

基于近红外光谱技术和人工神经网络的玉米品种鉴别方法研究

陈建¹, 陈晓¹, 李伟^{1*}, 王加华², 韩东海²

1. 中国农业大学工学院, 北京 100083

2. 中国农业大学食品科学与营养工程学院, 北京 100083

摘要 提出了一种采用近红外光谱技术结合人工神经网络对玉米品种进行鉴别的方法。在 3 800~10 000 cm^{-1} (波长 1 000~2 632 nm) 范围内采集四种玉米单粒完整籽粒的近红外漫反射光谱, 经 Savitky-Golay 平滑和多重散射校正预处理后, 对数据进行主成分分析, 再结合人工神经网络技术进行品种鉴别。主成分分析表明, 前 8 个主成分的累积贡献率达到 99.602%。以前 8 个主成分作为网络输入, 品种类型作为输出, 建立三层 LMBP 神经网络模型。每个品种各取 30 粒共 120 个样本用于建模, 10 粒共 40 个样本用于预测。模型对建模集 120 个样本鉴别率为 100%, 对预测集 40 个样本的鉴别率为 95%。实验结果说明该方法能快速无损地鉴别玉米品种, 为玉米的品种鉴别提供了一种新方法。

关键词 近红外光谱; 主成分分析; 人工神经网络; 玉米鉴别

中图分类号: O657.3 **文献标识码:** A **文章编号:** 1000-0593(2008)08-1806-04

引言

当前应用于玉米品种鉴定的方法主要有籽粒形态鉴定、幼苗鉴别、田间小区种植、蛋白质电泳和 DNA 分子标记等^[1, 2]。籽粒形态鉴定使用范围窄, 人为影响大, 识别率低; 幼苗鉴定只适用于某些差异较大的品种, 可鉴定的品种少; 田间小区种植所需时间长, 成本高, 受季节限制; 蛋白质电泳和 DNA 标记准确率高, 但样品一般要经过物理或化学方法处理, 分析速度慢, 成本高, 样品易遭破坏^[3]。机器视觉和图像处理技术在玉米品种的鉴定中也得到了广泛应用, 加快了检测速度, 但仍然克服不了形态鉴定使用范围窄、识别率低和化学方法破坏种子、使用化学试剂等缺陷^[3, 4]。

近红外光谱分析技术具有速度快、成本低、无污染以及不破坏样品等优点, 在食品、中草药和农产品快速鉴别中得到了许多成功的应用^[5-9]。国内外学者已利用近红外光谱技术定量分析玉米成分含量^[10-12], 检测玉米损伤^[13], 但是在玉米品种鉴别上的应用研究还未见报道。本研究将近红外光谱技术结合人工神经网络应用到单粒玉米种子的鉴别中, 实现了玉米品种快速无损的鉴别。

1 材料与方法

1.1 实验材料与设备

实验用样品高油 115、368、农大 108、4967 等四种玉米种子, 由中国农业大学农学院提供。产地甘肃, 2005 年采集, 自然晾晒干。从每个品种中选取 40 粒正常籽粒, 共计 160 个样本用于实验。

实验所用仪器为美国 Thermo Nicolet 公司 ANTARIS 傅里叶变换近红外光谱仪, 仪器采用 InGaAs 检测器, 光谱范围 833~2 632 nm, 最小分辨率 2 cm^{-1} 。分析软件为仪器自带 TQ Analyst V6.2。

1.2 单粒玉米光谱采集

在室温下, 采集单粒完整玉米籽粒漫反射光谱。把单粒玉米盛于直径 50 mm 的旋转样品池, 胚面向下, 在样品池内部放置一个直径 10 mm 的套杯, 以保证单籽粒样品置于样品池的光学中心, 采样光斑直径为 10 mm。使用漫反射大体积分球作为光谱采集方式, 在 3 800~10 000 cm^{-1} (波长 1 000~2 632 nm) 范围内采集, 间隔 2 cm^{-1} , 扫描 64 次取平均。

