

文章编号:1001-5078(2009)04-0407-04

· 红外技术 ·

杂交稻种品系与真伪的可见-近红外光谱鉴别

梁亮¹, 杨敏华¹, 刘志霄², 宋华湘¹

(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)

摘要:提出了一种基于可见-近红外光谱技术快速、无损鉴别杂交稻种品系与真伪的新方法。采集了5种稻种的光谱数据,各获取32个样本,随机分成训练集(125份)和检验集(35份)。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个主成分作为新变量。分别用模糊模式识别、BP-神经网络、Fisher多类线性判别以及Bayes多类逐步判别四种方法进行分析。对35个未知样的预测结果说明可见-近红外技术进行杂交稻种品系与真伪的快速、无损鉴别是可行的,且PCA结合Bayes多类逐步判别是一种优选方法。

关键词:可见-近红外光谱;稻种;鉴别;无损

中图分类号:O433.5 **文献标识码:**A

Discrimination of Lines and Authenticity of Hybrid Rice Seed with Visible-near Infrared Spectra

LIANG Liang^{1,2}, YANG Min-hua¹, LIU Zhi-xiao^{1,2}, SONG Hua-xiang¹

(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)

Abstract: A fast and non-invasive method based on visible-near infrared reflectance spectroscopy was put forward for discriminating lines and authenticity of hybrid rice seed. Five different varieties of rice seed were analyzed using a FieldSpec®3 visible-near infrared spectrometer, and 32 samples were used for each variety of rice. All samples were divided randomly into two groups, one group with 125 samples used as calibrated set, and another with 35 samples used as validated set. The samples data were pretreated with the methods of S. Golay smoothing and standard normal variable (SNV), and then analyzed by principal component analysis (PCA). The anterior 9 principal components computed by PCA were used as the new variables, and analysed by Fuzzy pattern recognition, back-propagation artificial neural network (ANN-BP), Fisher multi-types linear discriminant and Bayes multi-types stepwise discriminant. Then, the 35 unknown samples in the validated set were predicted. Therefore, the feasibility of discriminating the lines and authenticity of hybrid rice seed in rapid and non-invasive way by visible-near spectra technology was proved, and PCA combined with Bayes multi-types stepwise discriminant was confirmed as a preference method.

Key words: visual-near infrared spectra; rice seed; discrimination; noninvasive

1 引言

目前,杂交稻种鉴别主要有形态鉴定法、幼苗鉴定法、田间小区种植鉴定法以及蛋白质与DNA指纹鉴定法^[1-2]。其中形态鉴定法、幼苗鉴定法主要依靠经验,主观性太强;田间小区种植鉴定法虽较可

靠,但则耗时太长;而蛋白质与DNA指纹鉴定法则

基金项目:国家自然科学基金项目(No. 30570279);中南大学研究生创新项目(No. 1343-74334000022)资助。

作者简介:梁亮(1981-),男,中南大学信息物理工程学院博士研究生。E-mail: liangliang198119@163.com

收稿日期:2008-10-05

耗费较大,加之分析时用样太少,容易产生因取样误差而导致总体结果不可靠的问题。且染色法、发芽率测定以及蛋白质与 DNA 指纹鉴定法均为损伤性检测,需要破坏样品,因而在大批量取样检测中的应用受到很大限制。因此,探索一种快速、准确、无损鉴别杂交稻种品系与真伪的方法显得非常必要。

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,已在石油化工、探矿、制药以及纺织等领域获得了广泛的应用^[3]。近年来,可见-近红外光谱结合模式识别技术,进一步在茶叶、烟草等农产品^[4-5],蛇床子、丹参等中药材^[6-8]产地与种类鉴别获得了成功的应用,而对不同品牌的奶粉、咖啡、可乐的分类鉴定也取得了满意的结果^[9-11]。在水稻的相关研究中,近红外光谱技术也已应用于稻米直链淀粉、蛋白质、脂肪与氨基酸含量的分析以及稻谷年份的检测^[12-16],但是否可利用这一技术进行杂交稻种品系与真伪的鉴别还有待进一步探讨。因此,本文采用可见-近红外反射光谱技术对杂交稻种进行分析,为杂交稻种品系与真伪的快速、无损鉴别提供新的方法。

