

文章编号:1005-0523(2017)02-0105-07

基于LS-SVM高光谱成像鱼新鲜度鉴别

章海亮,叶青,罗微,刘雪梅

(华东交通大学电气与自动化工程学院,江西南昌330013)

摘要:基于高光谱成像光谱信息的鱼新鲜度(鱼不同冷冻时间以及冻融次数)鉴别。首先,提取鱼样本感兴趣区域(region of interest, ROI)光谱,分别采用蒙特卡罗无信息变量消除(Monte Carlo free information variable elimination, MCVE),连续投影算法(successive projections algorithm, SPA)和随机青蛙算法(random frog, RF)提取特征波长,三种算法分别得到90,31和49个特征变量,采用最小二乘支持向量机作为分类模型,将90,31和49个特征变量作为LS-SVM模型的输入变量建立分类模型,基于SPA-LS-SVM和MCVE-LS-SVM模型预测集识别率都达到了98%,而采用RF-LS-SVM建立的模型取得了较差的预测结果,模型预测集识别率都只是达到了88%。结果表明,SPA-LS-SVM作为分类模型优于其他模型,SPA选择的特征波长,不但可以简化模型,还可以提高模型的预测精度,基于高光谱成像技术可以用于鱼新鲜度(鱼不同冷冻时间以及冻融次数)鉴别。

关键词:蒙特卡罗无信息变量消除;连续投影;随机青蛙;LS-SVM

中图分类号:O433

文献标志码:A

DOI:10.16749/j.cnki.jecjtu.2017.02.015

随着人们消费水平的提高,多宝鱼由于其丰富的营养价值如低脂肪,高蛋白,富含多种维生素和矿物质,日益受到人们的青睐。但是由于鱼类产品普遍具有水分含量高和易腐败特点,蛋白质在酶和微生物的综合作用下,容易变质,致使鱼类产品品质下降,因此,鱼类产品的新鲜度的预测和判别就很重要^[1-6]。目前,一般采用微生物方法判别多宝鱼肉的新鲜度和剩余存储时间,这种方法尽管结果稳定可靠,但是这是一种破坏性试验,操作人员需要经过专业训练,由于操作过程中要用到化学试剂,造成环境破坏,鱼肉也不能继续食用,较长的检测时间也不能保证实时性^[7]。

近年来,采用高光谱成像技术对鱼类产品的实时和无损检测进行的大量研究,相比较于物理化学法,高光谱成像技术具有快速、无损、成本低和预测准确特点,在很多领域得到广泛应用^[8-11]。采用高光谱成像技术获得的数据,图像上的每一个像素点包含全波长范围的光谱信息,可以具体到单个像素级别的分析,大大提高了分析精细度,可以根据需要,在整个对象的面上寻找提取感兴趣区域,获得对象的光谱数据,而采用近红外光谱分析技术严格意义上说,只能提取对象单个小区域的平均光谱数据,不能同时提取对象整个面上区域的光谱数据,高光谱成像技术具有图像分析技术和光谱分析技术的双重优点,在鱼类产品新鲜度分析领域具有非常大的潜力,其数据矩阵结构原理如图1所示。从图1可知,高光谱数据块是一种三维数据,在每个波长上都有一幅灰度图像与之对应,在图像上每个像素点都包括全谱波长信息。

本文采用高光谱成像技术鉴别多宝鱼的新鲜度,具体研究目的如以下4点:①探寻高光谱成像数据的光谱信息和多宝鱼新鲜度之间的定性关系;②分别采用CARS,SPA和GA挑选多宝鱼的新鲜度特征波长;③比较不同的鉴别多宝鱼新鲜度的模型性能;④确定最优模型用于鉴别多宝鱼的新鲜度。鉴别过程思路如下,首先获取5个类别共160个样本高光谱成像数据,波长范围381~1 023 nm,对原始高光谱成像数据进行黑白板校正;提取样本感兴趣区域光谱,作为样本的光谱数据;根据2:1的比例建立样本建模集

收稿日期:2016-10-01

基金项目:国家自然科学基金项目(61565005);江西省科技支撑项目(20142BDH80021)

