

傅立叶变换近红外光谱在大豆蛋白质和粗脂肪检测中的研究

朱贞映, 袁建*, 鞠兴荣, 何荣, 后其军, 魏孟辉

(南京财经大学食品科学与工程学院, 南京 210046)

摘要: **目的** 本文以317份不同品种的大豆为原料,开展了大豆样品粉碎粒度的蛋白质和粗脂肪含量的近红外研究,以期建立大豆品质检测方法。**方法** 45份大豆粉碎样品经不同的过筛处理,对剩余的272份大豆样品在最优的粉碎粒度下建模分析。**结果** 大豆粉碎过60目建模效果最好,蛋白质和粗脂肪含量近红外检测模型的内部交叉验证决定系数 r^2 分别为0.959和0.939;剩余272份大豆样品蛋白质含量的近红外检测模型的内部交叉验证相关系数为0.909,粗脂肪含量的近红外检测模型的内部交叉验证相关系数为0.918,外部验证蛋白质和粗脂肪决定系数 R^2 分别为0.944和0.911。**结论** 近红外光谱技术可用于大豆品质指标的检测。

关键词: 红外光谱法;大豆;蛋白质;脂肪

Detection of soybean protein and crude fat by Fourier transform-near infrared spectroscopy

ZHU Zhen-Ying, YUAN Jian*, JU Xing-Rong, HE Rong, HOU Qi-Jun, WEI Meng-Hui

(College of Food Science and Engineering, Nanjing University of Finance & Economics, Nanjing 210046, China)

ABSTRACT: Objective To research protein and crude fat content of crushed soybean with Fourier transform-near infrared spectroscopy(FT-NIR). For this purpose,317different varieties of soybean samples were prepared. **Method** Forty-five soybean powder samples were treated by different sieving process, the rest of 272 soybean samples were treated by modeling analysis under optimal crushed size. **Results** The results showed that crushing over 60 mesh sieve crushed soybean samples had the best modeling effect, coefficient of determination in internal cross-validation of protein and crude fat content were 0.959 and 0.939. Protein and crude fat content of internal cross-validation correlation coefficientofthe rest of 272 soybean samples were 0.909 and 0.918. Meanwhile, the external validation determinationcoefficient were 0.944 and 0.911. **Conclusion** The results showed that FT-NIR technology can be used for soybean quality detection.

KEY WORDS: near infrared spectroscopy; soybean; protein; fat

基金项目: 国家科技支撑计划课题(2013BAD17B03)

Fund: Supported by National Key Technology R&D Subject(2013BAD17B03)

*通讯作者: 袁建, 教授, 主要研究方向为食品质量安全评价与控制、粮油深加工。E-mail: yjian_nj@163.com

*Corresponding author: YUAN Jian, Professor, College of Food Science and Engineering, No. 3, Wenyuan Road, Qixia District, Nanjing 210046, China. E-mail: yjian_nj@163.com

1 引言

东北是我国主要的大豆生产和出口地,作为我国主要农产品的大豆,其籽粒中的蛋白质含量达 30%~50%,是我们平时饮食中丰富而又廉价的蛋白质主要来源^[1],另含有 18%左右的脂肪,是食用植物油重要来源^[2]。目前,我国颁布的有关蛋白质和粗脂肪的标准测定方法是凯式定氮和索式抽提法。虽然其可靠性较高,但操作步骤过于繁琐,我们迫切需要一种准确而又快速的检测方法。

近红外光谱法(near infrared spectroscopy, NIRS)是近年全球发展起来的一种快速测定方法^[3],它利用有机物在近红外光谱区的振动吸收而快速测定样品中多种化学成分含量。蛋白质、脂肪糖类等含有的各种含氢基团的倍频与合频谱带恰好落在近红外区,以得到这些有机分子含氢基团的特征振动信息,从而测定其化学成分的含量^[4-5],Norris 是第一个将近红外光谱技术应用到农业领域^[6],后来该分析技术被广泛应用于小麦、玉米、芝麻等农产品的品质检测中^[7-13]。

本研究采用近红外漫反射光谱技术分析测定了东北 317 份大豆样品蛋白质、粗脂肪含量。首先,用 45 份大豆粉碎样品经不同的过筛处理建模(0、20、40、60、80 目),选择粉碎粒度建模结果最优的大豆样品;其次,对剩余的 272 份大豆样品在最优的粉碎粒度下建模分析;最后,对模型的预测能力进行外部验证,为以后近红外漫反射光谱技术在我国大豆品质分析中的应用提供参考。