2 实验部分

2.1 实验仪器

FieldSpec® 3 地物光谱仪一台,直径 115mm 玻璃皿若干。FieldSpec® 3 地物光谱仪为美国 ASD (Analytical Spectral Device) 公司新产品,波长范围 350 ~ 2500nm,采用间隔 1.4nm,光谱分辨率 3nm。

2.2 实验方法

2.2.1 样本来源与数据采集

于湖南长沙种子公司获取宜香 725、绵香 3 优 725 米、II 优 86、金优 191 杂交稻种 4 种,伪劣杂交稻种 1 种。每一稻种采集样本 32 份,合计 160 份。随机抽取 125 份(每种 25 份)作为训练集,其余 35 份(每种 7 份)作为预测集。稻种样本盛放于玻璃皿中进行光谱扫描。光谱仪探头置于样本正上方,探头视场角 20°,下部距样本 100mm;光源与水平面保持 45°角,距样本 350mm。每一样本重复测量 30 次取其均值。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 9.50 中进行处理。

2.2.2 光谱预处理

光谱经 S. Golay 平滑后,采用标准归一化 (SNV) 进行处理,以消除仪器噪声、基线漂移、样品不均匀等因素给光谱带来的影响。由于光谱在 380nm 以前与 2400nm 以后噪声较大,因此本研究

选用 380 ~ 2400nm 范围的反射光谱波段作为有效光谱数据进行分析。

2.2.3 光谱数据降维

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预测精度。为了消除噪声,提高信噪比,本研究采用主成分分析 (PCA) 以实现光谱数据的降维。数据经 PCA 降维后,既能有效除噪,又能最大限度地保留原始数据的结构特征,从而使降维所得的新变量并不丢失有价值的信息^[17]。

2.2.4 分析方法选取

为获得适当的分析方法,本研究分别采用模糊模式识别、BP-神经网络、Fisher 多类线性判别以及 Bayes 多类逐步判别四种方法对样本进行分析,经比较后选取较优方法。

3 结果与讨论

3.1 杂交稻种的可见-近红外漫反射光谱

图 1 为 5 种稻种可见-近红外漫反射光谱曲线。由图可知,稻种在 350 ~ 400nm 反射率较低,400nm 后迅速上升,于 1110nm 附近达到峰值,此后开始逐步下降,在 2500nm 降至 10% 左右。同时,由于淀粉、纤维素与水的吸收^[18],稻种光谱反射率在 990nm,1200nm,1460nm,1925nm 以及 2110nm 附近呈波谷形态。

由于杂交稻种间的亲缘关系很近,组成成分非常相似,因此,图中虽可看出不同稻种反射光谱的波形稍有差异,但单凭肉眼无法区分。下一步分析时,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。

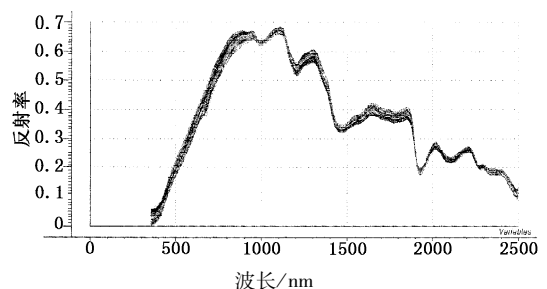


图 1 5 种稻种样本的可见-近红外反射光谱

3.2 PCA 分析与特征信息提取

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(如图 2 所示),以表征样本在该三维空间中的分布。由于 PC_1, PC_2, PC_3 对光谱矩阵的方差贡献率分别为 73.86%, 13.25% 与 5.04%,累加达 96.14%,因此,样本点在三维空间

