

基于近红外光谱的核桃仁蛋白质含量检测分析

马文强^{1,2} 张漫¹ 李忠新² 杨莉玲²

(1. 中国农业大学现代精细农业系统集成研究教育部重点实验室, 北京 100083;

2. 新疆农业科学院农业机械化研究所, 乌鲁木齐 830091)

摘要: 为了实现核桃仁蛋白质的快速无损检测, 采用近红外光谱技术, 建立了核桃仁蛋白质含量预测模型, 并对近红外光谱波段筛选方法进行了研究。首先针对3种不同粒度核桃仁样本, 采集了1 040~2 560 nm范围全波段信息, 采用多元散射校正法和标准正态化方法对原始光谱进行了预处理。然后, 采用间隔偏最小二乘算法筛选了光谱特征波段, 并建立了全波段和特征波段下核桃仁蛋白质含量偏最小二乘算法预测模型。通过对不同粒度核桃仁样本近红外光谱分析表明, 核桃仁粒度大小对核桃仁蛋白质含量预测效果并无显著影响。采用间隔偏最小二乘算法的波段筛选, 核桃整仁样本验证集的均方根误差和相关系数分别为0.021和0.913, 表明该方法能够优化模型质量并降低模型复杂度。

关键词: 核桃仁; 蛋白质含量; 近红外光谱技术; 光谱筛选; 间隔偏最小二乘算法

中图分类号: O657.33; S664.1 **文献标识码:** A **文章编号:** 1000-1298(2017)S0-0407-05

Detection and Analysis of Walnut Protein Content Based on Near Infrared Spectroscopy

MA Wenqiang^{1,2} ZHANG Man¹ LI Zhongxin² YANG Liling²

(1. Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China

2. Agricultural Mechanization Institute, Xinjiang Academy of Agricultural Sciences, Urumqi 830091, China)

Abstract: In order to establish a rapid and nondestructive method for the detection of protein in walnut kernel, a prediction model for walnut protein content was established based on near infrared spectroscopy. Near infrared band selection method was also studied for optimizing protein content predicting model. Firstly, near infrared spectroscopy information of three different sized walnut samples in the full range of 1 040 ~ 2 560 nm was collected, and the pretreatment was completed by multiple scattering correction and standard normal method. Then the spectral characteristic bands were selected by interval partial least square algorithm, and the partial least squares prediction models were established at full bands ranging from 1 040 nm to 2 560 nm and selected feature bands, respectively. The spectral analysis results showed that the walnut size has no significant effect on the walnut protein content prediction using near infrared spectrum method. The root mean square error and the correlation coefficients of the optimized model were 0.021 and 0.913 in the whole walnut sample validation set, which indicated that the application of interval partial least algorithm could optimize the model quality and reduce the complexity.

Key words: walnut kernel; protein content; near infrared spectroscopy; spectral screening; interval partial least algorithm

收稿日期: 2017-07-15 修回日期: 2017-11-18

基金项目: 新疆自治区重点研发项目(2017B01003-4)

作者简介: 马文强(1986—), 男, 博士生, 新疆农业科学院助理研究员, 主要从事农产品智能化检测技术研究, E-mail: mwq4530@163.com

通信作者: 张漫(1975—), 女, 教授, 博士生导师, 主要从事智能检测技术研究, E-mail: cauzm@cau.edu.cn

引言

核桃仁富含蛋白质、不饱和脂肪酸和多种微量元素,是一种老幼皆宜的坚果食品和重要的木本油料作物,我国核桃年产200万t以上,产量居世界首位^[1]。蛋白质含量是判别核桃仁品质的重要指标,传统的检测方法为凯氏定氮法,对核桃仁具有破坏性,检测时间也较长。因此开发一种快速、无损、可在线检测的核桃仁蛋白质含量检测方法是核桃精加工产业的迫切需求。