2 材料与方 法

2.1 材料与设备

从东北地区收集到的不同品种的大豆样品共 317 份。每种样品取一定量经粉碎过筛处理,备用,其余完整颗粒用自封袋包装保存。所有实验样品都经过预处理,如除杂,去生霉未熟等。

MB3600 高效 FTIR 傅立叶变换近红外光谱仪(ABB 公司,扫描光谱范围为 3700~15000cm⁻¹);快速消解仪(瑞士步琪有限公司);脂肪测定仪(瑞士步琪有限公司);高速万能粉碎机(天津市泰斯特仪器有限公司)。

2.2 方 法

2.2.1 化学分析测定方法

按照国标 GB/T 5511^[14]采用凯氏定氮法测定大豆中的蛋白质含量;按照国标 GB/T 5512^[15]采用索氏抽提法测定大豆中的粗脂肪含量。每个样品重复测定 3 次,取其平均值作为最终结果。

2.2.2 大豆样品集的划分

随机挑选总样品中的 45 个大豆品种,对 45 个大豆品种粉碎过筛,粒度分别为 0、20、40、60、80 目,建立相应的模型,找到最好的建模粒度。剩下的 272 个大豆样品的 70%(190 份)作为定标样品集,用于模型的建立,其余 30%(82 份)用作外部验证,验证集用于定标模型预测的性能评价。

2.2.3 近红外光谱收集

傅立叶变换近红外光谱仪开机预热 30 min 后,对定标样品 317 个大豆样品进行光谱扫描与存储,扫描光谱范围为 4000~12600 cm⁻¹,分辨率 16 cm⁻¹,扫描次数 60 次,每个样品重复 3 次以消除客观因素引起的差异。

2.2.4 近红外模型与评价

对大豆粉碎样品最优数学方法和回归技术的研究,实验采用 Horizon MB 化学计量软件建模分析,对定标集样品做预处理后经适当的光谱数学处理,例如多元散射校正(multiple scattering correction),偏差校正(offset correction)、归一化(standard normal variate)等,建立最佳的大豆蛋白质含量和粗脂肪含量的检测模型。

采用内部交叉验证和外部验证两种方法评价模型的预测性能。内部交叉验证通过预测残差总和(PRESS)、交互验证标准误差(SECV)和交叉验证决定系数 R^2 评价检测模型的优越性, PRESS 和 SECV 越小, R^2 越大,模型预测性能越好。外部验证是验证集样品用于定标模型预测的性能评价,通过决定系数 (R^2)和显著性 P 值评价模型的预测性能, R^2 越大, P 值小于 0.05,模型预测性能越好。

3 结果与讨论

3.1 大豆建模粉碎粒度的选择

选用 45 个不同品种的大豆样品,采用国标方法分别测定大豆的蛋白质和粗脂肪含量,蛋白质含量范围为 36.09%~41.36%,粗脂肪含量范围为 17.90%~

20.57%。对这 45 个大豆样品过筛处理建模, 近红外建模结果显示, 大豆样品经粉碎处理后, 最优的建模粒度为粉碎过筛 60 目, 此时的 PRESS 和 SECV 最小, 交叉验证决定系数 R^2 最大, 大豆蛋白质建模结果 PRESS、SECV 和 R^2 分别为 0.902、0.143 和 0.959, 大豆粗脂肪建模结果 PRESS、SECV 和 R^2 分别为 0.225、0.071 和 0.939; 如表 1 所示, 随着大豆样品粉碎粒度的精细, 建模效果越来越好, 60 目达到最好的建模结果, 之后建模效果开始下降, 40 目建模优于 80 目, 后续实验选用粉碎 60 目为大豆蛋白质和粗脂肪建模。

3.2 大豆的蛋白质和粗脂肪的化学测定值

根据国标对大豆各指标化学测定值进行统计分析, 实验所选用的 272 个样品, 大豆蛋白质和粗脂肪含量范围分别为 37.039~43.561%、15.775~25.897%, 平均蛋白质为 40.580%, 平均粗脂肪为 21.892%; 实验选用的 190 个定标集样品中大豆蛋白质和粗脂肪

含量范围分别为 37.984~42.938%、15.775~25.897%; 平均蛋白质为 40.571%, 平均粗脂肪为 22.001%; 实验选用的 82 个验证集样品中大豆蛋白质和粗脂肪含量范围分别为 37.039~43.561%、17.077~25.006%, 平均蛋白质为 40.601%, 平均粗脂肪为 21.639%。从粗脂肪和蛋白的偏度和峰度值看都符合正态分布, 且粗脂肪比蛋白质符合更好的正态分布, 其值分布范围广, 相互之间的差异性突出, 为建立优质的模型提供了有利条件。

3.3 大豆的近红外光谱图与预处理

图 1 为经过傅立叶变换近红外光谱仪扫描得到 190 个大豆样品的近红外光谱图, 由图可以看出各个样品的谱线具有基本相同的变化趋势, 但又各具差异, 显示了不同样品间的差别, 不同样品的谱线可能存在漂移和噪音等, 对不同样品进行基线校准、平滑处理和多元散射校正(MSC)等, 最终得到大豆样品的近红外光谱图, 如图 2。

表 1 大豆蛋白质和脂肪含量的过筛结果分析

Table 1 Sieving result analysis of soybean protein and fat content

		0 目	20 目	40 目	60 目	80 目
蛋白质	SECV	0.296	0.174	0.152	0.143	0.169
	PRESS	3.948	1.360	1.041	0.902	1.255
	R^2	0.932	0.939	0.953	0.959	0.945
粗脂肪	SECV	0.228	0.080	0.077	0.071	0.081
	PRESS	2.333	0.287	0.269	0.225	0.285
	R^2	0.895	0.909	0.934	0.939	0.927