1.3 LMBP 网络模型

人工神经网络是由人工建立的以有向图为拓扑结构的动态系统, 通过对连续或断续的输入作状态响应而进行信息处

收稿日期: 2007-05-26, 修订日期: 2007-08-28

基金项目: 国家自然科学基金项目(30471011)和高校博士学科点专项科研基金项目(20050019005)资助

作者简介: 陈建, 1981 年生, 中国农业大学工学院硕士研究生 e-mail: gxy5@cau.edu.cn * 通讯联系人

理^[14]。采用误差反向传播(back propagation, BP)算法的前馈神经网络,能够实现输入与输出之间的高度非线性映射。虽然神经网络在各方面都已得到广泛的应用,但标准的BP算法存在收敛速度慢、易陷入局部最小等问题^[15],本研究采用 Levenber-Marquardt(LM)算法优化标准 BP 网络,加快收敛速度。LM算法是一种利用标准的数值优化技术的快速算法,它在高斯牛顿法和梯度下降法之间进行平滑调和,在远离最小值时逐渐切换到高斯牛顿法。它既有高斯牛顿法的局部收敛性,又具有梯度下降法的全局特性,可以加快网络的收敛速度^[16, 17]。

本研究建立了一个三层的 LMBP 网络模型,各层传递函数均为 S 型函数。网络输入层节点数为 8,隐含层节点数为 8,输出层节点数为 1(玉米不同的品种值),目标误差为 0.01,学习速率为 0.2,设定训练迭代次数为 5 000 次。为了使识别效果更为理想,对输出值进行归一化处理 and 反归一化处理。

2 实验结果与分析

2.1 样本的近红外漫反射光谱

四种玉米种子的原始光谱曲线如图 1 所示。横坐标为波

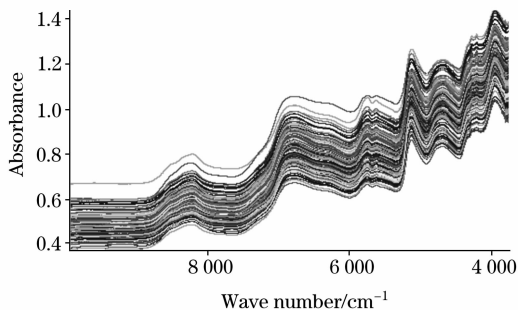


Fig. 1 Original NIR spectra of all individual corn kernels

长,范围是 3 900~10 000 cm^{-1} ,纵坐标为光谱漫反射率,可以看出不同玉米品种的光谱曲线没有明显区别。

为了去除高频随机噪声以及样本不均导致的基线漂移的影响,应用 TQ6.2 软件对原始光谱进行预处理。先用卷积平滑法(Savitzky-Golay),选用平滑窗口为 9,滤除高频噪声;再对光谱进行多元散射校正(MSC, multiplicative scatter correction)处理,消除基线漂移的影响;最后转存为吸光矩阵。图 2 是处理后的光谱曲线。

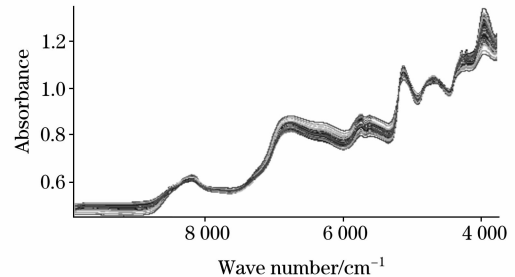


Fig. 2 NIR spectra after Savitzky-Golay and MSC

2.2 光谱数据的主成分分析

主成分分析的目的是将数据降维,以消除众多信息共存中相互重叠的信息部分^[18]。方法是通过对实测的多个指标相关矩阵内部结构关系的研究,构造少数几个综合的主成分指标,它们都是原指标的线性组合,不仅保留原指标的主要信息,且互不相关,同时比原指标具有更好的性质。

实验采集的玉米图谱从 1 000~2 632 nm 共有 3 216 个数据点,数据量大,冗余信息多。利用 Matlab 对数据进行主成分分析。累积贡献率反应了前 n 个主成分对原数据信息的表征能力^[19]。经过计算,前 8 个主成分的累积贡献率已达到 99.602%(如表 1 所示),说明这 8 个变量很好的表征了原数据的信息。这样数据阵就从 $160 \times 3 216$ 减少到 160×8 (8 个主成分)。

Table 1 PCs and reliabilities

主成分	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
累积贡献率/%	73.478	93.068	96.775	97.97	98.858	99.185	99.455	99.602

2.3 基于 8 个主成分建立 LMBP 玉米品种鉴别模型

把所有样本分为建模集和预测集,从每个玉米品种各随机选取 30 粒共计 120 个样本用作建模,10 粒共 40 个样本用作预测。神经网络模型要求输入变量少而精,即减少各输入变量的相关性,同时尽量保持原始信息^[20]。因此在主成分分析基础上,把前 8 个主成分作为 LMBP 网络的输入节点。输出节点为 1 即玉米的品种值(将高油 115、368、农大 108、4967 品种值分别设为 1, 2, 3, 4)。通过多次实验隐含层节点为 8 的时候,模型预测效果最好。这样就建立了一个 8(输入节点)-8(隐含层节点)-1(输出节点)的三层 LMBP 网络模型。模型对建模集 120 个样本的拟合残差为 $9.782 69 \times 10^{-7}$,识别率为 100%;对预测集 40 个样本品种识别率达到 95%(见表 2)。其中模型对高油 115 和 4 967 识别率达到 100%,368 和农大 108 都有一个被误判。