近红外光谱技术(Near infrared spectroscopy, NIRS)可以对样品进行定性或定量分析,具有分析速度快、结果稳定、对物料无破坏性等特点,适合于农产品品质快速无损检测,目前已在蔬菜、水果、肉类、谷物等多种农产品中得到了广泛的应用^[2-7]。迟茜等^[8]在900~1700 nm近红外光谱范围内,实现了猕猴桃早期隐性损伤的识别,平均正确识别率达到了97.9%。EVA等^[9]利用近红外反射光谱对杏仁内氰糖苷进行无损检测,利用主成分分析法和偏最小二乘法进行建模,并实现了苦杏仁与甜杏仁的分类,分类准确率大于95%。韩东海等^[10]采用近红外光谱技术建立了西瓜可溶性固形物含量的快速无损检测模型,瓜顶部位和赤道部位模型相关系数分别为0.953和0.924, RMSEP分别为0.510和0.766。匡静云等^[11]对原料乳近红外光谱进行了特征波段筛选建模,针对原料乳中蛋白含量与脂肪含量,模型相关系数和均方根误差分别为0.9883、0.9878和1.83%、1.85%。在坚果、谷物等检测方面,已有针对松子、板栗、玉米、大豆、花生等^[12-17]籽粒的无损检测研究,而针对核桃仁蛋白质含量的近红外光谱检测研究^[18-19],都采用对核桃仁进行粉碎处理后再采集光谱的方法。

本文拟在1040~2560 nm范围内采集3种不同粒度的核桃仁光谱信息,利用多元散射校正和标准正态化方法进行光谱预处理,通过偏最小二乘法(PLS)建立核桃仁蛋白质含量预测模型,以分析不同粒度大小对核桃仁蛋白质含量预测的影响;并进一步针对核桃整仁,采用间隔偏最小二乘算法进行光谱特征波段筛选,以实现优化模型质量,降低模型复杂度,以期为实现核桃仁蛋白质快速无损在线检测提供理论参考。

1 材料与方法

1.1 实验样品

实验用的核桃选用新疆“温185”核桃,含水率约为8%,采用手工破壳的方式进行破壳取仁,并保

持核桃仁的完整性。为了研究核桃仁在不同粒度下的近红外光谱特征,将剥好的核桃仁样品分成3组,每组0.5kg,分别制备成整仁(完整度在1/2以上)、碎仁(完整度在1/16以下)和粉碎仁(使用粉碎机粉碎至40目以下)。

1.2 仪器设备与光谱信息测量

实验仪器使用德国BRUKER公司的MPA傅里叶变换漫反射近红外光谱仪(图1)。仪器光源为卤钨灯,配有用于固体样品近红外漫反射光谱信息采集的样品池。光谱测量方式采用最大积分球漫反射方式,样品池为内径 $\phi 51$ mm旋转样品池,测量光谱范围为1040~2560 nm,分辨率为2.1 nm,扫描次数为64次。

测量步骤为:①打开光谱仪预热30 min以确保仪器的测量稳定性,并进行测量参数的设置。②将待测量的核桃仁样品放入样品池进行近红外光谱测量,测量的同时,样品随样品池自动匀速旋转360°。③测量后将样品按不同粒度和测量顺序进行编号处理,以进一步用于蛋白质含量测定。④每次测量后,对样品池进行超声波清洗。每种核桃仁样品取样测量34次,共获取102组光谱数据。

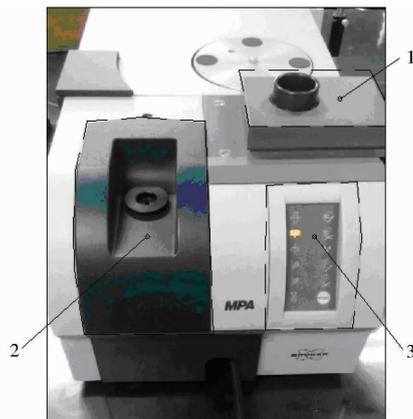


图1 BRUKER MPA漫反射近红外光谱仪
Fig.1 BRUKER MPA diffuse reflectance near infrared spectrometer

1.漫反射光谱采集样品池 2.透射光谱采集样品池 3.工作状态指示面板

1.3 蛋白质含量测定

蛋白质含量测定按照食品安全国家标准GB5009.5—2010中的要求,采用标准凯氏定氮法进行测量,测量仪器为FOSS全自动凯氏定氮仪。测量时,每份样品从经过光谱测量并粉碎处理的待测量样品中精确称取0.3 g并进行硝化处理,每一批次测量取待测核桃样品17个,样品空白3个,并按仪器说明书要求进行操作。测量结束后,使用每批次空白样品的测量均值进行误差校正。

