基于近红外光谱技术的海参品质快速检测

邹小波*, 薛 瑾, 石吉勇, 徐艺伟, 翟晓东, 胡雪桃, 崔雪平

(江苏大学食品与生物工程学院, 镇江 212013)

摘 要: 目的 应用近红外光谱技术建立海参产地区分和胶原蛋白快速检测的方法。**方法** 总计 43 个海参样品来自大连、福建、连云港、山东 4 个地区。首先采集样品的近红外光谱图, 经过标准正态变量(standard normal variables, SNV)预处理, 利用不同定性判别模型对海参产地进行区分。通过分光光度计法测定海参的胶原蛋白含量, 利用偏最小二乘法(partial least squares, PLS)、区间偏最小二乘法(interval partial least squares, iPLS)、向后区间偏最小二乘法(backwards interval partial least squares, BiPLS)和联合区间偏最小二乘法(synergy interval partial least squares, SiPLS)建立了海参胶原蛋白含量的预测模型。**结果** 产地区分模型中最小二乘支持向量机(least-squares support vector machine regression, LS-SVM)的识别率最高, 校正集识别率为 100%, 预测集识别率为 95.35%; 海参胶原蛋白预测模型中 BiPLS 的预测效果较好, 校正集相关系数 R_c 为 0.9002, 预测集相关系数 R_p 为 0.8517。**结论** 近红外光谱技术可实现对海参的产地区分和胶原蛋白的快速检测。

关键词: 海参; 产地区分; 近红外光谱分析技术; 胶原蛋白; 向后区间偏最小二乘法; 最小二乘支持向量机

Rapid determination of sea cucumber quality based on near-infrared spectroscopy

ZOU Xiao-Bo*, XUE Jin, SHI Ji-Yong, XU Yi-Wei, ZHAI Xiao-Dong, HU Xue-Tao, CUI Xue-Ping

(*School of Food and Biological Engineering*, *Jiangsu University*, *Zhenjiang* 212013, *China*)

ABSTRACT: Objective To establish a method for determination of collagen content of sea cucumber from different geographical origins by near-infrared spectroscopy (NIR). **Methods** Forty-three sea cucumber samples were collected from Dalian, Fujian, Lianyungang and Shandong. Near-infrared spectra of the samples were collected and pretreated by standard normal variables (SNV). Then, the geographical origin of the sea cucumber was determined by using qualitative discriminant models based on the pretreated spectra. The collagen content of sea cucumber was determined with UV spectrophotometry. Partial least squares (PLS), interval partial least squares (iPLS), backwards interval partial least squares (BiPLS) and synergy interval partial least squares (SiPLS) were used to build quantitive prediction models for the collagen content. **Results** The optimal qualitative discriminant model for the geographical origins of sea cucumber was least-squares support vector machine regression (LS-SVM), and its recognition rates of calibration set and prediction set were 100% and 95.35%, respectively. The best quantitive

基金项目: 国家自然科学基金项目(31671844)、国家科技支撑项目(2015BAD17B04)、“十三五”国家重点研发计划项目(2016YFD0401104)、江苏省自然科学基金项目(BK20160506)

Fund: Supported by the National Natural Science Foundation of China (31671844), the National Science and Technology Support Program (2015BAD17B04), the National Key Research and Development Program of China (2016YFD0401104) and the Natural Science Foundation of Jiangsu Province (BK20160506)

***通讯作者:** 邹小波, 教授, 主要研究方向为农产品、食品品质无损检测。E-mail: zou_xiaobo@ujs.edu.cn

***Corresponding author:** ZOU Xiao-Bo, Professor, School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China. E-mail: zou_xiaobo@ujs.edu.cn

prediction model for collagen content was BiPLS, and its calibration coefficient (*R*c) and prediction coefficient (*R*p) were 0.9002 and 0.8517, respectively. **Conclusion** Near-infrared spectroscopy can be used to rapidly determine the geographical origins and the collagen content of sea cucumber.

