

文章编号: 1001 - 9014 (2009) 05 - 0353 - 04

基于可见 / 近红外反射光谱的稻米品种与真伪鉴别

梁亮^{1,2,3}, 刘志霄², 杨敏华¹, 张佑祥², 汪承华²

(1. 中南大学 信息物理工程学院, 湖南 长沙 410083;

2 吉首大学 生物资源与环境科学学院, 湖南 吉首 416000;

3 中南林业科技大学 林业遥感信息工程研究中心, 湖南 长沙 410004)

摘要: 利用可见 / 近红外光谱技术对市场上 5 种稻米进行了鉴别。以 ASD FieldSpec® 3 地物光谱仪采集了 5 种稻米的光谱数据, 各获取 35 个样本, 随机分成训练集 (150 份) 和检验集 (25 份), 并分别采取全波段与特征波段 (400 ~ 500nm、910 ~ 1400nm 与 1940 ~ 2300nm) 两种方法建立模型进行分析。光谱经 S-Golay 平滑和标准归一化 (SNV) 处理后, 以主成分分析法 (PCA) 降维。将降维所得的前 9 个主成分数据作为 BP 人工神经网络 (BP-ANN) 的输入变量, 稻米品种作为输出变量, 建立 3 层 BP-ANN 鉴别模型。利用 25 个未知样对模型进行检验, 结果表明两类模型预测准确率均高达 100%, 其中特征波段模型比全波段模型具有更高的预测精度, 说明利用可见 / 近红外技术结合 PCA-BP 神经网络分析法进行稻米品种与真伪的快速、无损鉴别是可行的, 且提取特征波段是优化模型的有效方法之一。

关键词: 可见 / 近红外光谱; 稻米; 主成分分析; BP 人工神经网络; 鉴别

中图分类号: S123, S511 **文献标识码:** 28260

DISCRIMINATION OF VARIETY AND AUTHENTICITY FOR RICE BASED ON VISUAL / NEAR INFRARED REFLECTION SPECTRA

LIANG Liang^{1,2,3}, LIU Zhi-Xiao², YANG Min-Hua¹, ZHANG You-Xiang², WANG Cheng-Hua²

(1. School of Info-Physics and Geomatics Engineering, Central South University, Changsha 410083, China;

2 College of Biology and Environmental Sciences, Jishou University, Jishou 416000, China;

3 Research Center of Forestry Remote Sensing and Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China)

Abstract: Five different varieties of rice bought from the supermarket were identified with visual/near infrared spectroscopy (NIRS) technology. The spectra were collected by using ASD FieldSpec® 3 spectrometer, and 35 samples were obtained for each variety of rice. All the samples were divided randomly into two groups, one group with 150 ones used as calibrated set, and the other with 25 ones as validated set. Then the samples were analyzed with the whole wave band and the characteristic wave band (400 ~ 500nm, 910 ~ 1400nm and 1940 ~ 2300nm) models, respectively. The samples data were pretreated by the methods of S-Golay smoothing and standard normal variable (SNV), and then the pretreated spectra data were analyzed with principal component analysis (PCA). The anterior 9 principal components computed by PCA were used as the input variables of back-propagation artificial neural network (BP-ANN) which included one hidden layer, while the values of rice varieties were used as the output variables of BP-ANN, and then the three layers BP-ANN discrimination model was built. The 25 unknown samples in the validated set were predicted by ANN-BP model. The results show that the discriminating rates are 100% in both models, and the accuracy of characteristic wave band model is higher than that of the whole wave band model. The results indicate that it is feasible to discriminate the variety and authenticity of rice by visible / near infrared reflectance spectra as a rapid and non-contact way, and selecting characteristic wave band is one of the validated methods to improve the precision of the discrimination model.