作者简介:章海亮(1977—),男,副教授,博士,主要研究方向为高光谱成像技术及其应用。

通讯作者:刘雪梅(1978—),女,副教授,博士,主要研究方向为高光谱成像技术及其应用。

和预测集;基于全谱和特征光谱分别建立鉴别模型;基于分类准确率(correct classification rate, CCR)确定最优鉴别模型。

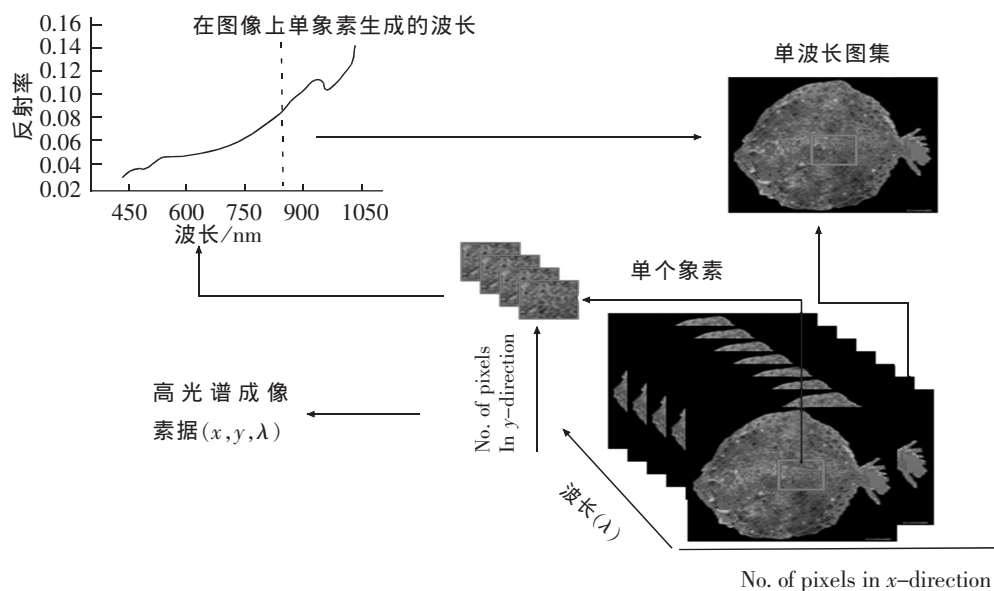


图1 高光谱成像数据矩阵

Fig.1 Hyperspectral imaging data matrix

1 材料与方法

1.1 样本来源及光谱的获取

试验用多宝鱼购于杭州农副产品水产交易中心,样本重量规格在 400~600 g 之间。在水产交易中心现场进行样本前处理,具体过程如下:每次在养殖水池捞出一条鲜活多宝鱼,鱼冰水休克,棒击头部致死,去除鱼内脏和头尾,在冰水中洗干净。每条多宝鱼均匀切割成 4 份,作为 4 个样本。采用塑料袋放置每个样本,将塑料袋放置在盛满新鲜碎冰的干净白色泡沫保温箱中,保温箱底部开有滴水孔,运输过程中部分碎冰化成水从滴水孔中流走,运输专车在 0.5~1 h 内将样本快速运送至实验室。160 样本分 5 类,分别用 1~5 表示 5 个类别,详见表 1 所示。

表 1 建模集和预测集统计信息

Tab.1 Statistical information of calibration and prediction sets

数据集	1 代表低温存贮 20 d 冷冻解冻 1 次	2 代表低温存贮 20 d 冷冻解冻 2 次	3 代表低温存贮 130 d 冷冻解冻 1 次	4 代表低温存贮 130 d 冷冻解冻 2 次	5 代表新鲜 样本	总数
建模集	20	20	20	20	30	110
预测集	10	10	10	10	10	50
总数	30	30	30	30	40	160

注:低温表示 $-20\text{ }^{\circ}\text{C}$,新鲜代表未冷藏未冷冻-解冻 0 次,1 表示在 $0\sim 4\text{ }^{\circ}\text{C}$ 冰箱冷藏环境过夜解冻,其它类似,4 表示过夜解冻,放回低温冷柜在冷冻 3 d 后再过夜解冻。