3 结论

本研究利用近红外光谱技术结合人工神经网络技术对四种玉米种子进行快速无损鉴别。采集了四种单粒玉米近红外漫反射光谱,经平滑处理和多元散射校正(MSC)后,用主成分分析法对数据进行降维,把前 8 个主成分作为输入建立三层 LMBP 神经网络模型,模型对未知样本的识别率达到 95%。该方法具有快速、无损、无污染、成本低等优点,避免了人工鉴别中主观因素的影响,克服了其他鉴别方法破坏种子、过于复杂、耗时的缺点。实验表明该方法可以作为玉米品种鉴别的新方法加以推广应用,并为其他农作物种子的品种鉴别提供了借鉴。

Table 2 Prediction result for unknown samples by LMBP model

预测样本	真实值	预测值	预测样本	真实值	预测值	预测样本	真实值	预测值	预测样本	真实值	预测值
(1)	1	1	(11)	2	2.231 7	(21)	3	2.998 7	(31)	4	4
(2)	1	1	(12)	2	1.933 1	(22)	3	2.465 7	(32)	4	3.999 9
(3)	1	1	(13)	2	1.689 1	(23)	3	3.004 2	(33)	4	3.604 3
(4)	1	1	(14)	2	1.813 9	(24)	3	2.964 2	(34)	4	4
(5)	1	1.008 9	(15)	2	1.825 3	(25)	3	2.827 1	(35)	4	3.956 6
(6)	1	1	(16)	2	2.368 5	(26)	3	2.971 6	(36)	4	3.999 8
(7)	1	1	(17)	2	2.072 6	(27)	3	3.063 1	(37)	4	4
(8)	1	1	(18)	2	1.853 1	(28)	3	2.991 7	(38)	4	4
(9)	1	1	(19)	2	2.286 9	(29)	3	3.076 5	(39)	4	4
(10)	1	1.000 1	(20)	2	2.711 5	(30)	3	3.044 6	(40)	4	4

Note: 真实值 1, 2, 3, 4 分别代表高油 115、368、农大 108、4967

参 考 文 献

- [1] WAN Pu-lin(万普林). Xandai Nongye Keji(现代农业科技), 2005, (7): 13.
- [2] WANG Xi, ZHANG Bao-shi(王 玺, 张宝石). Seed(种子), 2002, (1): 43.
- [3] LIU Yan-de, YING Yi-bin, CHENG Fang(刘燕德, 应义斌, 成 芳). Transactions of the Chinese Society for Agricultural Machine(农业机械学报), 2003, 34(5): 161.
- [4] Loren Steenhoek, Cyrille Precetti. 2000 ASAE Annual International Meeting, Technical Papers; Engineering Solutions for a New Century, 2000. 1125.
- [5] Leon Lorenzo, Daniel Kelly J, Downey Gerard. Applied Spectroscopy, 2005, 59(5): 593.
- [6] Young-Ah Wooa, Hyo-Jin Kima, Keum-Ryon Zeb, et al. Journal of Pharmaceutical and Biomedical Analysis, 2005, 36: 955.
- [7] HE Yong, FENG Shui-juan, LI Xiao-li, et al(何 勇, 冯水娟, 李晓丽, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(11): 2021.
- [8] ZHAO Jie-wen, CHEN Quan-sheng, ZHANG Hai-dong, et al(赵杰文, 陈全胜, 张海东, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(9): 1601.
- [9] HE Yong, LI Xiao-li(何 勇, 李晓丽). J. Infrared Millim. Waves(红外与毫米波学报), 2006, 25(6): 192.
- [10] Campbell M R, Mannis S R, Port H A, et al. Cereal Chemistry, 1999, 76(4): 552.
- [11] Tesfaye M Baye, Tom C Pearson, A Mark Settles. Journal of Cereal Science, 2006, 43: 236.
- [12] WEI Liang-ming, JIANG Hai-ying, LI Jun-hui, et al(魏良明, 姜海鹰, 李军会, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(9): 1404.
- [13] Harper Raquel. Biophotonics International, 2005, 12(12): 23.
- [14] Furrer David, Thaler Stephen. Advanced Materials and Processes, 2005, 163(11): 42.
- [15] Jui-Hong Horng. Advances in Engineering Software. In press.
- [16] WANG Yun-song, XU Hong-guo(王赞松, 许洪国). Journal of Jilin University(Engineering and Technology Edition)(吉林大学学报, 工学版), 2003, 33(4): 79.
- [17] WU Ling-yun(吴凌云). Information Technology(信息技术), 2003, 27(7): 42.
- [18] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍, 袁洪福, 徐广通, 等). Modern Near Infrared Spectrum Analysis Technique(现代近红外光谱分析技术). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2000. 129.
- [19] DENG Zu-xin. Data Analytics and SAS System(数据分析方法和 SAS 系统). Shanghai: Shanghai University of Finance & Economics Press(上海财经大学出版社), 2006. 291.
- [20] QI Xiao-ming, ZHANG Lu-da, DU Xiao-lin, et al(齐小明, 张录达, 杜晓林, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(5): 870.