的分布可大体反应样本在超维空间中的分布特征,表征出不同稻种的聚类结果。从图2中可看出,各种稻种有一定的聚类趋势,可各稻种样本分布较为分散且存在交叉,欲取得精确的结果,还需利用更多的主成分分数以要建立鉴别能力更强的模型。

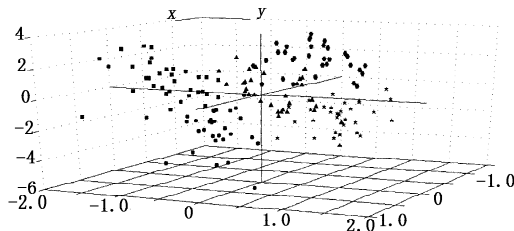


图2 5种稻种前3个主成分的得分聚类

●宜香725; ■绵香3优725; ▲II优86; ◆金优191; ★伪劣杂交稻种

根据统计学的原理,在下一步建模的过程中,如果所选取的主成分过少,将会损失某些变量的信息,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,则会将光谱中的噪音信息引入模型,产生过拟合现象而导致模型预测的准确率下降。本研究中通过交互验证确定最佳主成分数,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。将光谱数据进行PCA分析后,所得前12个主成分的累积可信度如图3所示。其中前9个主成分的累积可信度达99.82%,已包含了光谱数据绝大部分的特征信息。

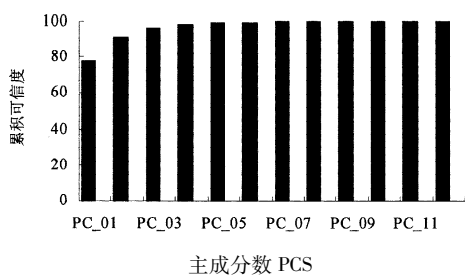


图3 PCA前12个主成分的累积可信度

3.3 四种方法鉴别结果的比较分析

将宜香725、绵香3优725米、II优86、金优191以及伪劣杂交稻种分别赋值为1,2,3,4,5,利用训练集中的125个样本,以PCA降维得到的前9主成分作为输入变量,在DPS 9.50中分别利用模糊模式识别、BP神经网络、Fisher多类线性判别以及Bayes多类逐步判别对样本进行分析。

训练过程中,BP神经网络各层间采用Sigmoid激励函数,网络结构设置为9-6-5三层,迭代次数设置为1000次;Bayes多类逐步判别的临界值F取为1.74;模糊模式识别与Fisher多类线性判别则直

接进行分析。结果表明,四种方法对训练集样本的拟合结果很好,正确率均为100%。

在对未知样本判定分析时,神经网络模型给出定量预测值,并以此得出类别归属;模糊模式识别直接通过模糊函数判定未知样本类别归属及对所类别的最大贴近度;Fisher多类线性判别与Bayes多类逐步判别则给出判别结果与概率。四种方法对35个未知样本的判别结果如表1所示。其中模糊模式识别有7样本误判,判定的准确率最低,为80%;神经网络模型有6个样本误判,预测的准确率稍高,为82.9%;Fisher多类线性判别有5个样本误判,判别正确率为85.7%;Bayes多类逐步判别对未知样本判定的准确率最高,达到了100%。

相关文献表明,在对有机物进行分类鉴别时,由于其可见-近红外光谱间存在严重的共线性关系,因此,BP-神经网络模型因其强大的非线性建模能力而可取得较优的结果^[4-11]。但上述比较分析表明,Bayes多类逐步判别的准确率达100%,比BP神经网络模型具有更好的鉴别效果。其原因是杂交稻种的可见-近红外光谱间虽然存在共线性关系,但本研究在建模前已将其进行了PCA降维。由于PCA具有排除众多信息共存中相互重叠的信息,将降维后变量变为原变量的线性组合这一功能^[17],故降维后的新变量相互正交,互不相关,因此BP-神经网络算法在本研究中并无明显优势。根据分析结果,利用可见-近红外光谱进行杂交稻种的品系与真伪鉴别,PCA结合Bayes多类逐步判别是一种优选方法。