1.2 数据采集系统

本试验的鱼样本数据采集系统包括高光谱摄像机(N10E, Specim, Finland), 150 W 卤素灯线光源 2 个 (Oriol Instruments, Irvine, Cal.), 铁皮暗箱隔绝室内灯光影响, 步进导轨平台由步进电机(IRCP0076, Isuzu

Optics Corp, Taiwan, China) 驱动, 近红外光谱仪, 可接收的光谱波长范围为 381~1 023 nm, CCD 成像相机, 控制装置和计算机。

1.3 高光谱数据采集

为了获取到鱼样本不变形、不失真和清晰的高光谱成像数据, 需要调节 CCD 相机的曝光时间和步进导轨平台移动速度, 本研究中的步进导轨平台移动速度为 3.2 mm/s, 曝光时间设置为 0.08 s, 近红外光谱仪至样本间距离为 45 cm。采集数据时用纸巾吸去鱼样本表面的多余水分, 将样本置于步进导轨平台上面, 垂直于近红外光谱仪移动, 获取三维鱼样本高光谱成像数据块。分析软件为 Matlab 2010, Unscrambler 10.1 及 Origin8.5 软件平台。为了减少 CCD 相机暗电流及各波长下光源强度分布不均匀的影响, 对原始数据分析前需要对原始高光谱成像数据()进行反射率校正。

1.4 特征波长选择算法

为了简化模型, 提高模型运行效率和精度, 本研究分别采用 MCVE、SPA 和 RF 算法选择特征波长, 关于 MCVE^[12-14]、SPA^[15-16]和 RF^[17-20]原理详见文献。

2 试验结果与分析

2.1 鱼样本的光谱

通过鱼样本高光谱图像上感兴趣区域提取光谱, 感兴趣区域大小选择在 80×80 左右的矩形区域内, 得到的光谱波长范围为 381~1 023 nm, 共计 512 个波长, 由于在 429 nm 波长之前的光谱, 存在部分噪声且有用信息较少, 本研究取 429~1 023 nm 波长范围作为原始波长范围, 图 3 中每一条光谱为样本感兴趣区平均光谱, 从光谱曲线可知, 在整个波长范围内, 最大反射率不高于 0.25, 从光谱曲线还可以看出, 近红外区域的反射率要高于可见光区域的反射率, 在 960 nm 附近, 近红外光谱存在较大的吸收度, 这是由于鱼样本的 C-H 功能键在此区域存在吸收峰。

2.2 Random frog 算法选取特征波长

随机青蛙算法(random frog, RF)是一种非常有用的特征波长提取算法, 在每次建模运算时, 该算法会评价参与建模变量的被选频率值, 被选频率高的变量组合, 用于替代原始光谱用于进一步的建模分析。本文随机青蛙返回结果如图 3 所示, 图 3 中横虚线分别表示取被选频率大于横线对应频率的特征波长用于建模, 或者说取某条横线之上的变量用于建模, 本研究采用随机青蛙算法获取到的变量数量为 49 个。表 2 是采用随机青蛙选择的变量具体波长点。