2 结果与讨论

2.1 核桃蛋白质含量测定结果分析

表 1 为 3 种不同粒度的核桃仁样品蛋白质含量的测量统计数据,所选样品的蛋白质含量分布在 17.6% ~ 23.9% 之间,均值为 21.05%,标准差为 1.46%。表明所选样品的蛋白质含量具有较好的数据差异性,可以满足采用近红外光谱数据进行建模的要求。

表 1 核桃仁蛋白质含量测定数据

Tab. 1 Protein content measurement data of

walnut kernel

样品类型	样品数	蛋白质含量/%			
		平均值	最大值	最小值	标准差
整仁	34	20.74	23.08	17.85	1.26
碎仁	34	21.36	23.59	19.36	1.33
粉碎仁	34	21.06	23.90	17.60	1.76
全部样品	102	21.05	23.90	17.60	1.46

2.2 光谱预处理与特征分析

图 2a 为采集的核桃仁近红外光谱图。由图 2a 可以看出随着核桃仁样本粒度的减小,其相对吸光度增大,但总体特征基本一致。

为了进一步消除近红外光谱信息中的噪声,降低样本粒度大小对光谱数据的影响,采用多元散射校正和标准正态化的组合方法对原始光谱信息进行预处理。多元散射校正 (Multivariate scatter correction, MSE) 可有效地消除散射影响,提高光谱信息的信噪比;标准正态化处理 (Standard normal variate transformation, SNV) 可以提高不同光谱信息的一致性,增强光谱间差异,并抑制随机噪声。图 2b 为预处理后的核桃仁近红外光谱图,可以看出预处理后的光谱信息的一致性和特征峰 (谷) 都得到了增强。

2.3 全光谱偏最小二乘法 (PLS) 建模

预处理后的光谱信息通过随机抽样法划分为建模集和验证集,其中每种样品建模集样本数为 26 个,验证集样本数为 8 个。采用偏最小二乘算法 (Partial least squares, PLS) 在采集的全光谱范围内对 3 种样品的蛋白质含量建立预测模型,并采用均方根误差 (RMSE) 和相关系数 (R) 对模型的准确性和相关性进行评价。建模结果如表 2 所示。由表 2 可以看出,建立的预测模型对 3 种样品的蛋白质含量的预测都取得了较高的精度,说明样品粒度大小对其蛋白质含量的预测影响较小,采用核桃整仁进行蛋白质含量近红外光谱预测具有较好的可行性;同时预测模型相关系数 (R) 并不理想,说明在采集

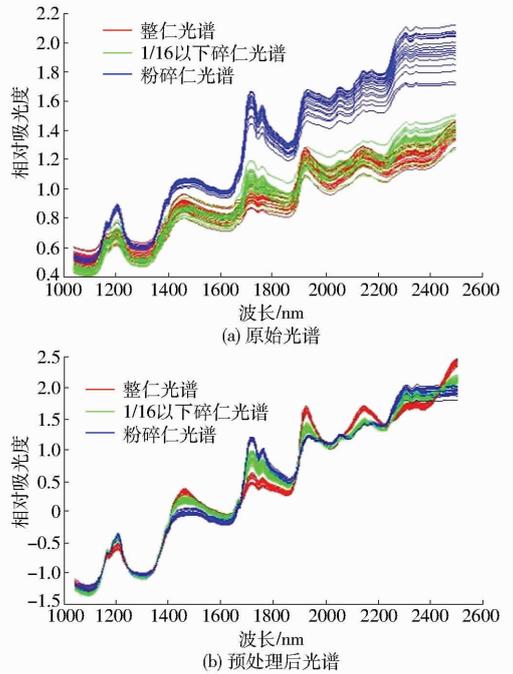


图 2 核桃仁近红外光谱图

Fig. 2 Walnut kernel near infrared spectrum

表 2 全光谱核桃仁蛋白质 PLS 建模统计

Tab. 2 PLS modeling statistics of walnut kernel protein in full spectrum

不同粒度 样品	建模集		验证集	
	RMSE	R	RMSE	R
整仁	0.014	0.907	0.025	0.467
碎仁	0.033	0.936	0.039	0.576
粉碎仁	0.069	0.879	0.043	0.471