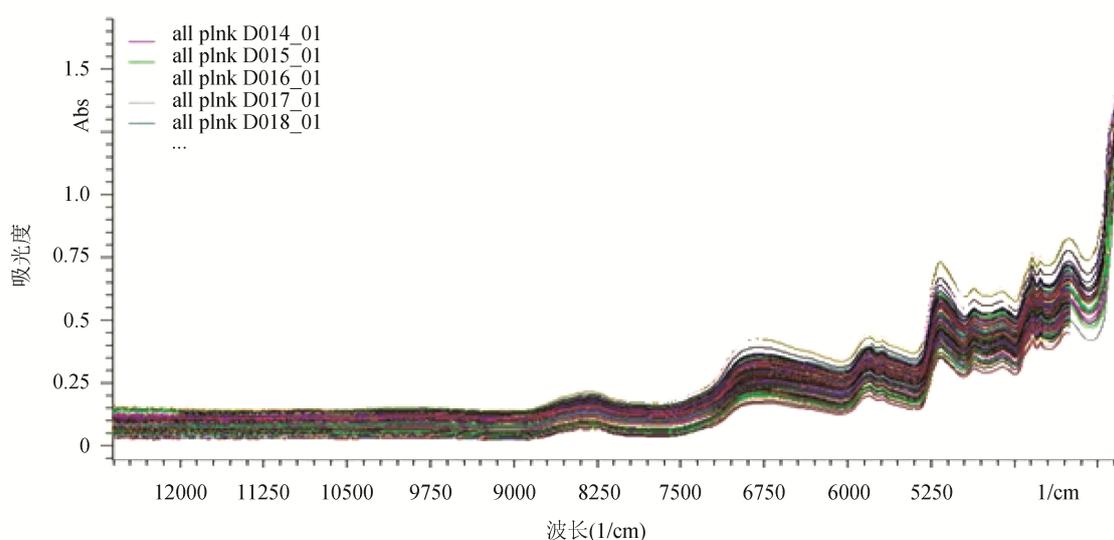


图 1 大豆的近红外光谱图

Fig. 1 Near infrared spectrum of soybean

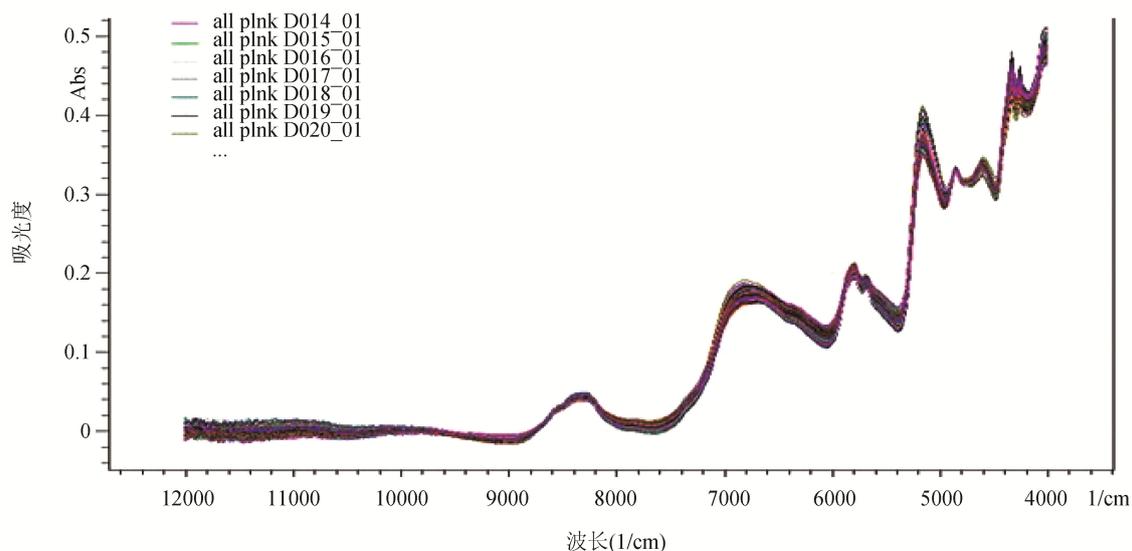


图 2 经基线校准、平滑和标准多元散射校正处理的近红外光谱图

Fig. 2 Near-infrared Spectra after baseline correction, smoothing and standard multiplicative scatter correction process

3.4 近红外测定模型的建立与优化

3.4.1 校正方法的选择

利用 Horizon MB 化学计量软件建模分析, 对定标集样品做预处理后经适当的数学处理, 例如多元散射校正(MSC)、趋势变换法(Detrending)、偏差校正(Offset Correction)、归一化(Standard Normal Variate)等, 建立最佳的大豆蛋白质和粗脂肪含量检测模型。由表 2 可知, 对于大豆蛋白质采用偏差校正的多元散射校正方法最好, PRESS 为 16.062, SECV 为 0.291, 交叉验证决定系数 R^2 为 0.909, 同样, 对于大豆粗脂肪采用归一化的多元散射校正方法最好, PRESS 为 64.674, SECV 为 0.584, 交叉验证决定系数 R^2 为 0.918。

3.4.2 回归技术的选择

由以上校正方法和表 4 回归技术的分析, 最终大豆蛋白质的建模可知采用偏差校正的多元散射校正方法和偏最小二乘法(PLS)回归技术, 大豆粗脂肪的建模可以采用归一化的多元散射校正方法和偏最小二乘法(PLS)回归技术。

3.4.3 近红外定标集建模结果

图 3 为采用最优的校正方法和回归技术建立的最佳大豆近红外检测模型, 由图可知蛋白质含量的近红外检测模型的交叉验证相关系数 R^2 为 0.909, 粗脂肪含量的近红外, 检测模型的交叉验证相关系数 R^2 为 0.918。

3.5 近红模型的外部验证

3.5.1 外部验证结果与分析

将各个验证集样品的化学测定值与通过模型预测的值进行比较(图 4)对模型进行外部验证, 由表 5 和表 6 可知经外部验证, 蛋白质含量的近红外检测模型的预测相关系数 R^2 达 0.944, 预测平均偏差为 0.227%; 大豆粗脂肪含量的近红外检测模型的预测相关系数 R^2 也达到 0.911, 预测平均偏差为 0.479%, 说明回归的拟合度非常高, 通过对大豆蛋白质和粗脂肪外部验证进行 Durbin-Watson 统计量分析(Durbin-Watson 检验常用来检测残差是否存在自相关, 其取值范围介于 0~4 之间, 其值越接近 2, 说明变量之间是相互独立的), 大豆蛋白质和粗脂肪的 Durbin-Watson 值都很接近 2, 说明模型的残差不存在自相关, 可以认为回归方程可以概括因变量的变化, 其中蛋白质和脂肪其预测偏差平均值都在原始国标实验误差要求内。由表 6 可以得到回归部分的显著性 P 值都是 0.000, 小于显著水平 0.05, 由此可以判断由模型得到的样品预测值对样品的真实值解释非常显著, 综合外部验证的结果分析知定标模型对外部验证集样品预测的性能是可信的, 说明大豆蛋白质和粗脂肪含量的近红外模型均具有良好的代表性和预测能力。

表 2 不同校正方法对大豆蛋白质近红外检测模型的影响
Table 2 Effect of different correction methods on soybean protein NIR model

成分	校正方法	PRESS	SECV	R^2
蛋白质	MSC	16.102	0.291	0.909
	Detrending	23.824	0.354	0.865
	Offset Correction	32.574	0.414	0.815
	Standard Normal Variate	35.276	0.431	0.800
	MSC/ Detrending	28.821	0.389	0.836
	MSC/ Offset Correction	16.062	0.291	0.909
	MSC/ Standard Normal Variate	16.090	0.291	0.909
	Detrending/ Offset Correction	24.376	0.358	0.862
	Detrending/ Standard Normal Variate	26.697	0.375	0.848
	Offset Correction/ Standard Normal Variate	25.507	0.366	0.855

表 3 不同校正方法对大豆粗脂肪近红外检测模型的影响
Table 3 Effect of different correction methods on soybean fat NIR model

成分	校正方法	PRESS	SECV	R^2
粗脂肪	MSC	64.763	0.584	0.918
	Detrending	73.103	0.620	0.907
	Offset Correction	64.798	0.587	0.918
	Standard Normal Variate	85.513	0.671	0.891
	MSC/ Detrending	68.699	0.601	0.913
	MSC/ Offset Correction	64.763	0.584	0.918
	MSC/ Standard Normal Variate	64.647	0.584	0.918
	Detrending / Offset Correction	74.856	0.628	0.905
	Detrending / Standard Normal Variate	73.245	0.621	0.907
	Offset Correction/ Standard Normal Variate	85.513	0.671	0.891

表 4 不同回归技术对大豆近红外检测模型的影响
Table 4 Effect of different regression on soybean NIR model

成分	回归技术	PRESS	SECV	R^2
蛋白质	PLS / Offset Correction/MSC	16.062	0.291	0.909
	SIMPLS / Offset Correction/MSC	16.062	0.291	0.909
	PCR / Offset Correction/MSC	16.767	0.297	0.905
粗脂肪	PLS / Standard Normal Variate/MSC	64.763	0.584	0.918
	SIMPLS / Standard Normal Variate/MSC	64.763	0.584	0.918
	PCR / Standard Normal Variate/MSC	123.715	0.807	0.843

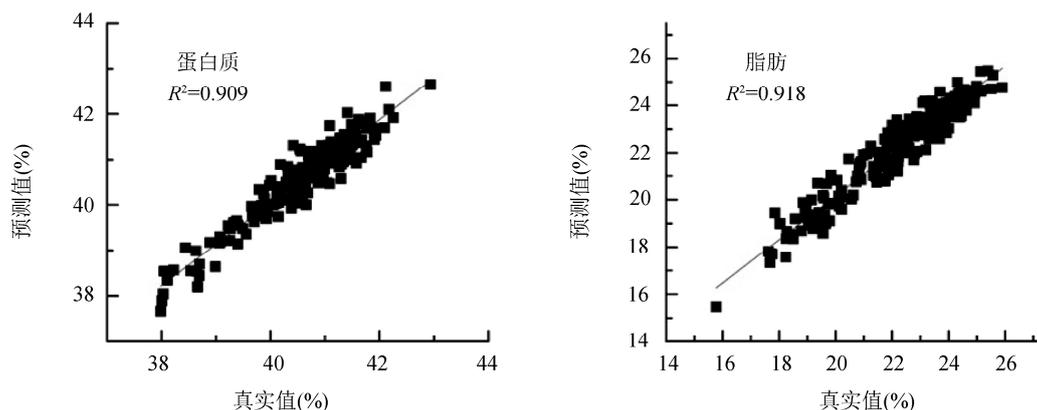


图 3 大豆蛋白质和粗脂肪定标集模型

Fig. 3 Calibration set model of soybean protein and fat

表 5 外部验证真实值与预测值对比

Table 5 Comparison between actual and predicted values by external verification

	偏差最小值	偏差最大值	平均偏差	偏差标准差
蛋白质	0.002	0.625	0.227	0.189
粗脂肪	0.006	1.356	0.485	0.329

表 6 外部验证评价模型的统计量

Table 6 Statistics of external authentication Evaluation Model

	R	R ²	调整 R ²	预测平均偏差	Durbin-Watson	显著性 P 值
蛋白质	0.972	0.944	0.943	0.227	2.026	0.000
粗脂肪	0.955	0.911	0.910	0.479	1.930	0.000