Key words: visual/near infrared spectra; rice; principal component analysis (PCA); BP-artificial neural network; discrimination

引言

近年来国内大米市场混乱,各种低档、劣质稻米冒充品牌稻米销售,严重侵犯了消费者的利益,甚至危及生命安全。而目前,稻米品质与品种的鉴别方法主要为感觉法与化学分析法。感觉法通过形状、色泽以及气味对大米品质进行鉴定,主要依赖鉴别者的经验,缺乏严格的科学依据;常规化学分析法则存在耗时长、耗资大以及操作复杂等缺点而难以满足市场经营中快速鉴别的需要。

而可见近红外反射光谱技术因其快速、高效以及无损的特点,已被广泛地应用于石油化工、探矿、制药以及纺织等领域^[1]。近年来,可见近红外光谱结合模式识别技术,进一步在杨梅、燕麦等农产品的分类中获得了成功的应用^[2,3],对丹参、白术、蛇床子等中药材产地与真伪鉴别也取得了满意的结果^[4-6]。在稻谷的相关研究中,可见近红外光谱技术也已用于其直链淀粉、蛋白质、脂肪以及氨基酸含量与稻谷储存年份的分析^[7-12]。但目前,利用可见/近红外光谱技术对稻米的研究主要应用于营养成分的定量分析,而进行稻米品种与真伪鉴别的文献尚未见有报道。因此,本文采用可见近红外反射光谱技术对稻米的品种与真伪进行鉴别,从而为稻米品种的快速无损鉴别提供了新的方法。

1 材料与方法

1.1 试验仪器

试验仪器主要包括 FieldSpec® 3 地物光谱仪、恒温箱以及直径为 75mm 玻璃皿若干。FieldSpec® 3 地物光谱仪为美国 ASD (Analytical Spectral Device) 公司的新产品,波长范围 350 ~ 2500nm,光谱分辨率在 350 ~ 1400nm 为 3nm,1400 ~ 2500nm 为 10nm。光谱数据在 ASD View Spec Pro 中以 ASCII 码形式导出,再导入 Unscramble 9.7 与 DPS (data procession system for practical statistics) 软件中进行处理。

1.2 样本来源与数据采集

于超市购买了泰国香米、夜郎贡米、珍珠米、泰国糯米四种稻米样本;于湘西质量技术监督局获得劣质米(珍珠米伪品)样本。每种稻米采集样本 35 份,合计 175 份。随机抽取 150 份(每种 30 份)作为训练集,其余 25 份(每种 5 份)作为预测集。样本于 45℃ 下恒温烘干 24h 后,于玻璃皿中进行光谱扫描,米样上表面均与玻璃皿上端保持平齐。测样于暗室

中进行,以卤素灯为唯一光源。光纤探测器头部与样本之间保持垂直,下部距离样本 0.5cm,每一样本重复测量 30 次取其均值。

1.3 光谱预处理

由于光谱在小于 400nm 与大于 2300nm 的波段噪声较大,因此本研究选用 400 ~ 2300nm 波段作为有效光谱数据进行分析。将有效光谱经 S-Golay 平滑后,采用标准归一化(SNV)方法进行除噪。

1.4 数据降维与波段选择

如果将光谱数据直接作为变量输入进行建模,不但会因变量太多而增加建模难度,而且会引入噪声而降低模型的预测精度。为了避免这一问题,本研究采用主成分分析(PCA)以实现光谱数据的降维。同时,如能选用特征波段进行建模,将信噪比低的波段删除,可能会获得比全波段建模更好的结果。因此,本研究将对全波段与特征波段两种建模方法进行比较分析以获得较优模型。

1.5 人工神经网络模型

在光谱分析中,人工神经网络是一种重要的模式识别方法,其中多层误差反向传播神经网络方法(back-propagation, BP)应用尤广,具有强大的非线性建模能力,特别适合解决复杂的映射问题^[13,14]。而作为一种有机物,稻米内部的理化性质与其可见/近红外反射光谱之间正是一种复杂的映射关系,因此,本研究将全波段与特征波段降维后的数据分别导入 DPS 中,采用 BP 算法建立不同稻米的鉴别模型。