的全光谱范围内存在较多与核桃蛋白质含量相关度较低的光谱信息,需要进一步对光谱信息的特征波段进行筛选。

2.4 间隔偏最小二乘法 (IPLS) 光谱特征筛选

对全波段光谱数据进行特征光谱筛选,可以有效消除全光谱范围的冗余数据,提高光谱数据信噪比,同时降低计算数据量,提升运算速度。由于利用全波段光谱建立偏最小二乘算法建模时,各个数据段在采用时对模型的贡献率不同,因此可以将全波段光谱数据划分为 n 等份,分别建立每一个数据段的偏最小二乘模型并进行交叉验证,再根据模型的最小均方根误差和相关系数对划分的波段进行筛选,这种光谱筛选方法称作间隔偏最小二乘法 (Interval partial least squares, IPLS)。本文采用上述方法对核桃整仁样品的特征光谱进行筛选。

光谱数据的间隔宽度对特征光谱筛选的效果有较大的影响。间隔宽度过大有可能使信息量较高的光谱数据被排除;间隔宽度小,将使筛选出的波段增多,计算量增大,达不到光谱筛选的目的。因此采用

逐步增加间隔宽度的方法,以模型最佳交叉验证均方根误差 (REMSECV) 为指标,确定最佳间隔数目,步骤为:

(1) 以 4 个光谱数据点为初始间隔宽度。

(2) 在全光谱范围内以步长为 1 个光谱数据点移动间隔。

(3) 建立每一个间隔的 PLS 模型,并记录模型的交叉验证均方根误差 (REMSECV),最后确定最佳的 REMSECV 值。

(4) 将间隔宽度增加 4 个光谱数据点,重复步骤(2)、(3),直到间隔宽度达到 36 个光谱数据点为止。

图 3 为不同间隔宽度进行 IPLS 建模的影响,由图 3 可以看出,当间隔宽度取 12 个光谱数据点时,REMSECV 值达到最小。

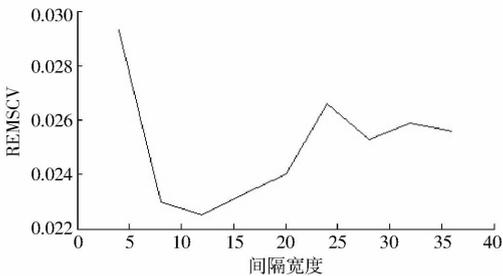


图 3 不同间隔宽度 IPLS 建模结果

Fig. 3 Modeling results of IPLS with different interval widths

确定间隔宽度后需要进一步确定参与建模的间隔数目,以完成特征光谱筛选,步骤如下:

(1) 根据确定的间隔宽度,进行全光谱数据划分。

(2) 对划分后的间隔进行 PLS 处理,并进行交叉验证后记录每一个间隔的相关系数 R 。

(3) 按照相关系数 R 由大到小的顺序对划分的间隔进行排序。

(4) 根据排序后的顺序依次增加间隔进行 PLS 建模处理,并用验证集计算模型的 RMSE 和 R 值。

图 4 是间隔宽度为 12 个光谱数据点时,采用 IPLS 方法进行特征光谱筛选与建模的结果。由图 4 可以看出,当参与建模的间隔数目为 2 时,模型的均方根误差和相关系数分别达到了最小值 0.021 和最

大值 0.913,因此当参与建模的间隔数目为 2 时,模型具有最好的预测精度和相关性。筛选出的对应特征波段为 1 056 ~ 1 081 nm 和 1 503 ~ 1 528 nm。筛选出的特征波段对应了 N—H 键的基频和倍频吸收谱带^[20],说明采用 IPLS 方法有效地筛选出了核桃仁蛋白质对应的近红外光谱特征波段,消除了冗余光谱信息,从而简化了模型。