KEY WORDS: sea cucumber; geographical origins identification; near-infrared spectroscopy; collagen; backwards interval partial least squares; least-squares support vector machine regression

1 引 言

海参(sea cucumber)属棘皮动物门(echinodermata)海参纲(holothuroidea)生物, 是生活在海边至 8000 米的海洋动物, 距今已有六亿多年的历史[1,2], 海参具有很高的食用和药用价值, 自古以来就被视为名贵的滋补品[3]。GB 31602-2015《食品安全国家标准干海参》[4]中将海参蛋白质含量列为重要理化指标。而海参胶原蛋白是海参蛋白的主要活性成分, 含量约达海参蛋白的 70%。其胶原蛋白中所富含的甘氨酸和碱性氨基酸是生血、养血(血红素和珠蛋白的生物合成)和促进钙质吸收的物质基础[5,6], 具有很高的营养价值, 被普遍地用在功能食品和蛋白饮料中[7]。

由于各产地的海参有效成分含量不同, 价格差异较大, 且外观特征差异不明显, 致使市场上常常出现真伪混淆现象, 因此海参产地的快速区分对食品安全管理具有十分重要的意义。传统的海参胶原蛋白的检测方法是用分光光度计进行检测, 化学检测法虽然测量误差小, 但操作复杂, 难以实现现场快速检测, 因此本文提出了胶原蛋白的快速检测方法。

近红外光谱分析技术可有效地实现对海参的真伪、掺假、产地等的定性区分[8,9], 但对海参中化学成分的定量测定研究报道较少。本文首先采用分光光度计法测定海参中胶原蛋白的含量, 再基于近红外光谱结合区间偏最小二乘法(interval partial least squares, iPLS)、向后区间偏最小二乘法(backwards interval partial least squares, BiPLS)和联合区间偏最小二乘法(synergy interval partial least squares, SiPLS)校正模型, 实现对不同产地海参中胶原蛋白含量的定量分析。

2 材料与方法

2.1 仪器与试剂

Antaris Ⅱ型傅里叶变换近红外光谱仪(美国 Thermo Scientific 公司); FW100 型高速粉碎机(北京永光明医疗器械公司)

43 个干海参样品购自海参专卖店, 均为淡干海参, 来自 4 个产地(大连 13 个、福建 10 个、连云港 11 个、山东 9 个)。

柠檬酸钠缓冲溶液(pH 6.0)、氯胺 T 溶液、高氯酸溶液、*L*-羟脯氨酸、盐酸(分析纯, 国药集团化学试剂有限公司)。

2.2 近红外数据采集

将海参用高速粉碎机磨粉处理得海参粉末。获得的干海参粉末样品装入圆柱形石英样品杯中后敦实(样品高度 2 cm)放置在积分球的窗口上采集样品数据。采集数据时使用 OPUS® 6.5 气氛补偿功能, 将环境中的水蒸气的光谱吸收自动扣除。为进一步消除高频噪声信号, 还对所有的样品光谱数据进行了 3 点数据平滑操作。

试验采用 Antaris 傅里叶变换近红外光谱仪扫描海参近红外光谱。扫描波数范围为 4000~10000 cm^{-1}, 扫描次数为 16 次, 波数间隔为 3.856 cm^{-1}(共 1557 个波数点), 以仪器内置背景为参比, 积分球漫反射。本实验中共扫描 43 个海参样本的光谱信息, 每个海参样本分成 3 份, 每份分别扫描 1 次光谱信息并测其对应的胶原蛋白含量。

2.3 海参胶原蛋白含量测定

海参体壁胶原蛋白含量的测定: 称取干海参粉末约 12 mg 于安培瓶中, 加入 2 mL(6 mol/L)盐酸溶液, 用酒精喷灯封口后, 置于 130℃下水解 4 h, 过滤, 转移至试管中, 用 2 mol/L 氢氧化钠调 pH 至中性后, 定容至 10 mL。取上述水解液 1 mL 按照以上标准曲线制作方法进行测定, 在 560 nm 处测吸光度。由 *L*-羟脯氨酸标准曲线得出样品 *L*-羟脯氨酸的浓度, 经换算得胶原蛋白浓度[10]。

标准曲线制作: 准确称取 *L*-羟脯氨酸, 加入 1 mL 盐酸(0.01 mol/L)中溶解, 定容至 100 mL, 4 ℃冰箱储存。吸取 *L*-羟脯氨酸储存液 10 mL 到 100 mL 容量瓶中, 用蒸馏水定容到 100 mL, 浓度为 100 μg/mL(现配)。吸取此标准工作液 3.0、6.0、9.0、12.0、15.0、18.0、21.0 mL, 分别用蒸馏水定容到 100 mL, 浓度分别为 3.0、6.0、9.0、12.0、15.0、18.0、21.0 μg/mL。以蒸馏水为空白, 分别吸取上述标准液 1 mL, 加入 1 mL 柠檬酸缓冲液和 1 mL 氯胺 T 溶液在室温(25 ℃)氧化 20 min 后, 加入高氯酸 1 mL 放置 10 min, 加入对二甲基氨基苯甲醛 1 mL, 65 ℃水浴显色 20 min, 冷却后测 $A_{560\ nm}$。