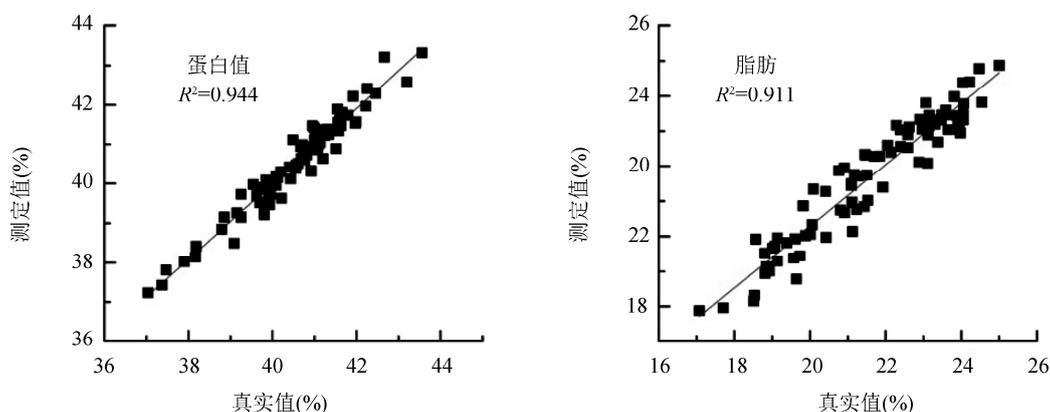


图 4 大豆验证集样品的蛋白质和粗脂肪测定值与真实值的关系

Fig. 4 Relationship between protein and fat estimated value and the real value of soybean validation set of samples

4 结论

利用常规分析方法测定大豆蛋白质、脂肪含量是一个繁琐的过程, 样品需前处理, 既费时又费力, 且

需消耗水、电、化学试剂, 同时还会造成环境污染; 而近红外分析技术具有快速、简便、无化学试剂污染等优点, 更重要的是只要建立了相应指标的模型, 就能同时对一份大豆材料进行相应指标含量的数据输出。

本文基于傅立叶变换近红外光谱分析技术,开展了选取最优数学的数学处理方法和回归技术分别建立了大豆蛋白质和粗脂肪含量的近红外检测模型,并对模型进行内部交叉验证和外部验证。首先,选用45个不同品种的大豆样品,样品经粉碎处理,最优的建模粒度为粉碎过筛60目,此时的PRESS和SECV最小,交叉验证决定系数 R^2 最大;其次,对剩余的272份大豆样品在最优的粉碎粒度下建模分析,建模结果显示,大豆蛋白质含量的近红外检测模型的内部交叉验证相关系数 R^2 为0.909,粗脂肪含量的近红外检测模型的内部交叉验证相关系数 R^2 为0.918,其外部验证通过对模型预测性能的决定系数 R^2 和显著性 P 值验证,外部验证蛋白质和粗脂肪 R^2 分别为0.944和0.911,说明拟合性很好可以认为回归方程能够概括预测值对真实值的变化,同样,方差分析 P 值小于0.05,由此可以判断由模型得到的样品预测值对样品的真实值解释非常显著,综合外部验证的结果分析知定标建立的模型对外部验证集样品预测的性能是可信的,说明应用近红外光谱分析技术建立大豆主要成分含量的近红外检测模型是可行的,可用于大豆成分含量的快速检测。