2 结果与讨论

2.1 稻米样本的可见近红外漫反射光谱

图 1 为 5 种稻米部分可见近红外漫反射光谱曲线。从图中可看出,不同品种稻米的反射光谱的波形有稍许差异,而同种稻米样本的光谱则有一定的聚集趋势,但凭肉眼难以准确区分。分析时,先将光谱数据转化为 ASCII 码,在 Unscramble 9.7 中完成预处理后进行 PCA 分析。

2.2 PCA 分析与特征波段提取

训练集样本经 PCA 降维分析后,分别以前 3 个主成分 PC_1、PC_2、PC_3 作为 x、y、z 坐标,建立各样本的三维得分图(图 2),以表征样本在该三维空间中的分布。由于前 3 个主成分对光谱矩阵的累积方差贡献达 91.82%,因此,样本在三维空间的分布可大体反应其在超维空间的分布特征,表征出不同稻米的聚类结果。从图 2 中可看出,各种稻米有良好

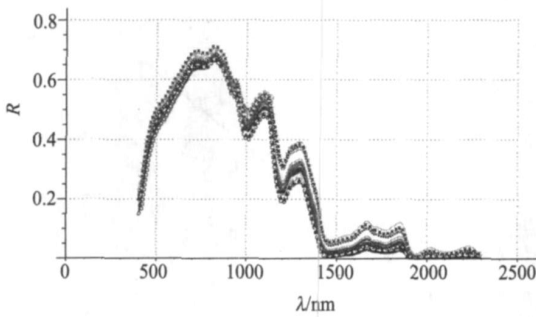


图1 5种稻米样本的可见/近红外反射光谱
Fig.1 The visual/near infrared reflection spectra for 5 varieties of rice

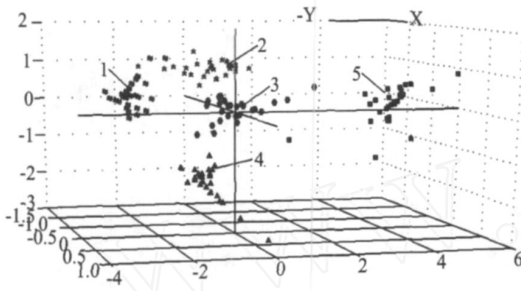


图2 5种稻米前3个主成分的得分聚类
Fig.2 Score cluster plot using top three principal components (PCs) for 5 varieties of rice

注:1. 泰国香米;2. 夜郎贡米;3. 珍珠米;4. 泰国糯米;5. 珍珠米伪品
Note:1. Thailand fragrant rice; 2. Ye-lang tribute rice; 3. Pearl rice; 4. Thailand sticky rice; 5. Tinpot pearl rice

的聚类趋势,可进行定性分析,但要取得精确的定量分析结果,还需要建立鉴别能力更强的模型。

由于 PCA 分析的前 3 个主成分已包含了绝大部分的分类信息,因此,可根据不同波段的光谱对前 3 个主成分的贡献值分析出稻米分类的特征波段。在 Unscramble 9.7 中,得出前 3 个主成分的 X 加载图,以表征各波段对模型前 3 个主成分的贡献大小(图 3)。以波谱区域对 PC₁ 的贡献值为主要参考指标,并综合考虑对 PC₂ 与 PC₃ 的贡献值,得出 400~500nm、910~1300nm 与 1940~2300nm 三个波段为稻米鉴别的特征波段。根据水、淀粉、蛋白质的吸收特征^[15,16],可分析出 910~1400nm、1940~2300nm 两段特征波段主要反映了不同稻米营养成分的差异,而 400~500nm 这一位于可见光部分的特征波段则主要反映了形状与颜色的差异。