2.4 近红外光谱获取与预处理

本实验共得到 43 个海参的 129 条光谱信息。在光谱预处理中, 原始光谱通常伴随着高频随机噪声、基线漂移、样本颗粒大小和光散射噪声等信息夹入, 影响建立模型的可靠性, 需采用一定的预处理方法进行消除[11]。本研究采

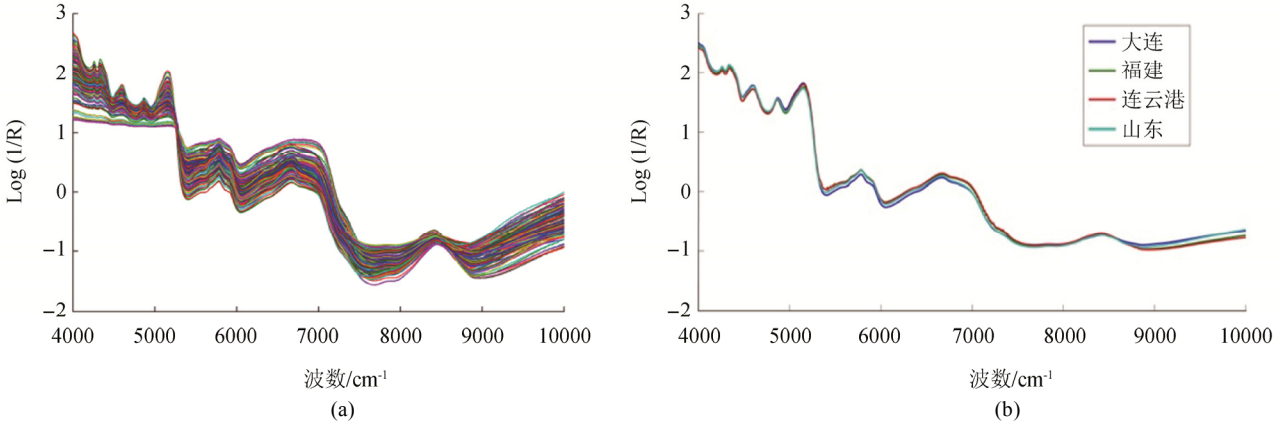

图 1　海参样品 SNV 处理近红外光谱(a)和平均近红外光谱(b)

Fig. 1　Spectrum of sea cucumber after SNV preprocessing (a) and average spectrum of sea cucumber sample (b)

用标准正态变量(standard normal variatetransformation, SNV)[12]对光谱进行有效信息提取, SNV 处理后的光谱曲线如图 1(a)所示。由图 1(a)可知, SNV 预处理减少了样本信号噪声, 提高了信号的分辨度和灵敏度。与原始光谱相比, SNV 预处理可以提取光谱有效信息, 减少误差影响, 提高建模准确度。图 1(b)为 4 个产地海参样品的平均光谱图(*R* 表示近红外光的反射强度)。由图 1(b)可以看出每种产地海参都具有相同的趋势, 而在相同的波长下又有不同的吸收度, 因此可以进行定性识别。

3　结果与分析

3.1　海参胶原蛋白含量

4 种产地海参胶原蛋白含量如表 1, 可得山东的海参胶原蛋白含量最高, 平均含量为 221.08 mg/g, 因为各个产地海参所处海域不同, 其中蛋白质含量差异很大, 以至于质量存在差异。因此也强调了快速、准确地测定胶原蛋白含量对于评价海参质量的重要性。

表 1　4 种产地海参胶原蛋白含量

Table 1　Collagen content of sea cucumber from 4 geographical origins

类别	数量	最小值(mg/g)	最大值(mg/g)	平均值(mg/g)
大连	13	166.76	243.26	198.06±20.533
福建	10	131.28	169.79	150.33±12.328
连云港	11	145.47	196.54	170.19±11.934
山东	9	196.54	254.50	221.08±17.043

3.2　海参产地区分

3.2.1　主成分分析

本实验采用主成分分析(principal component analysis, PCA)对近红外数据进行分析, 可将分散在一组变量上的信息集中到某几个综合指标上, 采用较少的信息集中地对海参进行清晰表征[13]。经主成分分析处理后, 取前 3 个主成分因子得分向量作图, 结果如图 2 所示, 前 3 个主成分的贡献率分别为 71.49%、23.97%、2.42%, 累计贡献率达到 97.88%。从图 2 可看出 4 种产地的海参样品在该得分图中大体可以聚为 4 类, 部分存在交叉, 但还需进一步的模式识别, 将不同产地的海参样品区分。