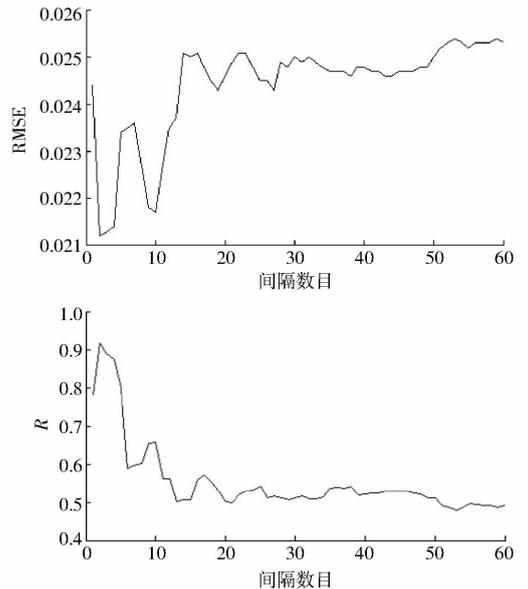


图 4 不同间隔数目 IPLS 建模结果

Fig. 4 Modeling results of IPLS with different interval numbers

3 结论

(1) 在全光谱范围内,3 种不同粒度核桃仁样品的蛋白质含量的 PLS 预测模型都取得了较高的精度,表明样品粒度大小对其蛋白质含量的预测影响较小,因此采用核桃整仁进行蛋白质含量近红外光谱预测具有较好的可行性。

(2) 间隔偏小二乘算法能够较好地实现核桃仁近红外特征光谱筛选,利用筛选出的特征光谱建立的蛋白质预测模型,针对验证集的均方根误差和相关系数分别为 0.021 和 0.913,有效地消除了冗余光谱信息,优化了模型质量,为核桃仁蛋白质快速无损在线检测提供思路和依据。

参 考 文 献

- 李忠新,杨莉玲,阿不力孜·巴斯提,等. 中国核桃产业发展研究[J]. 中国农机化学报,2013,34(4):23-28.
LI Zhongxin, YANG Liling, ABULIZI Basiti, et al. Research on the development of walnut in China[J]. Journal of Chinese Agricultural Mechanization, 2013, 34(4):23-28. (in Chinese)
- JIE Dengfei, XIE Lijuan, RAO Xiuqin, et al. Using visible and near infrared diffuse transmittance technique to predict soluble solids content of watermelon in an on-line detection system[J]. Postharvest Biology and Technology, 2014, 90:1-6
- DIMITRIS A, GERARD D, AMALIA G M S. Rapid non-destructive detection of spoilage of intact chicken breast muscle using near-infrared and Fourier transform mid-infrared spectroscopy and multivariate statistics[J]. Food Bioprocess Technol, 2012(5):338-