在建立模型过程中,如果所选取的主成分过少,将会因不充分拟合而导致模型预测准确度降低;而若选用的主成分过多,则会产生过拟合现象而导致模型预测的准确率下降。因此,本研究通过交互验证

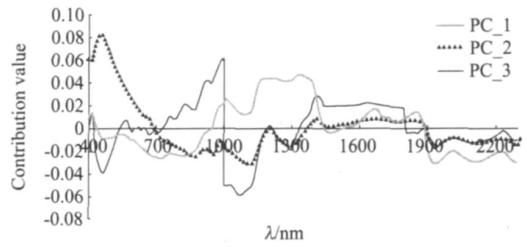


图3 不同谱区对 PCA 分析前 3 个主成分的贡献值
Fig.3 Contribution value of spectral region to top three principal components from PCA analysis

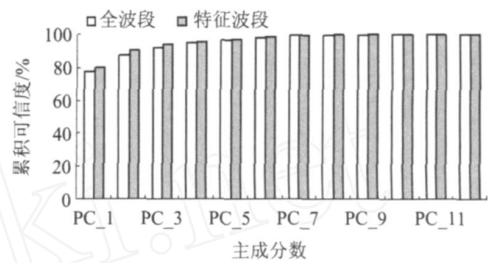


图4 两种模型训练集前 12 个主成分的累积可信度
Fig.4 Accumulative reliabilities plot of the top 12 principal components of two kinds of PCA model

确定最佳主成分数,即在累积可信度(累积方差贡献)变化不大的情况下选取较少的主成分数。将全光谱与特征光谱分别进行 PCA 分析后,所得前 12 个主成分的累积可信度如图 4 所示。由图可知,两种方法前 9 个主成分的累积可信度在 99.5% 以上,包含了光谱数据绝大部分的特征信息。

2.3 BP神经网络模型预测结果与比较分析

利用训练集中的 150 个样本,以 PCA 降维得到的前 9 个主成分作为 BP 神经网络的输入变量,在 DPS 中建立 PCA-BP 神经网络预测模型。

建模分析时,泰国香米、夜郎贡米、珍珠米、泰国糯米、劣质米分别赋值为 1.0000、2.0000、3.0000、4.0000、5.0000;BP 网络各层间采用 Sigmoid 激励函数,其中 Sigmoid 参数取 0.9,动态参数取 0.6,最小训练速度设为 0.1,允许误差设为 0.0001,最大迭代次数设为 3000 次。通过调节隐含层的节点数反复地验证以优化网络结构,得到最佳的 BP 网络结构为 9-6-5 三层 BP 神经网络模型。用模型对预测集的 25 个未知样进行预测,结果表明两类模型对所有样本预测的正确率均达 100% (表 1)。

为获得最佳建模方法,分析了两种方法对 25 个未知样的预测结果。两类模型对预测集的拟合结果与标准值之间的回归关系见图 5。由图可知,两回归方程的斜率都接近于 1,但特征波段模型预测集决

表 1 两种模型对 25 个未知样本的预测结果

Table 1 Prediction results for 25 unknown samples by two BP-ANN models

样本号	标准值	预测值 (全波段)	预测值 (特征波段)
(1)	1.0000	1.0069	1.0153
(2)	1.0000	1.0233	1.0343
(3)	1.0000	1.0033	1.0116
(4)	1.0000	1.0054	1.0135
(5)	1.0000	1.0040	1.0120
(6)	2.0000	2.0670	2.0027
(7)	2.0000	2.1387	2.0692
(8)	2.0000	2.0578	2.0138
(9)	2.0000	2.0851	2.0095
(10)	2.0000	2.0697	2.0024
(11)	3.0000	2.9325	2.9863
(12)	3.0000	3.0064	3.0041
(13)	3.0000	2.9817	2.9540
(14)	3.0000	2.9933	3.0553
(15)	3.0000	2.9558	2.9705
(16)	4.0000	4.0511	3.9270
(17)	4.0000	4.0292	3.9496
(18)	4.0000	4.1280	4.0816
(19)	4.0000	4.0350	3.9267
(20)	4.0000	4.0641	3.9421
(21)	5.0000	4.9876	4.9976
(22)	5.0000	4.9716	4.9912
(23)	5.0000	4.9544	4.9968
(24)	5.0000	4.9821	4.9945
(25)	5.0000	4.9851	4.9963