参考文献

- [1] 李里特,李再贵,殷丽君.大豆加工与利用[M].北京:化学工业出版社,2003:20-26.
Li LT, Li RG, Yin LJ. Soybean processing and utilization [M]. Beijing: ChemInd Press, 2003:20-26.
- [2] 邱丽娟,常汝镇.大豆种质资源描述规范和数据标准[M].北京:中国农业出版社,2006.
Qiu LJ, Chang RZ. Soybean germplasm description specifications and data standards [M]. Beijing: Chin Agric Press, 2006.
- [3] 陆婉珍.现代近红外光谱分析技术[M].中国石化出版社,2007.
Lu WZ. Modern Near Infrared Spectroscopy[M]. Chin Petrochemical Press, 2007.
- [4] 刘晓庚.近红外光谱技术在粮油储藏及其品质分析中的应用[J].粮食储藏,2007,36(01):29-35.
Liu XG. Application of near infrared spectroscopy in grain and oil storage and quality analysis [J]. Grain Storage, 2007, 36(01): 29-35.
- [5] 姚鑫森,张瑞英,李霞辉,等.近红外透射光谱法(NITS)分析大豆品质的研究[J].大豆科学,2006,25(4):417-420.
Yao XM, Zhang RY, Li XH, *et al.* The study of soybean quality analysis by near infrared transmission spectroscopy (NIRS) [J]. Soybean Sci, 2006, 25(4): 417-420.
- [6] Williams P C, Stevenson S G, Starkey P M, *et al.* The application of near infrared reflectance spectroscopy to protein-testing in pulse breeding programmes [J]. J Sci Food Agri, 1978, 29(3): 285-292.
- [7] 郭蕊,王金水,金华丽,等.近红外光谱分析技术测定芝麻水分含量的研究[C]2010年中国农业工程学会农产品加工及贮藏工程分会学术年会暨华南地区农产品加工产学研研讨会论文摘要集.2010:366-369.
Guo R, Wang JS, Jin HL, *et al.* Study on determination of the sesame moisture content by Near Infrared Spectroscopy [C]. 2010 Chinese Society of Agricultural Engineering Agricultural Products Processing and Storage Engineering Branch Annual Conference and processing of agricultural products in South China research seminar papers Abstracts.2010: 366-369.
- [8] Kovalenko IV, Rippke GR, Hurburgh CR. Determination of Amino acid composition of soybeans (glycine max) by near-infrared spectroscopy [J]. J Agric Food Chem, 2006, 54(10): 3485-3491.
- [9] 王若兰,王春华,黄亚伟.小麦脂肪酸值的近红外光谱快速测定研究[J].现代食品科技,2013,2(2):393-396.
Wang RL, Wang CH, Huang YW. Rapid determination study of wheat fatty acid value by near infrared spectroscopy [J]. Mod Food Sci Tech, 2013, 2(2): 393-396.
- [10] Fei L, Fan Z, Jin Z, *et al.* Determination of acetolactate synthase activity and protein content of oilseed rape (*Brassica napus* L.) leaves using visible/near [J]. Anal Chim Act, 2008, 629(1-2): 56-65.
- [11] 齐波,张宁,赵团结,等.利用高光谱技术估测大豆育种材料的叶面积指数[J].作物学报,2015,7(07):1073-1085.
Qi B, Zhang N, Zhao TJ, *et al.* Using hyperspectral technology to estimate leaf area index of soybean breeding material. 2015, 7(07): 1073-1085.
- [12] 王立琦,崔月,张欢,等.基于SVM的大豆油脂色泽近红外光谱分析[J].中国粮油学报,2015,08期(08):115-119.
Wang LQ, Cui Y, Zhang H, *et al.* Near infrared spectroscopy analysis of SVM-based soybean oil color [J]. J Chin Cereal Oil Assoc, 2015, 8(08):115-119.
- [13] Costa GBD, Fernandes D, Gomes AA, *et al.* Using near infrared spectroscopy to classify soybean oil according to expiration date [J]. Food Chem, 2016, 196: 539-543..

- [14] GB/T5511-2008 谷物和豆类氮含量测定和粗蛋白质含量计算凯氏法[S].
GB/T5511-2008 Kjeldahl method of Calculating cereals and legumes nitrogen content and crude protein content[S].
- [15] GB/T5512-2008 粮油检测粮食中粗脂肪含量测定[S].
GB/T5512-2008 Cereals, Oils and fat content in food detection assay[S].

(责任编辑: 金延秋)

作者简介



朱贞映, 硕士, 主要研究方向为食品安全。

E-mail: 616687280@qq.com



袁建, 教授, 硕士生导师, 主要研究方向为食品质量安全评价与控制、粮油深加工。

E-mail: yjian_nj@163.com