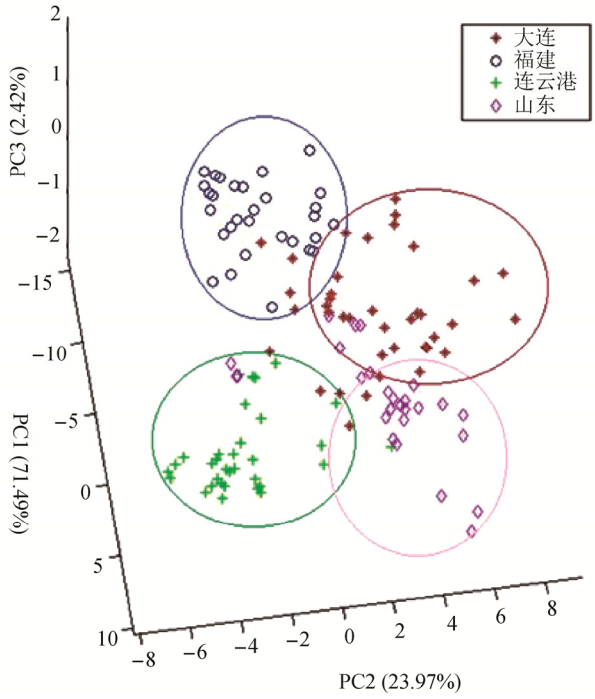

图 2　三维主成分得分图

Fig. 2　Three-dimensional projection of PCA result

3.2.2　基于线性判别分析海参产地的区分

线性判别分析(linear discriminant analysis, LDA)是用于判断样品所属类型的统计分析方法, 可以通过找出数据的特征向量然后将其投影到一个更低维的方向, 使得数据的组与组之间分开, 实现样本的分类[14]。在这项研究中(如图 3), LDA 的校正集和预测集的识别率分别为 95.35%和 88.37%。

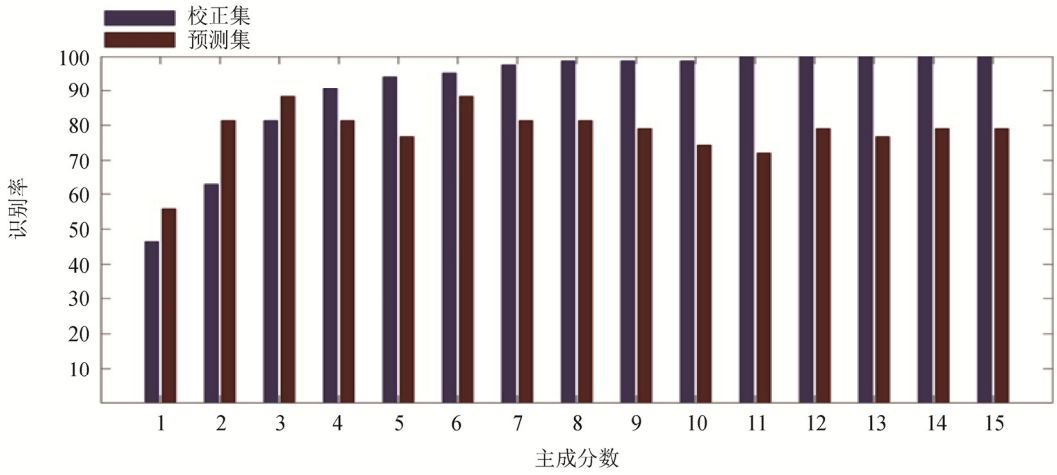

图 3 基于 LDA 的各主成分识别率

Fig. 3 Discrimination rate of each principal component based on LDA

3.2.3 基于 *K* 最近邻法海参产地的区分

K 最近邻法(*K*-nearest neighbors, KNN)主要是靠同类样本在模式空间相互靠近为依据的分类方法[15]。在 KNN 模型识别的过程中, 输入的主成分数对模型的识别结果有较大的影响, 研究选取了 10 个主成分数和 10 个 *K* 值同步对 KNN 模型进行优化。从图 4 可以看出, 当主成分数为 4, 当 *K* 值为 2 时, KNN 模型校正集可取得最佳结果, 识别率为 94.19%; 模型预测集识别率为 81.40%。

