江西农业大学学报 2010,32(4):0723 - 0728 http://xuebao.jxau.edu.cn
Acta Agriculturae Universitatis Jiangxiensis E - mail: ndxb7775@sina.com

基于光谱技术识别不同农药污染脐橙的研究

黎 静[1],薛 龙[2],刘木华[1*],王 晓[1],罗春生[1]

(1. 江西农业大学 工学院,江西 南昌 330045; 2. 华东交通大学 机电学院,江西 南昌 330013)

摘要: 用遗传算法(Genetic Algorithm,GA)搜寻可识别被不同农药污染脐橙的可见/近红外光谱的最佳特征光谱区间及波长,并建立了支持向量机(Support Vector Machines,SVM)定性分析模型。实验供试农药为灭多威、氰戊菊酯和氧乐果3种。通过GA来搜寻整个波段范围(460~1 800 nm),将得到的9个最佳特征光谱区间所包含的波长(共318个)作为SVM建模的输入变量,对识别被3种农药污染脐橙的准确率为100%。并继续应用GA优化,得到71个特征波长,此时建立的SVM模型的识别准确率为99.57%。虽然识别的准确率有所下降,但是模型的复杂程度得到了很大的优化,其输入变量减少到71个。实验结果表明利用可见/近红外光谱技术结合SVM方法可以有效识别被不同农药污染的脐橙。

关键词: 农药污染; 脐橙; 可见/近红外光谱; 无损检测; 支持向量机; 遗传算法
中图分类号: O436; S123 **文献标志码:** A **文章编号:** 1000 - 2286(2010)04 - 0723 - 06

Recognition of Different Pesticide Contamination in Navel Oranges Based on Spectra Technology

LI Jing[1],XUE Long[2],LIU Mu-hua[1*],WANG Xiao[1],LUO Chun-sheng[1]

(1. College of Engineering,JAU, Nanchang 330045, China; 2. College of Mechanical and Electronical Engineering, East China Jiaotong University, Nanchang 330013,China)

Abstract: Genetic algorithm (GA) was used to search for the best characteristic spectral ranges and wavelengths of visible/near - infrared spectra (Vis/NIRs), a qualitative analysis model of support vector machine (SVM) was set up to recognize navel oranges contaminated with different pesticides. The pesticides in the experiment were Methomyl, fenvalerate and omethoate. