第30卷,第5期
2010年5月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 30, No. 5, pp1248-1251
May, 2010

基于近红外光谱的商品玉米品种快速鉴别方法

邬文锦[1]，王红武[2]，陈绍江[2]，郭婷婷[3]，王守觉[3]，苏 谦[1]，孙 明[1]，安 冬[1*]

1. 中国农业大学信息与电气工程学院，北京 100083
2. 中国农业大学国家玉米改良中心，北京 100094
3. 中国科学院半导体研究所，北京 100083

摘 要 现有的玉米种子品种鉴别方法检测时间长，费用高，不易大批量快速鉴别。提出了一种基于近红外光谱数据快速鉴别商品玉米品种的新方法。先使用傅里叶变换近红外光谱仪获得从4 000到12 000 cm^{-1}波段范围的37个商品玉米品种籽粒的漫反射光谱数据。对原始光谱进行矢量归一化预处理以消除噪声干扰，为了找到玉米品种籽粒的光谱特征波段，提出一种基于标准差的方法，进而对寻找到的玉米籽粒特征波段光谱做主成分分析(PCA)，取能反映玉米品种99.98%光谱信息的前10个主成分。最后使用仿生模式识别(BPR)方法建立了37个玉米品种鉴别模型，对于每个品种的25个样本，随机挑选15个样本作为训练样本，其余10个样本作为第一测试集，其他品种共900个样本作为第二测试集。该鉴别模型对于37个玉米品种的平均正确识别率为94.3%。该方法的进一步研究有利于建立以近红外光谱为基础的物理指纹品种鉴别技术。

关键词 近红外光谱；仿生模式识别；玉米商品籽粒；品种鉴别

中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2010)05-1248-04

引 言

玉米是一种重要的农作物，玉米种子品种鉴别是目前农业生产、作物育种和种子检验的重要问题之一。现有的种子品种鉴别常用方法有形态学方法、荧光扫描鉴定法、化学鉴定法和电泳鉴定法等。形态学方法所需鉴别时间长，且精度不高；荧光扫描鉴定法、化学鉴定法和DNA分子标记鉴定法鉴别精度高，但所需时间长，且鉴别成本较高，过程烦琐[1,2]。

近红外光谱区的波长范围为780～2 500 nm，通过近红外光谱，可以检测样品中有机分子含氢基团的特征信息[3-5]。目前近红外光谱在农产品检测中的应用已经相当成熟[6-12]，在玉米中的应用也有很多报道。如芮玉奎等报道了近红外光谱在转基因玉米检测识别中的应用；李伟等报道了近红外光谱在4个玉米品种鉴别中的应用[13]；Mark等报道了近红外光谱在玉米种子成分检测中的应用[14]等。但现有报道中鉴定的品种种类较少，绝大多数在10个品种以内，大规模的品种鉴定还未见报道，因此探索怎样利用近红外光谱数据进行大规模的品种鉴定具有重要的理论和现实意义。

本文首先使用傅里叶变换近红外光谱仪获得37个商品玉米品种籽粒的漫反射光谱，对原始光谱进行矢量归一化预处理，利用一种基于标准差的方法寻找玉米籽粒光谱特征波段，进而对寻找到的玉米籽粒特征波段光谱做主成分分析(PCA)，试图提出一种基于近红外光谱数据快速鉴别玉米种子品种的新方法，为近红外光谱技术在玉米品种中的定性分析做一些理论研究。

1 材料与方法

1.1 仪器设备

德国BRUKER公司的VECTOR22/N型傅里叶变换漫反射近红外光谱仪，谱区范围4 000～12 000 cm^{-1}，扫描次数为64次，分辨率为8 cm^{-1}。

1.2 样品来源与光谱获取

所有试验商品玉米种子由国家科技支撑计划高产优质多抗玉米项目组提供。玉米籽粒样品均为北京种植的同年份的玉米品种(见表1)，田间种植每品种4行区，设3次重复。

收稿日期：2009-06-22，**修订日期**：2009-09-26

基金项目：国家自然科学基金项目(60805011)，教育部博士点基金项目(200800191028)，国家科技支撑计划项目(2006BAD01A03)和现代玉米产业技术体系项目资助

作者简介：邬文锦，1984年生，中国农业大学信息与电气工程学院研究生 e-mail：wanna1106@163.com

*通讯联系人 e-mail：andong@semi.ac.cn

Table 1　37 varieties of the corn samples

编号	品种名称	编号	品种名称	编号	品种名称
1	中试 703	14	科育 4501	27	DH2826
2	中试 705	15	豫单 2268	28	耕玉 801
3	中试 909	16	LD686	29	LD8074
4	ND603	17	弘大 8 号	30	LD8078
5	ND8807	18	07H130366	31	LD3070
6	MC3558	19	LD688	32	聊玉 0815
7	豫单 811	20	LD9067	33	中玉 688
8	CA558	21	LD9059	34	先行 2 号
9	MC716	22	LD6076	35	中试 4340
10	豫单 2773	23	LD6077	36	郑单 958
11	豫单 606	24	DH8601	37	五岳 5508
12	豫单 2269	25	聊玉 0812		
13	郑单 2091	26	LD689		

成熟后收获玉米籽粒，风干脱水保存。小区选取中间两行籽粒扫描光谱，样品扫描前均在 40 ℃烘干 72 h。样品盛放在统一尺寸的样品杯中。放置时，最底层玉米籽粒一半胚乳向上，一半胚乳向下。对同一品种的玉米籽粒多次取样，共测量 25 次。

1.3　原始光谱矢量归一化预处理

为降低同一品种玉米籽粒若干次测量之间的差别，本文采用矢量归一化方法对原始光谱进行预处理。具体计算步骤为：(1) 对一条原始光谱，计算其平均吸光度值；(2) 用原始光谱值减去平均吸光度值，得到处理后的光谱值；(3) 计算处理后的光谱值的平方和，再开平方根，设这个值为 m；(4) 将处理后的光谱值除以 m，完成一条光谱的矢量归一化。

1.4　特征光谱波段选择

直接利用玉米籽粒的全波段近红外光谱数据(4 000～12 000 cm^{-1})进行品种鉴定，存在两个问题：(1) 全波段光谱数据量较大，计算速度较慢；(2) 全波段光谱中有相当多的与品种鉴定关系不大的数据，这些数据会对鉴定造成较大干扰，影响鉴定精度。所谓特征光谱波段就是在这些波段内，不同品种玉米籽粒光谱之间的差异较大，相同品种玉米籽粒光谱之间的差异较小。本文在文献[15]的基础上，提出了一种基于标准差的方法寻找玉米籽粒特征光谱波段，具体算法为：对全波段近红外光谱中的每个波数(cm^{-1})计算参数 $Q_s = \frac{S_{ter}}{S_{tra}}$，其中 S_{ter}(不同品种间离散程度) 为每类玉米籽粒在此波数上吸光度平均值的标准差；S_{tra} (同一品种内离散程度)为各类玉米籽粒在此波数上吸光度标准差的平均值。本文选取 $Q_s > 2$ 的波数区域为特征光谱波段。

1.5　主成分分析(PCA)

PCA 是常用的一种降维方法，它根据方差最大原则对表征原始数据集的多个自变量进行线性组合，从而用数量较少的新的综合自变量替代原始自变量，达到降维的目的。玉米籽粒特征波段光谱数据经 PCA 算法处理后，前 10 个主成分的累积贡献率达到 99.98 %，故本文选取前 10 个主成分。

1.6　仿生模式识别(BPR)

BPR 是王守觉提出的一种模式识别理论[15,16]，其原理完全不同于传统模式识别。传统模式识别把不同类样本在特征空间中的最佳划分作为目标，最具代表性的就是支持向量机(SVM)；而 BPR 则以同一类样本在特征空间中的最佳覆盖为目标。对于未经训练的类别样本，传统模式识别会将其错误的归入某一经过训练的类别；而仿生模式识别则会正确拒识该样本，不会将其归入任何一个经过训练的类别，这点在进行玉米品种鉴定时尤为重要。

BPR 首先分析同类样本在特征空间中的分布情况，然后使用多权值神经网络构造出复杂的几何形体对其进行覆盖。本文采用一个三权值神经元为基本覆盖单元，用多个三权值神经元的组合来实现不同品种玉米籽粒的神经网络覆盖模型。

2　结果与讨论[17,18]

2.1　特征光谱区域的选择

图 1 是使用基于标准差的方法寻找 37 个商品玉米品种特征光谱波段的结果。横坐标是波数(4 000～12 000 cm^{-1})，纵坐标是 Q_s 值。从图 1 中可以看出，波数在4 952.3～5 280 和 9 580.6～10 016.5 cm^{-1}时 $Q_s > 2$。但是当波数在 9 580.6～10 016.5 cm^{-1}时，原始光谱噪声较大，且 Q_s 值大于 2 不多，因此本文没有选择此波段为特征光谱波段。当波数在 4 952.3～5 280 cm^{-1}时，原始光谱有很高的吸收峰，且在此范围内的光谱能够反映玉米籽粒中蛋白质、淀粉等与玉米品种有关的成分的信息，本文选择 4 952.3～5 280 cm^{-1}作为特征光谱波段。

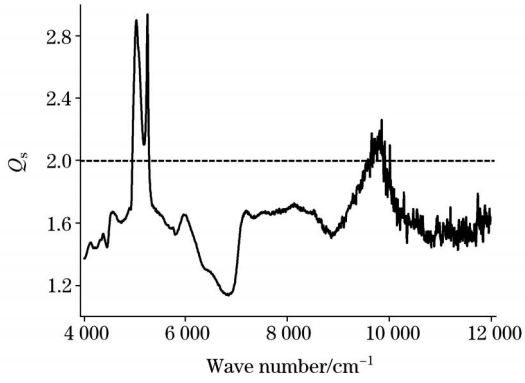

Fig. 1　Relationship of wavenumber and Q_s

2.2　基于仿生模式识别的玉米品种鉴别模型

每个品种的玉米籽粒都经过 25 次取样、测量，得到 25 条原始光谱，37 个玉米品种共得到 925 条原始光谱。对所有原始光谱进行矢量归一化预处理后，使用基于标准差的方法确定这 37 个玉米品种的特征光谱波段为 4 952.3～5 280.1 cm^{-1}，进而对特征光谱波段数据做 PCA，取前 10 个主成分作为建立及测试玉米品种鉴别模型使用的样本(每个样本是一个 10 维向量，共 925 个样本)。

每个品种的玉米籽粒都有 25 个样本，从中随机挑选 15 个作为训练集，其余 10 个作为第一测试集，其他品种玉米籽粒的样本(36 类共 900 个)作为第二测试集。第一测试集用于

检测每个品种的鉴别模型对于本品种样本的正确识别率，第二测试集用于检测每个品种的鉴别模型对于其他品种样本的正确拒识率。

本文应用仿生模式识别原理，使用一个三权值神经元为基本覆盖单元，用 13 个三权值神经元的组合来实现不同品种玉米籽粒的神经网络覆盖模型，实验结果见表 2。

由以上实验结果说明：(1) 仿生模式识别能够有效地拒识非本类样本，克服了传统模式识别的致命弱点；(2) 对于小训练样本集的多分类问题，仿生模式识别具有很大优越性。

Table 2 Recognition results of 37 classification on the two testing sets based on BPR