3.2.4 基于 BP 神经网络和最小二乘支持向量机海参产地的区分

BP 神经网络(BP-neural network, BP-ANN)是一种反向传递并修正误差的多层映射神经网络, 它是各种神经网络模型中最具有代表意义的一种神经网络模型[16]。此外, BP-ANN 是非线性建模手段, 对于组织复杂的检测对象效果显著, 故本研究采用 BP-ANN 建立区分 4 个产地海参的数学模型, 结果表明, 该神经网络模型的校正集识别率为 100%, 预测集识别率为 75.32%。

最小二乘支持向量机(least squares support vector machine, LS-SVM)是在经典 SVM 的基础上改进的, 除了能够进行线性分类外, 支持向量机可以使用所谓的核技巧, 有效地进行非线性分类[17]。本研究采用最小二乘支持向量机建立了区分 4 个产地海参的数学模型。最小二乘支持向量机在应用之前, 需要优化 γ 和 σ 两参数。结果表明(如图 5), 当主成分为 6 时, 校正集的识别率是 100%, 预测集的识别率是 95.35%, 模型最优, 可见支持向量机对区分不同产地海参有着良好的应用性能。

3.2.5 比较分析结果

通过前述可以发现, 不同的模型对海参的产地区分都存在差异, 从表 2 可以看出, 在 4 种检测模式中, LS-SVM 模型在 PCs=5 时, 校正集识别率是 100%, 预测集识别率是 95.35%, 较 LDA、KNN 和 BP-ANN 都具有优势。原因在于对内部化学成分, 组织结构复杂的海参来说, 基于线性判别方法的 LDA、KNN 不适合对其进行判别。虽然 BP-ANN 的校正集识别率为 100%, 但其预测集识别率为 75.32%, 明显低于 LS-SVM。由于 BP-ANN 遵循经验风险最小化准则(ERM), 经过训练得到的最佳模型可能导致预测结果的降低, 而 LS-SVM 遵循结构风险最小化(SRM)准则构造决策超平面, 使得样本之间的分类间隔最大, 并且能够同时兼顾经验风险和置信范围, 训练得到的模型具有良好的泛化性。故 LS-SVM 法对 4 种产地海参进行区分的效果最好。

表 2 基于 LS-SVM、BP-ANN、KNN 和 LDA 模式识别的识别率

Table 2 Discrimination rates of LS-SVM, BP-ANN, KNN and LDA models

模型	识别结果	
	校正集	预测集
LS-SVM	100%	97.67%
BP-ANN	100%	75.32%
KNN	94.19%	81.40%
LDA	95.35%	88.37%

3.3 海参胶原蛋白的偏最小二乘法(partial least squares, PLS)、iPLS、BiPLS 和 SiPLS 模型建立与评价

以校正集和预测集的相关系数 *R*c 和 *R*p、交互验证均

方根误差 RMSECV 及预测均方根误差 RMSECP 作为评价各种建模方法的有效指标。4 种海参全光谱 PLS 模型及通过 iPLS、BiPLS 和 SiPLS 挑选特征波段后建立的 PLS 模型结果如表 3。结果显示, 通过 BiPLS 挑选特征波段建立的模型效果最好, 胶原蛋白含量测定模型见图 6。海参的全光谱 PLS 和 iPLS 模型结果与联合区间(BiPLS 和 SiPLS)建立的 PLS 模型相比结果较差, 这是因为全光谱波段中的光谱噪声会影响 PLS 模型的精度; iPLS 模型只有一个区间的光谱信息, 反映的色谱信息不足; 而 BiPLS 和 SiPLS 是挑选几个子区间进行建模, 可以去除干扰大的区间, 因此建模效果比 PLS 模型好。

表 3　海参胶原蛋白的 PLS、iPLS、BiPLS 和 SiPLS 模型预测结果
Table 3　Prediction results of collagen content of sea cucumber by PLS, iPLS, BiPLS and SiPLS models

模型	*R*c	RMSECV	*R*p	RMSECP
全光谱 PLS	0.8859	1.48	0.8382	1.67
iPLS	0.8805	1.51	0.7738	1.99
BiPLS	0.9002	1.38	0.8517	1.62
SiPLS*	0.8993	1.39	0.8511	1.63
SiPLS**	0.8884	1.44	0.8513	1.70