注: (1)~(5), 泰国香米; (6)~(10), 夜郎贡米; (11)~(15), 珍珠米; (16)~(20), 泰国糯米; (21)~(25), 珍珠米伪品

Note: (1)~(5), Thailand fragrant rice; (6)~(10), Ye-lang tribute rice; (11)~(15), Pearl rice; (16)~(20), Thailand sticky rice; (21)~(25), Tinpot pearl rice

定系数 ($R^2 = 0.9994$) 比全波段模型 ($R^2 = 0.9988$) 稍高, 而预测标准误差 ($SEP = 0.0390$) 与预测误差均方根 ($RMSEP = 0.0383$) 则比全波段模型 ($SEP = 0.0519$; $RMSEP = 0.0550$) 稍低, 说明特征波段模型具有更好的预测效果, 是一种优选方法。

3 结论

对 5 种稻米的分析结果表明, 采用可见近红外光谱技术进行稻米品种与真伪的鉴别是可行的, 从而为稻米品种与真伪的快速、无损鉴别提供了一种新方法。比较分析结果表明, 利用特征波段所建立的模型比全波段模型具有更高的预测精度, 说明特征波段提取是进行模型优化的有效手段。

本研究采用了 FieldSpec® 3 地物光谱仪, 波段范围在 350~2500nm 范围内, 比一些研究^[2-8]所用的前一代产品 Handheld Field Spec (波段范围 325~1075nm) 有更宽的波段范围。而分析表明, 910~

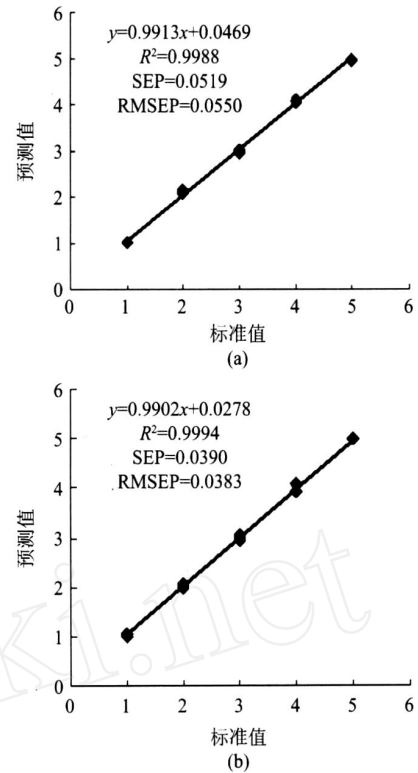


图 5 全波段模型 (a) 与特征波段模型 (b) 对未知样本预测值与标准值之间的关系

Fig 5 Standard value versus predicted value by whole wave band model (a) and characteristic wave band model (b) in validation set

1420nm、1940~2300nm 为稻米鉴别的主要特征波段, 说明在有机物的鉴别方面, FieldSpec® 3 比 Handheld Field Spec 可提供更丰富的分类信息。

REFERENCES

- [1] CHU Xiao-Li, YUAN Hong-Fu, LU Wan-Zhen. Present situation and prospect of near infrared spectrometer in China and abroad [J]. *Analytical Instrumentation* (褚小立, 袁洪福, 陆婉珍. 近红外光谱仪国内外现状与展望. 分析仪器), 2006 (2): 1—10.
- [2] LU Yan-De, LUO Ji, CHEN Xing-Miao. Analysis of soluble solid content in Nanfeng mandarin fruit with visible near infrared spectroscopy [J]. *J. Infrared Millim. Waves* (刘燕德, 罗吉, 陈兴苗. 可见近红外光谱的南丰蜜桔可溶性固形物含量定量分析. 红外与毫米波学报), 2008, 27 (2): 119—122.
- [3] HAN Liang-Liang, MAO Pei-Sheng, WANG Xin-Guo, et al. Study on vigour test of oat seeds with near infrared reflectance spectroscopy [J]. *J. Infrared Millim. Waves* (韩亮亮, 毛培胜, 王新国, 等. 近红外光谱技术在燕麦种子活力测定中的应用研究. 红外与毫米波学报), 2008, 27 (2): 86—90.
- [4] LU Xiu-hua, ZHANG Xue-Gong, SUN Su-qin. Auto-discrimination and characteristic wave band selection of Chinese traditional medicine origin based on near infrared spec-

(下转 391 页)

- 王有年,等. 桃果实品质评价因子的选择. *农业工程学报*, 2006, **22** (8): 235—239.
- [2] HAN Dong-Hai, WANG Jia-Hua. Review of nondestructive measurement of fruit quality by means of near infrared spectroscopy[J]. *Chinese Journal of Lasers* (韩东海, 王加华. 水果内部品质近红外光谱无损检测研究进展. *中国激光*), 2008, **35** (8): 1123—1131.
- [3] LU Yan-De, LUO Ji, CHEN Xing-Miao. Analysis of soluble solid content in nanfeng mandarin fruit with visible near infrared spectroscopy[J]. *J. Infrared Millim. Waves* (刘燕德, 罗吉, 陈新苗. 可见近红外光谱的南丰蜜桔可溶性固形物含量定量分析. *红外与毫米波学报*), 2008, **27** (2): 119—122.
- [4] NIE Li-Xing, WANG Gang-Li, LI Zhi-Meng, et al. Qualitative and quantitative analysis of Tongren Wuji Baifeng pills by near infrared spectroscopy[J]. *J. Infrared Millim. Waves* (聂黎行, 王钢力, 李志猛, 等. 近红外光谱法对同仁乌鸡白凤丸的定性和定量分析. *红外与毫米波学报*), 2008, **27** (3): 205—210.
- [5] Ying Yi-bin, Liu Yan-de. Nondestructive measurement of internal quality in pear using genetic algorithms and FT-NIR spectroscopy[J]. *Journal of Food Engineering*, 2008, **84**: 206—213.
- [6] Durand A, Devos O, Ruckebusch C, et al. Genetic algorithm optimization combined with partial least squares regression and mutual information variable selection procedures in near-infrared quantitative analysis of cotton viscose textiles[J]. *Analytica Chimica Acta*, 2007, **595**: 72—79.
- [7] Jiang J H, Berry R J, Ozaki Y, et al. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data [J]. *Anal. Chem.*, 2002, **74**: 3555—3565.
- [8] Nørgaard L, Saudland A, Wagner J, et al. Interval partial least squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy [J]. *Applied Spectroscopy*, 2000, **54**: 413—419.
- [9] Zou Xiao-bo, Zhao Jie-wen, Li Yan-xiao. Selection of the efficient wavelength regions in FT-NIR spectroscopy for determination of SSC of Fuji' apple based on BiPLS and FiPLS models [J]. *Vibrational Spectroscopy*, 2007, **44**: 220—227.
- [10] ZHU Xiang-Rong, LIN a, SHI Xin-Yuan, et al. Analysis of geniposide and total nitrogen content in Qingkailing injection intermediate by near infrared spectroscopy and synergy interval partial least squares[J]. *Chin. J. Chinese Universities* (朱向荣, 李娜, 史新元, 等. 近红外光谱与组合的间隔偏最小二乘法测定清开灵四混液中总氮和槲子苷的含量. *高等学校化学学报*), 2008, **29** (5): 906—911.
- (上接 356页)
- tra[J]. *China Science Bulletin* (刘沐华, 张学工, 孙素琴. 中药材产地的近红外光谱自动鉴别和特征谱段选择. *科学通报*), 2005, **50** (4): 393—398.
- [5] WU Yong-Jun, LI Wei, Xiang Bing-Ren, et al. Identification of traditional Chinese medicine Baizhi with near-infrared spectrum[J]. *Journal of Chinese Medicinal Materials* (吴拥军, 李伟, 相秉仁, 等. 近红外光谱技术用于白芷类中药的鉴定研究. *中药材*), 2001, **24** (1): 26—28.
- [6] LU Gou-Lin, CAI Jin-Na, LI Wei, et al. Near-Infrared spectroscopy technique used in the classification of the *Cnidium Monnieri* (L.) Cusson [J]. *Computer and Applied Chemistry* (刘国林, 蔡金娜, 李伟, 等. 近红外光谱技术在中药蛇床子分类中的应用. *计算机与应用化学*), 2000, **17** (2): 109—110.
- [7] Delwiche S R, Bean M M, Miller R E, et al. Apparent amylose content of milled rice by near-infrared reflectance spectrophotometry[J]. *Cereal Chemistry*, 1995, **72**: 182—187.
- [8] Bao J S, Cai Y Z, Corke H. Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy [J]. *Food Science*, 2001, **66**: 936—939.
- [9] Wu J G, Shi C H, Zhang X M. Estimating the amino acid composition in the milled rice powder by near-infrared reflectance spectroscopy[J]. *Field Crops Research*, 2002, **75**: 1—7.
- [10] WANG Hai-Lian, WAN Xiang-Yuan, HU Pei-Song, et al. Quantitative analysis of fat content in brown rice by near infrared spectroscopy (NIRS) technique[J]. *Scientia Agricultura Sinica* (王海莲, 万向元, 胡培松, 等. 稻米脂肪含量近红外光谱分析技术研究. *中国农业科学*), 2005, **38** (8): 1540—1546.
- [11] BI Jing-Cui, ZHANG Wen-Wei, XIAO Ying-Hui, et al. Analysis for protein content in rice by near infrared reflectance spectroscopy (NIRS) technique[J]. *Acta Agronomica Sinica* (毕京翠, 张文伟, 肖应辉, 等. 应用近红外光谱技术分析稻米蛋白质含量. *作物学报*), 2006, **32** (15): 709—715.
- [12] SHAO Yong-Ni, CAO Fang, HE Yong. Discrimination years of rough rice by using visible/near infrared spectroscopy based on independent component analysis and BP neural network [J]. *J. Infrared Millim. Waves* (邵咏妮, 曹芳, 何勇. 基于独立组分析法和BP神经网络的可见/近红外光谱稻谷年份的鉴别. *红外与毫米波学报*), 2007, **26** (6): 433—436.
- [13] Haykin S. *Neural network—a comprehensive foundation* [M]. New York, USA: Macmillan College Publishing Company, 1994, 1—44.
- [14] QI Xiao-Ming, ZHANG Lu-Da, DU Xiao-Lin, et al. Quantitative analysis using NIR by building PLS-BP model [J]. *Spectroscopy and Spectral Analysis* (齐小明, 张录达, 杜晓林, 等. PLS-BP法近红外光谱定量分析研究. *光谱学与光谱分析*), 2003, **23** (5): 870—872.
- [15] Workman J. *Handbook of Organic Compounds: NIR, IR, Raman, and UV-VIS Spectra Featuring Polymers and Surfactants* [M]. San Diego, USA: Academic Press, 2001: 77—197.
- [16] Sasic S, Ozaki Y. Short-wave near-infrared spectroscopy of biological fluids. 1. In Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment[J]. *Anal. Chem.*, 2001, **73**: 64—71.