表1 四种方法对35个未知样本的预测结果

样本号	标准值	模糊模式识别	BP-神经网络模型	Fisher多类线性判别	Bayes多类逐步判别
		类别归属(贴近度)	类别归属(预测值)	类别归属(概率)	类别归属(后验概率)
(1)	1	2(0.9044)	1(1.0339)	1(0.9790)	1(1.0000)
(2)	1	1(0.9949)	1(1.0249)	1(0.5790)	1(1.0000)
(3)	1	1(1.0000)	1(1.0018)	1(0.9471)	1(1.0000)
(4)	1	1(0.2393)	1(1.0038)	1(0.9994)	1(1.0000)
(5)	1	1(1.0000)	1(1.0092)	2(0.9959)	1(0.9977)
(6)	1	1(0.9361)	1(1.0407)	3(0.5097)	1(1.0000)
(7)	1	1(0.9463)	1(1.0913)	1(0.9999)	1(1.0000)
(8)	2	2(1.0000)	2(2.3955)	2(0.9998)	2(1.0000)
(9)	2	2(1.0000)	2(1.9382)	2(0.9092)	2(1.0000)

续表

(10)	2	2(0.9908)	2(1.9507)	2(0.8617)	2(1.0000)
(11)	2	2(1.0000)	2(1.9734)	2(0.9325)	2(0.9991)
(12)	2	2(1.0000)	2(2.2990)	2(0.9265)	2(0.9977)
(13)	2	2(0.9981)	2(2.3911)	2(0.9696)	2(0.9827)
(14)	2	2(1.0000)	2(1.9560)	2(1.0000)	2(1.0000)
(15)	3	1(0.8280)	1(1.1221)	3(0.9818)	3(1.0000)
(16)	3	3(0.7618)	5(4.6156)	5(0.6737)	3(1.0000)
(17)	3	3(0.8654)	1(1.1840)	2(0.5513)	3(1.0000)
(18)	3	3(1.0000)	3(3.0274)	3(0.8419)	3(1.0000)
(19)	3	2(1.0000)	4(4.0204)	3(0.9975)	3(1.0000)
(20)	3	3(1.0000)	3(3.0037)	3(0.9963)	3(1.0000)
(21)	3	2(0.4670)	2(2.3958)	3(0.9037)	3(1.0000)
(22)	4	4(0.5868)	4(3.6899)	4(0.9999)	4(1.0000)
(23)	4	1(0.9780)	4(3.6762)	4(0.9973)	4(1.0000)
(24)	4	4(0.7486)	4(4.0480)	4(1.0000)	4(1.0000)
(25)	4	4(0.9250)	4(4.0363)	4(0.9999)	4(1.0000)
(26)	4	4(0.8716)	1(1.3457)	4(0.8937)	4(1.0000)
(27)	4	3(0.8272)	4(4.0342)	4(1.0000)	4(1.0000)
(28)	4	4(0.9382)	5(4.8126)	4(0.9850)	4(1.0000)
(29)	5	5(1.0000)	5(4.9621)	5(1.0000)	5(1.0000)
(30)	5	5(0.9784)	5(4.9304)	5(1.0000)	5(1.0000)
(31)	5	5(1.0000)	5(4.9864)	5(1.0000)	5(1.0000)
(32)	5	1(0.9803)	5(4.9568)	5(1.0000)	5(1.0000)
(33)	5	5(0.9232)	5(4.6574)	3(0.4686)	5(0.9987)
(34)	5	5(0.8375)	5(4.9922)	5(1.0000)	5(1.0000)
(35)	5	5(0.9096)	5(4.9694)	5(1.0000)	5(1.0000)