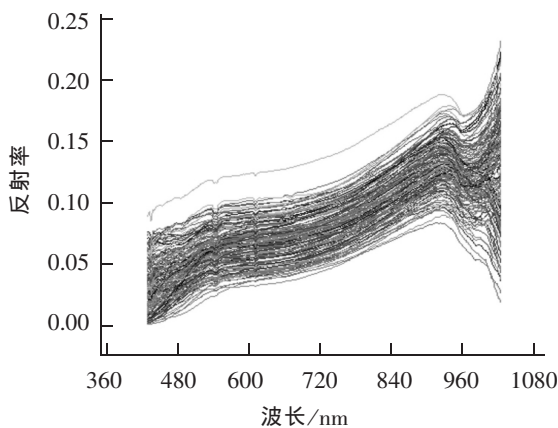


图 2 160 个鱼样本光谱
Fig.2 Spectra of 160 fish samples

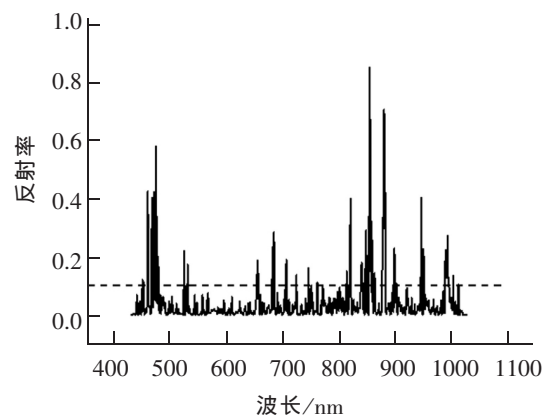


图 3 Random frog 变量被选频率
Fig.3 Frequency of variable selections by RF

表2 Random frog 选择特征波长
Tab.2 Effective wavelengths selected by RF

算法	数量	选择波长 /nm
随机青蛙	49	448,451,458,465,466,469,470,472,474,476,522,529,652,654,679,682,703,721,743,748,758,811,816,818,837,838,843,845,851,852,855,860,874,876,877,878,895,899,943,947,948,985,987,988,989,991,992,1 000,1 009

2.3 MCUVE 关键变量提取

图4表示基于MCUVE算法挑选特征波长结果图。图4表示RMSEC和RMSEP随建模因子数量变化趋势, RMSEC值随建模变量数量变化趋势和每个变量稳定性分布。从图4(a)可以看出,采用8个建模因子数量时,得到的RMSEC和RMSEP值最小,后面随着建模因子数量的增加,模型复杂度也随之增加,降低了模型精度;图4(b)中RMSECV值起初呈急剧减小趋势,表明采样过程中,无信息变量被不断消除,然后RMSECV值变化不明显,表明变量数量变化不明显,最后RMSECV值逐渐增大,表明一些关键变量被消除;图4(c)中,两天虚线之间的波长是要被剔除的,不能参与建模分析,阈值1.57的确定是结合图4(b)和图4(c)决定,由图4(b)可知,保留变量数量为90个,在由图4(c)可知,阈值在1.57处,虚线之外的波长变量数量正好为90个。在MCVE算法中,一些无用变量被消除,一些关键变量被保留,本研究中,90个关键变量被提取出来用于替代原始光谱进行进一步的分析建模。