- 347.
- 4 KHANMOHAMMADI M, KARAMI F, MIR-MARQUÉS A, et al. Classification of persimmon fruit origin by near infrared spectrometry and least squares-support vector machines[J]. *Journal of Food Engineering*, 2014, 142:17 – 22.
- 5 田潇瑜,徐杨,彭彦昆,等. 基于光谱技术的牛肉多品质参数快速检测模型[J/OL]. *农业机械学报*,2013,44(增刊1):171 – 176. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=2013s131&flag=1. DOI:10.6041/j.issn.1000-1298.2013.S1.031.
- TIAN Xiaoyu, XU Yang, PENG Yankun, et al. Rapid detection model of beef quality based on spectroscopy[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2013,44(Supp.1):171 – 176. (in Chinese)
- 6 杨玮,李民赞,孙红,等. 温室黄瓜叶片近红外图像消噪算法与含氮量快速检测[J/OL]. *农业机械学报*,2013,44(7):216 – 221. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20130738&flag=1. DOI:10.6041/j.issn.1000-1298.2013.07.038.
- YANG Wei, LI Minzan, SUN Hong, et al. De-noising algorithm of multispectral images and nonlinear estimation of nitrogen content of cucumber leaves in greenhouse[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2013,44(7):216 – 221. (in Chinese)
- 7 胡晓男,彭云发,罗华平,等. 无信息变量消除法在筛选新疆红枣总酸近红外特征波长中的应用[J]. *食品工业*,2015(5):232 – 235.
- HU Xiaonan, PENG Yunfa, LUO Huaping, et al. The application that UVE method screens the NIR characteristic wavelengths of southern Xinjiang red jujube total cid[J]. *Science and Technology of Food Industry*, 2015(5):232 – 235. (in Chinese)
- 8 迟茜,王转卫,杨婷婷,等. 基于近红外高光谱成像的猕猴桃早期隐性损伤识别[J/OL]. *农业机械学报*,2015,46(3):235 – 241. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20150334&flag=1. DOI:10.6041/j.issn.1000-1298.2015.03.034.
- CHI Qian, WANG Zhuanwei, YANG Tingting, et al. Recognition of early hidden bruises on kiwifruits based on near-infrared hyperspectral imaging technology[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015,46(3):235 – 241. (in Chinese)
- 9 EVA B, JOSE M A, FRANS B, et al. Fast and robust discrimination of almonds (*Prunus amygdalus*) with respect to their bitterness by using near infrared and partial least squares-discriminant analysis[J]. *Food Chemistry*,2014,153:15 – 19.
- 10 韩东海,常冬,宋耀辉,等. 小型西瓜品质近红外无损检测的光谱信息采集[J/OL]. *农业机械学报*,2013,44(7):174 – 178. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20130730&flag=1. DOI:10.6041/j.issn.1000-1298.2013.07.030.
- HAN Donghai, CHANG Dong, SONG Shuhui, et al. Information collection of mini watermelon quality using near-infrared non-destructive detection[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2013,44(7):174 – 178. (in Chinese)
- 11 匡静云,管骁,刘静. 原料乳中蛋白质与脂肪的近红外光谱快速定量研究[J]. *分析科学学报*,2015,31(6):783 – 786.
- KUANG Jingyun, GUAN Xiao, LIU Jing. Rapid determination of protein and fat contents in raw milk by near infrared spectroscopy analysis[J]. *Journal of Analytical Science*, 2015,31(6):783 – 786. (in Chinese)
- 12 仇逊超,曹军. 近红外光谱波段优化在东北松子蛋白质定量检测中的应用[J]. *现代食品科技*,2016,32(11):303 – 309.
- QIU Xunchao, CAO Jun. The application of bands optimization of near infrared spectra in the quantitative detection of protein in northeastern pine nuts[J]. *Modern Food Science and Technology*, 2016,32(11):303 – 309. (in Chinese)
- 13 刘洁,李小昱,李培武,等. 基于近红外光谱的板栗水分检测方法[J]. *农业工程学报*,2010,26(2):338 – 341.
- LIU Jie, LI Xiaoyu, LI Peiwu. Determination of moisture in chestnuts using near infrared spectroscopy[J]. *Transactions of the CSAE*, 2010,26(2):338 – 341. (in Chinese)
- 14 JIANG Hui,LIU Guohai,XIAO Xiahong,et al. Classification of Chinese soybean paste by Fourier transform near-infrared (FT – NIR) spectroscopy and different supervised pattern recognition[J]. *Food Anal. Methods*, 2012, 5:928 – 934.
- 15 NICOLA B, VINCENZA P, PAOLA B, et al. Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy[J]. *Agricultural and Food Chemistry*,2005,53:8128 – 8134.
- 16 郑田甜,孙腾飞,曹增辉,等. 花生种子品质可见-近红外光谱分析[J]. *光谱学与光谱分析*,2015,35(3):622 – 625.
- ZHENG Tiantian, SUN Tengfei, CAO Zenghui, et al. Quality analysis of peanut seed by visible/near-infrared spectra[J]. *Spectroscopy and Spectral Analysis*, 2015,35(3):622 – 625. (in Chinese)
- 17 姜训鹏,雷恒,李海涛,等. 不同饲料原料日粮纤维水平的近红外测定方法[J]. *农业机械学报*,2016,47(增刊):353 – 358. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=2016s054&flag=1. DOI:10.6041/j.issn.1000-1298.2016.S0.054.
- JIANG Xunpeng,LEI Heng,LI Haitao,et al. Measurement of fiber content in different feed ingredients using near-infrared spectroscopy method[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016,47(Supp.):353 – 358. (in Chinese)
- 18 汪庆平,黎其万,董宝生,等. 近红外光谱法快速测定山核桃品质性状的研究[J]. *西南农业学报*,2009,22(3):873 – 875.
- WANG Qingping, LI Qiwan, DONG Baosheng, et al. Study on the prediction of oils of *Carya cathayensis* sargent using near infrared reflectance spectroscopy[J]. *Southwest China Journal of Agricultural Sciences*,2009,22(3):873 – 875. (in Chinese)
- 19 徐艳飞,回瑞华,候岩冬. 新疆种植核桃与野生核桃蛋白质含量的测定[C]//中国化学会第26届学术年会分析化学分会场论文集,2008.
- 20 李民赞. 光谱分析技术及其应用[M]. 北京:科学出版社,2007.