量近红外光谱预测模型的研究[J]. 光谱学与光谱分析,2010,30(9): 2479 - 2483
- [13] 代 芬,洪添胜,岳学军,张 昆,洪 涯. 砂糖橘可溶性总糖可见—近红外光谱无损检测[J]. 农业机械学报,2011,42(4): 133 - 138
- [14] 代 芬,李 岩,冯 栋. 小波去噪在基于近红外光谱的砂糖橘水分检测的应用[J]. 湖南科技学院学报,2011,32(8): 36 - 39
- [15] 石吉勇,邹小波,赵杰文,殷晓平,陈正伟. 一种近红外光谱特征子区间选择新算法[J]. 光谱学与光谱分析,2012,30(12): 3199 - 3202

(责任编辑 王媛媛)