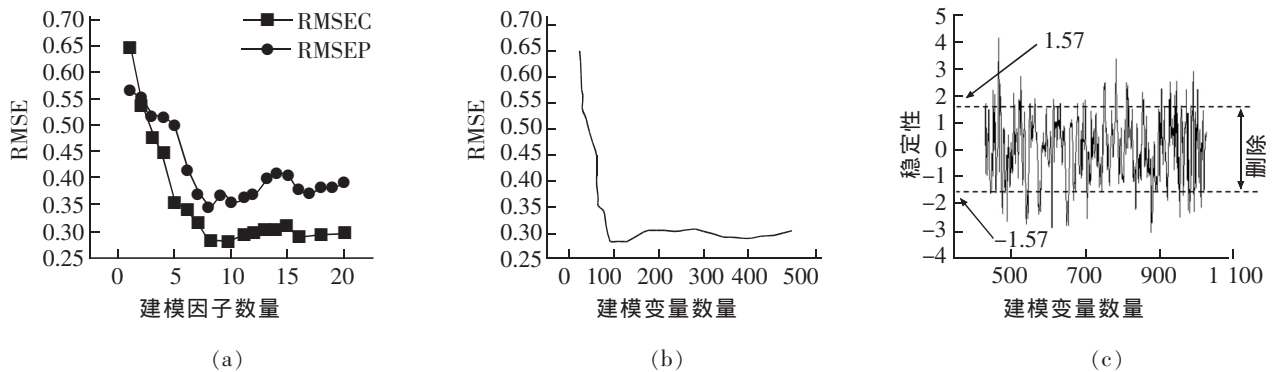


图4 MCUVE 关键变量提取结果图

Fig.4 Key variables selection results by MCVE

2.4 连续投影算法选取特征波长

原始波长数由470个经过遗传算法选择后变为31个波长,采用连续投影算法后,建模变量大大减少,极大的简化了模型。说明一点,图5是SPA算法得到的特征波长在全谱波长上的具体位置信息,SPA算法选择的特征波长详见表3。图6表示模型的RMSEC值变化趋势,随着SPA算法选择的特征变量数量的增加, RMSEC的值起初急剧变小,说明选择的31个特征波长为最优特征波长,而第31个波长点以后的波长对模型的RMSEC值影响不大。

表3 SPA 选择特征波长
Tab.3 Effective wavelengths selected by SPA

算法	数量	选择波长 /nm
连续投影算法	31	431,433,434,435,436,437,440,445,451,453,458,469,476,504,513,531,610,661,765,852,907,965,983,989,992,997,1 001,1 007,1 009,1 014,1 019

2.5 基于光谱建立 LS-SVM 类别预测模型

基于 LS-SVM 的高光谱成像鱼不同冷冻时间以及冷冻解冻次数类别的鉴别模型。LS-SVM 模型在很多领域都得到了广泛应用,可以处理线性和非线性的问题,参数 γ 用于减小模型训练错误率和简化模型,参数 σ^2 用于定义非线性转换,从一个向量空间向高维空间转换^[21-22]。本研究采用的 LS-SVM 分类模型基于 LS-SVM 工具箱。

选取 429~1 023 nm 的波段作为样本光谱信息分析基础,把选取的 470 个波段输入建立 LS-SVM 类别预测模型,模型的预测集识别率为 96%,为了进一步降低模型的复杂度,对原始光谱采用 MCVE 算法,连续投影算法和 RF 算法分别得到 90,31 和 49 个波段变量,且把这 90,31 和 49 个波段变量分别作为 LS-SVM 模型的输入,如表 4 所示。

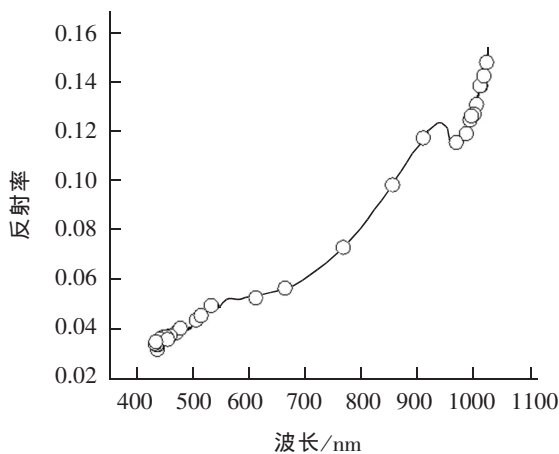


图 5 连续投影算法选择特征波长
Fig.5 SPA selected variables

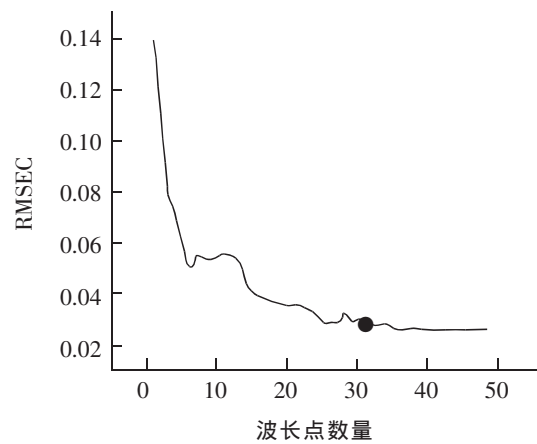


图 6 连续投影算法选择特征波长 RMSEP 变化趋势
Fig.6 Trend of RMSEP based on characteristic wavelengths selected by SPA

表 4 基于不同的波长建立鉴别模型的分类准确率

Tab.4 Correct classification rate of different models based on different wavelengths selection methods