品种名称	正确识别率/ %	正确拒识率/ %	品种名称	正确识别率/ %	正确拒识率/ %	品种名称	正确识别率/ %	正确拒识率/ %
中试 703	100	86.8	科育 4501	100	90.1	DH2826	100	87.0
中试 705	90	93.9	豫单 2268	100	94.1	耕玉 801	100	86.9
中试 909	100	87.9	LD686	100	91.4	LD8074	100	84.2
ND603	90	89.8	弘大 8 号	100	97.3	LD8078	90	86.9
ND8807	100	85.9	07H130366	90	92.4	LD3070	90	85.4
MC3558	100	82.1	LD688	90	80.1	聊玉 0815	80	80.0
豫单 811	80	83.7	LD9067	90	74.6	中玉 688	100	96.0
CA558	100	95.3	LD9059	100	86.6	先行 2 号	100	93.9
MC716	100	92.6	LD6076	80	83.1	中试 4340	100	88.1
豫单 2773	90	81.0	LD6077	100	89.9	郑单 958	100	88.9
豫单 606	80	82.1	DH8601	90	95.6	五岳 5508	90	81.8
豫单 2269	90	96.7	聊玉 0812	90	90.9	平均	94.3	87.9
郑单 2091	90	87.0	LD689	100	82.6			

3 结 论

本文提出了一种基于近红外光谱的玉米品种鉴别方法。先使用傅里叶变换近红外光谱仪获得 37 个玉米品种的籽粒漫反射光谱。对原始光谱进行矢量归一化预处理后，提出一种基于标准差的方法寻找玉米籽粒光谱特征波段，进而对寻找到的玉米籽粒特征波段光谱做主成分分析(PCA)，取前 10 个主成分。最后使用仿生模式识别(BPR)方法建立了 37 个玉米品种鉴别模型，该鉴别模型对于 37 个商品玉米品种籽粒的平均正确识别率为 94.3%，表明商品籽粒虽然有一定的变异性，但其对于多种类的品种鉴别仍然具有较高的准确度。

矢量归一化预处理能够降低同品种玉米籽粒多次测量间的光谱差异。基于标准差的特征光谱波段选择方法能够选择出不同品种玉米籽粒光谱之间差异较大，而相同品种玉米籽粒光谱之间差异较小的特征光谱区域，从而减少无用信息的干扰，提高模型的鉴别精度。主成分分析可降低数据维数，提高鉴别速度。仿生模式识别对于小样本、多分类问题具有特殊的优势，且能对未经训练类别的样本有效地拒识，对品种鉴定较为适合。

目前，利用近红外光谱数据对玉米中各种成分进行定量分析已非常普遍，但对玉米品种进行定性的品种鉴定还不多见，且现有报道中鉴定的品种种类较少，绝大多数在 10 个品种以内，大规模的品种鉴定还未见报道。该方法的提出对利用近红外光谱在玉米品种鉴别上的规模化应用具有重要参考价值。

参 考 文 献

[1] LI Shang-yu, CHEN Yang, WANG Chun-yan, et al(李尚禹, 陈 阳, 王春艳, 等). Journal of Molecular Science(分子科学学报). 2007, 23(3): 220.

[2] DING Nian-ya, LI Wei, FENG Xin-wei, et al(丁念亚, 黎 薇, 冯昕韡, 等). Computers and Applied Chemistry(计算机与应用化学), 2008, 25(4): 499.

[3] ZHAO Jie-wen, HU Huai-ping, ZOU Xiao-bo(赵杰文, 呼怀平, 邹小波). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2007, 23(4): 149.

[4] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). Foundation and Application of Near-Infrared Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.

[5] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍, 袁洪福, 徐广通, 等). Modern Near Infrared Spectroscopy Analytical Technology (Second Edition)(现代近红外光谱分析技术, 第 2 版). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2007.

[6] HAN Liang-liang, MAO Pei-sheng, WANG Xin-guo, et al(韩亮亮, 毛培胜, 王新国, 等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2008, 27(2): 86.