注:(1) - (7): 宜香 725; (8) - (14): 绵香 3 优 725; (15) - (21): II 优 86; (22) - (28): 金优 191; (29) - (35): 伪劣杂交稻种

4 结 论

以 FieldSpec[®] 3 地物光谱仪采集 5 种稻种的可见 - 近红外漫反射光谱, 数据经 PCA 降维后, 分别以模糊模式识别、BP 神经网络、Fisher 多类线性判别以及 Bayes 多类逐步判别进行分析, 得出以下结论: 采用可见 - 近红外技术进行稻种鉴别是可行的, 可作为杂交稻种品系与真伪快速、无损鉴别的新方法, 且光谱数据处理以 PCA 结合 Bayes 多类逐步判别的效果较好。

参考文献:

- [1] 翁醒华, 周姚宇, 唐榛. 水稻三系及其杂种 F₁ 的叶绿体同工酶研究[J]. 科技通报, 1995, 11(2): 102 - 106.
- [2] 段世华, 毛加宁, 朱英国. 用微卫星 DNA 标记对我国杂交水稻主要恢复系遗传差异的检测分析[J]. 遗传学报, 2002, 29(3): 250 - 254.

- [3] 褚小立, 袁洪福, 陆婉珍. 近年来我国近红外光谱分析技术的研究与应用进展[J]. 分析仪器, 2006, (2): 1 - 10.
- [4] 李晓丽, 何勇, 裘正军. 一种基于可见 - 近红外光谱快速鉴别茶叶品种的新方法[J]. 光谱学与光谱分析, 2007, 27(2): 279 - 282.
- [5] 段焰青, 者为, 李青青, 等. 用光谱主成分分析选择烟草近红外模型校正集样品[J]. 激光与红外, 2008, 38(7): 663 - 665.
- [6] 张晓慧, 刘建学. 近红外光谱技术鉴别连翘产地[J]. 激光与红外, 2008, 38(4): 342 - 344.
- [7] 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[J]. 科学通报, 2005, 50(4): 393 - 398.
- [8] 虞科, 程翼宇. 一种基于最小二乘支持向量机算法的近红外光谱判别分析方法[J]. 分析化学, 2006, 34(4): 561 - 564.
- [9] 王艳艳, 何勇, 邵咏妮, 等. 基于可见 - 近红外光谱的咖啡品牌鉴别研究[J]. 光谱学与光谱分析, 2007, 27(4): 702.
- [10] 黄敏, 何勇, 岑海燕, 等. 应用可见 - 近红外光谱技术快速无损鉴别婴幼儿奶粉品种[J]. 光谱学与光谱分析, 2007, 27(5): 916 - 919.
- [11] 裘正军, 陆江锋, 毛静渊, 等. 基于可见 - 近红外光谱的可乐品牌鉴别方法研究[J]. 光谱学与光谱分析, 2007, 27(8): 1543 - 1546.
- [12] Bao J-S, Cai Y-Z, Corke H. Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy[J]. J Food Sci, 2001, 66(7): 937 - 939.
- [13] Wu J G, Shi C H, Zhang X M, et al. Estimating the amino acid composition in the milled rice powder by near-infrared reflectance spectroscopy[J]. Field Crops Res, 2002, 75(1): 1 - 7.
- [14] 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[J]. 中国农业科学, 2005, 38(8): 1540 - 1546.
- [15] 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[J]. 作物学报, 2006, 32(15): 709 - 715.
- [16] 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[J]. 红外与毫米波学报, 2007, 26(6): 433 - 436.
- [17] 许禄, 邵学广. 化学计量学方法[M]. 北京: 科学出版社, 2004: 130 - 138.
- [18] Workman J. Handbook of organic compounds: NIR, IR, Raman, and UV-VIS spectra featuring polymers and surfactants[M]. San Diego: Academic Press, 2001: 77 - 197.