注: *代表 3 区间联合; **代表 4 区间联合

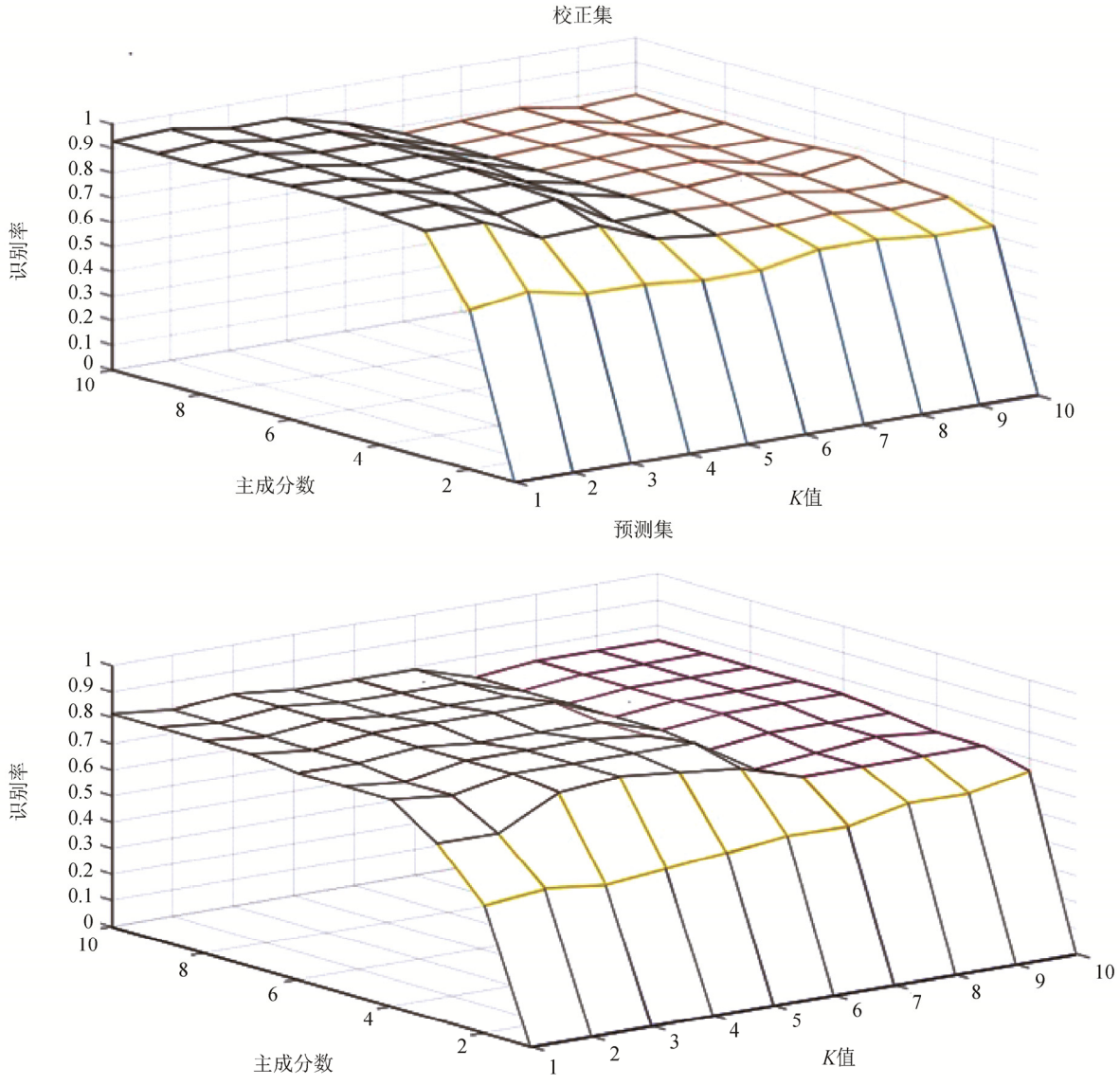

图 4　KNN 模型识别的校正集和预测集
Fig. 4　Discrimination results of calibration set and prediction set based on KNN