[7] HUANG Min, HE Yong, HUANG Ling-xia, et al(黄 敏, 何 勇, 黄凌霞, 等). Journal of Infrared and Millimeter Waves(红外与毫

米波学报), 2006, 25(5): 342.
[8] WANG Tie-gu, LIU Xin-xiang, KU Li-xia, et al(王铁固, 刘新香, 库丽霞, 等). Journal of Maize Sciences(玉米科学), 2008, 16(3): 57.
[9] SUN Guo-ming, LIU Guo-lin(孙国明, 刘国林). China Journal of Chinese Materia Medica(中国中药杂志), 2006, 31(23): 1996.
[10] HAO Yong, CAI Wen-sheng, SHAO Xue-guang(郝 勇, 蔡文生, 邵学广). Chemical Journal of Chinese Universities(高等学校化学学报), 2009, 30: 28.
[11] FANG Li-min, LIN Min(方利民, 林 敏). Chinese Journal of Analytical Chemistry(分析化学), 2008, 36(6): 815.
[12] ZHANG Hui(张 卉). Chinese Journal of Spectroscopy Laboratory(光谱实验室), 2007, 24(3): 380.
[13] CHEN Jian, CHEN Xiao, LI Wei, et al(陈 建, 陈 晓, 李 伟, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2008, 28(8): 1806.
[14] Tesfaye M Baye, Tom C Pearson, Mark A Settles. Journal of Cereal Science. 2006, 43: 236.
[15] WANG Shou-jue(王守觉). Acta Electronica Sinica(电子学报), 2002, 30(10): 1417.
[16] WANG Shou-jue, WANG Bai-nan(王守觉, 王柏南). Acta Electronica Sinica(电子学报), 2002, 30(1): 1.
[17] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Progress in Chemistry(化学进展), 2004, 16(4): 528.
[18] CAO Yu, ZHAO Xing-tao(曹 宇, 赵星涛). Acta Electronica Sinica(电子学报), 2004, 32(10): 1671.

Fast Discrimination of Commerical Corn Varieties Based on Near Infrared Spectra

WU Wen-jin[1], WANG Hong-wu[2], CHEN Shao-jiang[2], GUO Ting-ting[3], WANG Shou-jue[3], SU Qian[1], SUN Ming[1], AN Dong[1]*

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
2. National Maize Improvement Center of China, China Agricultural University, Beijing 100094, China
3. Institute of Semiconductor, Chinese Academy of Sciences, Beijing 100083, China

Abstract The existing methods for the discrimination of varieties of commodity corn seed are unable to process batch data and speed up identification, and very time consuming and costly. The present paper developed a new approach to the fast discrimination of varieties of commodity corn by means of near infrared spectral data. Firstly, the experiment obtained spectral data of 37 varieties of commodity corn seed with the Fourier transform near infrared spectrometer in the wavenumber range from 4 000 to 12 000 cm^{-1}. Secondly, the original data were pretreated using statistics method of normalization in order to eliminate noise and improve the efficiency of models. Thirdly, a new way based on sample standard deviation was used to select the characteristic spectral regions, and it can search very different wavenumbers among all wavenumbers and reduce the amount of data in part. Fourthly, principal component analysis (PCA) was used to compress spectral data into several variables, and the cumulate reliabilities of the first ten components were more than 99.98%. Finally, according to the first ten components, recognition models were established based on BPR. For every 25 samples in each variety, 15 samples were randomly selected as the training set. The remaining 10 samples of the same variety were used as the first testing set, and all the 900 samples of the other varieties were used as the second testing set. Calculation results showed that the average correctness recognition rate of the 37 varieties of corn seed was 94.3%. Testing results indicate that the discrimination method had higher precision than the discrimination of various kinds of commodity corn seed. In short, it is feasible to discriminate various varieties of commodity corn seed based on near infrared spectroscopy and BPR.

Keywords Near infrared spectral(NIRS); Biomimetic pattern recognition (BPR); Commerical corn seed; Discrimination

(Received Jun. 22, 2009; accepted Sep. 26, 2009)

* Corresponding author