序号	分类模型	波长数量	建模集			预测集		
			样本	失败个数	CCR*/%	样本	失败个数	CCR*/%
1	LS-SVM	470	110	0	100	50	2	96
3	MCCV-LS-SVM	90	110	0	100	50	2	98
5	RF-LS-SVM	66	110	0	100	50	7	98
7	SPA-LS-SVM	31	110	0	100	50	1	98

注:CCR*: Correct classification rate 分类准确率

从表 3 中可以看到,基于全谱的 LS-SVM 模型在预测时,基于光谱模型识别率分别为 96%,当把 MCVE 提取的 90 个参数变量作为 LS-SVM 模型输入时准确率为 98%,把 SPA 选择的特征变量作为 LS-SVM 模型的输入,识别率分别达到 98%,当把 RF 选择的光谱变量作为 LS-SVM 模型的输入时,模型精度为 88%,不难看出,为了简化模型和选择精度最高模型,本文把 SPA 选择的 31 光谱变量作为最优特征变量,把 LS-SVM 分类模型作为最优建模模型。

3 结论

本文基于高光谱成像技术对鱼不同冷冻时间以及冷冻解冻次数类别样本共5个类别进行分类识别。首先,获取鱼样本 ROI 光谱曲线,分别采用 MCVE,SPA 和 RF 算法提取特征波长,将3种算法分别得到90,31和49个特征变量作为模型 LS-SVM 模型的输入建立分类模型,基于 SPA-LS-SVM 和 MCVE-LS-SVM 模型预测集识别率都达到了98%,而采用 RF-LS-SVM 模型预测集识别率都只是达到了88%。结果表明,SPA-LS-SVM 和 MCVE-LS-SVM 作为分类模型优于 RF-LS-SVM 模型,SPA 和 CARS 选择的特征波长,不但可以简化模型,还可以提高模型的预测精度,基于高光谱成像技术可以用于鱼新鲜度(鱼不同冷冻时间以及冻融次数)鉴别。

参考文献:

- [1] 刘晓华,马俐珍,郭耀华,等. 不同保鲜剂对4℃冷藏鲶鱼鱼糜保鲜效果的影响[J]. 食品科学,2014(24):316-320.
- [2] RZEPKA M, OZOGUL F, SUROWKA K, et al. Freshness and quality attributes of cold stored Atlantic bonitogravad[J]. International Journal of Food Science and Technology, 2013, 48(6):1318-1326.
- [3] KIMIYA T, SIVERTSEN A H, HEIA K. VIS/NIR spectroscopy for non-destructive freshness assessment of Atlantic salmon filets[J]. Journal of Food Engineering, 2013, 116(3):758-764.
- [4] LIU D, ZENG X A, SUN D W. NIR spectroscopy and imaging techniques for evaluation of fish quality-A review[J]. Applied Spectroscopy Reviews, 2013, 48(8):609-628.
- [5] CHENG J, SUN D. Rapid Quantification analysis and visualization of escherichia coli loads in grass carp fish flesh by hyperspectral imaging method[J]. Food and Bioprocess Technology, 2015, 8(5):951-959.
- [6] CHENG J, SUN D, PU H. Combining the genetic algorithm and successive projection algorithm for the selection of feature wavelengths to evaluate exudative characteristics in frozen - thawed fish muscle[J]. Food Chemistry, 2016, 197:855-863.
- [7] 章海亮,朱逢乐,刘雪梅,等. 基于图像光谱信息融合的鱼不同冻藏时间及冻融次数鉴别[J]. 农业工程学报,2014(6):272-278.
- [8] HE H, SUN D. Hyperspectral imaging technology for rapid detection of various microbial contaminants in agricultural and food products[J]. Trends in Food Science & Technology, 2015, 46(1):99-109.
- [9] CHENG J, SUN D. Hyperspectral imaging as an effective tool for quality analysis and control of fish and other seafoods: Current research and potential applications[J]. Trends in Food Science & Technology, 2014, 37(2):78-91.
- [10] KHOJASTEHNZHAND M, KHOSHTAGHAZA M H, MOJARADI B, et al. Comparison of visible-near infrared and short wave infrared hyperspectral imaging for the evaluation of rainbow trout freshness[J]. Food Research International, 2014, 56:25-34.
- [11] ZHU F, ZHANG D, HE Y, et al. Application of visible and near infrared hyperspectral imaging to differentiate between fresh and frozen-thawed fish fillets[J]. Food and Bioprocess Technology, 2013, 6(10):2931-2937.
- [12] ANTANASIJEVI D, POCAJT V, PERI-GRUJI A, et al. Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis[J]. Journal of Hydrology, 2014, 519:1895-1907.
- [13] HUANG Z, SHA S, RONG Z, et al. Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed[J]. Industrial Crops and Products, 2013, 43:654-660.
- [14] HAN Q, WU H, CAI C, et al. An ensemble of Monte Carlo uninformative variable elimination for wavelength selection[J]. Analytica Chimica Acta, 2008, 612(2):121-125.
- [15] LIU K, CHEN X, LI L, et al. A consensus successive projections algorithm-multiple linear regression method for analyzing near infrared spectra[J]. Analytica Chimica Acta, 2015, 858:16-23.
- [16] GUAN X C, CHEN X J, JIANG J. Integration of modified uninformative variable elimination and successive projections algorithm for determination harvest time of laver by using visible and near infrared spectra[J]. African Journal of Agricultural Research, 2011, 6(27):5987-5991.