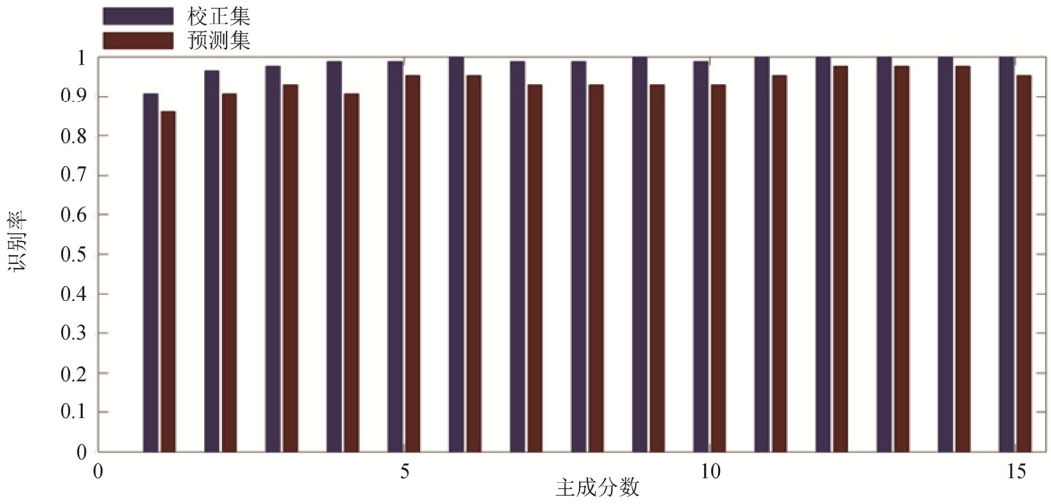

图 5 基于 LS-SVM 的各主成分的识别率

Fig. 5 Discrimination rate of each principal component based on LS-SVM

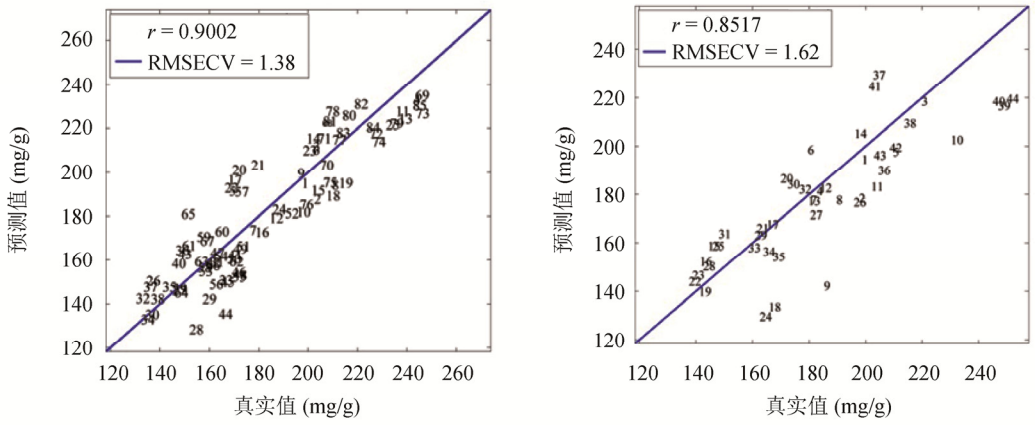

图 6 海参中胶原蛋白近红外光谱含量测定模型图

Fig. 6 Near infrared spectrum of collagen content of sea cucumber model diagram

4 结 论

本研究提出了使用近红外光谱技术对海参产地区分并检测海参中胶原蛋白的方法，组合了线性判别分析(LDA)、*K*最邻近算法(KNN)、最小二乘支持向量机(LS-SVM)模型和误差反向传播神经网络(BP-ANN)进行产地区分研究，与其他模型相比，LS-SVM 的识别率最好，然后通过联合区间 BiPLS 挑选特征波段建立海参光谱与胶原蛋白含量的 PLS 预测模型，对应的校正集相关系数 *R*c 为 0.9002，预测集相关系数 *R*p 为 0.8517，表明近红外光谱分析技术可无损、快速分析海参中的蛋白质。

参考文献

[1] 郭盈莹，丁燕，徐飞飞，等. 海参中主要生物活性成分研究进展[J]. 食品科学, 2014, 35(15): 335−344.
Guo Y, Ding Y, Xu FF, *et al*. Progress in research on main bioactive constituents of sea cu cumber [J]. Food Sci, 2014, 35(15): 335−344.

[2] 姜健，杨宝灵，邰阳. 海参资源及其生物活性物质的研究[J]. 生物技术通讯, 2004, 15(5): 537−540.
Jiang J, Yang BL, Tai Y. Studies on resources and bioactive substances of sea cucumber [J]. Lett Biotechnol, 2004, 15(5): 537−540.

[3] 张梅秀，王锡昌，刘源. 海参生物活性研究进展[J]. 天然产物研究与开发, 2012, 24(8): 1151−1159.
Zang MX, Wang XC, Liu Y.Advances on the research of sea cucumber biological activities [J]. Nat Prod Res Dev, 2012, 24(8):1151−1159.