Study on Discrimination of Corn Seed Based on Near-Infrared Spectra and Artificial Neural Network Model

CHEN Jian¹, CHEN Xiao¹, LI Wei^{1*}, WANG Jia-hua², HAN Dong-hai²

1. College of Engineering, China Agricultural University, Beijing 100083, China

2. College of Food Science & Nutritional Engineering, China Agricultural University, Beijing 100083, China

Abstract A new non-destructive and rapid method was developed to discriminate varieties of corn seeds. The method is based on near-infrared reflectance spectroscopy (NIRS) and artificial neural network (ANN). The corn seeds used for this study involved four varieties: Gaoyou115, Nongda368, Nongda108 and Nongda4 967. After collecting the near-infrared reflectance spectrum of each single seed in the range between 1 000 and 2 632 nm, the principal component analysis (PCA) was used to compress the NIR spectra, which had been preprocessed with Savitky-Golay and multiplicative scatter correction (MSC). The analysis results showed that the cumulate reliabilities of PC1 to PC8 (the first eight principal components) were 99.602%. A three-layer back-propagation neural network (BPNN) was developed for classification, which was trained by the Levenberg-Marquard algorithm to improve the network training speed and efficiency. The LMBP was activated by the sigmoid function, and normalization of targets was used to get the best discrimination result of network. The first eight principal components of the samples were applied as LMBPNN inputs, and the values of the type of corn seeds were applied as the outputs. In this model, 120 kernels were used as the training data set and 40 kernels were used as the test data set. Calculation results showed that the distinguishing rate of the four corn seed varieties was 95%. This model is reliable and practicable. The results demonstrated that this identification method was rapid and non-destructive, and could be used for classification.

Keywords Near infrared spectra (NIRS); Principal component analysis (PCA); Artificial neural network(ANN); Corn seed; Discrimination

* Corresponding author

(Received May 26, 2007; accepted Aug. 28, 2007)

2008 第 3 届亚太地区冬季等离子体光谱化学会议 2008 Third Asia-Pacific Winter Conference on Plasma Spectrochemistry (2008 APWC)

由日本化学会和日本分析化学学会主办的第 3 届亚太地区冬季等离子体光谱会议(2008 APWC)将于 2008 年 11 月 16~21 日在日本筑波国际会议中心举行。

征文范围:

1. ICP-MS 和 ICP-OES 的仪器与原理; 2. 样品的制备和导入; 3. ICP-MS 的碰撞反应池技术与应用; 4. 新仪器与等离子体源; 5. 辉光放电光源; 6. 环境、金属组学和生命科学中的元素形态; 7. ICP-MS 和 ICP-OES 的应用, 包括: 冶金、RoHS、高纯物质及半导体材料、核工业材料、环境、地质、海洋科学、食品、农业、生物学、临床和药物、质量控制和标准化等; 8. 同位素比和同位素稀释; 9. 激光烧蚀进样。

会议地点:

筑波国际会议中心(Tsukuba International Congress Center) (日本茨城县筑波市(Tsukuba, Ibaraki, Japan))

联系人: Prof. Naoki Furuta

Chuo University, Faculty of Science and Engineering, Department of Applied Chemistry, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Tel: +81-3-3817-1906, Fax: +81-3-3817-1699

e-mail: nfuruta@chem.chuo-u.ac.jp

会议使用英语交流

会议详细信息请登录网站查询 <http://envsun.chem.chuo-u.ac.jp/plasma/2008apwc>