- [17] LI X,SUN C,LUO L,et al. Determination of tea polyphenols content by infrared spectroscopy coupled with iPLS and random frog techniques[J]. Computers and Electronics in Agriculture,2015,112:28-35.
- [18] HU M,DONG Q,LIU B,et al. Estimating blueberry mechanical properties based on random frog selected hyperspectral data[J]. Postharvest Biology and Technology,2015,106:1-10.
- [19] YUN Y,LI H,E WOOD L R,et al. An efficient method of wavelength interval selection based on random frog for multivariate spectral calibration[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,2013,111:31-36.
- [20] LI H,XU Q,LIANG Y. Random frog:An efficient reversible jump Markov Chain Monte Carlo-link approach for variable selection with applications to gene selection and disease classification[J]. Analytica Chimica Acta,2012,740:20-26.
- [21] LIU X,LIU J. Measurement of soil properties using visible and short wave-near infrared spectroscopy and multivariate calibration[J]. Measurement,2013,46(10):3808-3814.
- [22] SHAO Y N,ZHAO C J,BAO Y D,et al. Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy[J]. Food and Bioprocess Technology,2012,5(1):100-107.

Classification of Fish Freshness Based on LS-SVM and Hyperspectra Imaging Technology

Zhang Hailiang, Ye Qing, Luo Wei, Liu Xuemei

(School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: This study investigated the feasibility of using near infrared hyperspectral imaging system (NIR-HIS) technique for non-destructive identification of fresh and frozen-thawed fish fillets. Hyperspectral images of freshness, storage time, and frozen-thawed times of fillets for turbot flesh were obtained in the spectral region of 381~1 023 nm. Reflectance values were extracted from each region of interest (ROI) of each sample. Monte Carlo free information variable elimination (MCVE) algorithm, successive projections algorithm (SPA) and random frog (RF) were carried out to identify the most significant wavelengths. Based on the ninety, thirty-one and forty-nine wavelengths suggested by MCVE, SPA and RF, respectively, two classified models (least squares-support vector machine, LS-SVM and SIMCA) were established. Among the established models, SPA-LS-SVM model performed well with the highest classification rate (100%) in calibration and 98% in prediction sets. SPA-LS-SVM and MCVE-LS-SVM models obtained better results 98% of classification rate in prediction set with thirty-one and ninety effective wavelengths respectively. The RF-LS-SVM model obtained poor results with 88% of classification rate in prediction set. The results showed that NIR-HIS technique can be used to identify the varieties of fresh and frozen-thawed fish fillets rapidly and non-destructively, and SPA was effective wavelengths selection method.

Key words: monte carlo free information variable elimination(MCVE); successive projections algorithm(SPA); random frog(RF); least squares-support vector machine(LS-SVM)

(责任编辑 姜红贵)