[4] GB 31602-2015 食品安全国家标准 干海参使用标准[S].
GB 31602-2015 Food safety national standard-Standard of dry sea cucumber [S].

[5] 李翠翠. 海参体壁胶原提取及加工过程组织形态学研究 [D]. 大连: 大连工业大学, 2008.
Li CC. Study on extractionan and structural morphous in the processing of collagen from the body wall of sea cucumber [D]. Dalian: Dalian

Polytechnic University, 2008.

[6] 李志超. 海参质量评定方法与加工工艺对品质的影响研究[D]. 大连: 大连海洋大学, 2014.
Li ZC. Studies on quality assessment methods and processing parameters optimization for sea cucumber [D]. Dalian: Dalian OceanUniversity, 2014

[7] 罗臻, 黄静敏, 陈血建, 等. 海参胶原蛋白的提取及其理化性质[J]. 安徽农业科学, 2015, (3): 145–146.
Luo Z, Huang JM, Chen XJ. *et al*. Extraction and physico-chemical properties of pepsin-solubilized collagen from sea cucumber [J]. J Anhui Agri Sci, 2015, (3): 145–146.

[8] Liu X, Xue C, Wang Y, *et al*. The classification of sea cucumber (*Apostichopus japonicus*) according to region of origin using multi-element analysis and pattern recognition techniques [J]. Food Control, 2012, 23(2): 522–527.

[9] 陶琳, 武中臣, 张鹏彦, 等. 近红外光谱法快速鉴定干海参产地[J]. 农业工程学报, 2011, 27(5): 364–366.
Tao L, Wu ZC, Zhang PY, *et al*. Rapid identification of dried sea cucumber products from different geographical areas by near-infrared spectroscopy [J]. Trans Chin Soc Agric Eng, 2011, 27(5): 364–366.

[10] 肖枫, 曾名勇, 董士远, 等. 海参胶原蛋白的研究进展[J]. 水产科学, 2005, 24(6): 39–41.
Xiao F, Zeng MY, Dong SY, *et al*. Research progress in sea cucumber collagen [J]. Fish Sci, 2005, 24(6): 39–41.

[11] 陈全胜, 郭志明, 赵杰文, 等. 基于净分析物预处理算法的绿茶中儿茶素的近红外光谱定量分析[J]. 红外与毫米波学报, 2009, 28(5): 357–361.
Chen QS, Guo ZM, Zhao JW, *et al*. Quantitative analysis of the catechins conients in green tea with near infrared spectroscopy and net analyte preprocessing algorithm [J]. J Infrar Millim Wave, 2009, 28(5): 357–361.

[12] Chen Q, Jiang P, Zhao J. Measurement of total flavone content in snow lotus (*Saussurea involucrate*) using near infrared spectroscopy combined with interval PLS and genetic algorithm [J]. Spectrochim Acta A Mol Biomol Spectrosc, 2010, 76(1): 50–55.

[13] 赵杰文, 蒋培, 陈全胜. 雪莲花产地鉴别的近红外光谱分析方法[J]. 农业机械学报, 2010, 41(8): 111–114.
Zhao JW, Jiang P, Chen QS. Discrimination of snow lotus from different geographical origins by near infrared spectroscopy [J]. Trans Chin Soc Agric Mach, 2010, 41(8): 111–114.

[14] 王晓慧. 线性判别分析与主成分分析及其相关研究评述[J]. 中山大学研究生学刊: 自然科学、医学版, 2007, (4): 50–61.
Wang XH. A summary of LDA, PCA and relative work [J]. J Grad Sun Yat-Sen Univ (Nat Sci Med Ed), 2007, (4): 50–61.

[15] Zou XB, Zhao J, Povey MJ, *et al*. Variables selection methods in near-infrared spectroscopy [J]. Anal Chim, 2010, 667(1–2): 14–32.

[16] Prasad R, Pandey A, Singh KP, *et al*. Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: A comparison of different transfer functions [J]. Adv Space Res, 2012, 50(3): 363–370.

[17] He HJ, Sun DW, Wu D. Rapid and real-time prediction of lactic acid bacteria (LAB) in farmed salmon flesh using near-infrared (NIR) hyperspectral imaging combined with chemometric analysis [J]. Food Res Inter, 2014, 62(8): 476–483.

(责任编辑: 杨翠娜)

作者简介

邹小波, 博士生导师, 教授, 主要研究方向为农产品、食品品质无损检测。
E-mail: zou_xiaobo@ujs.edu.cn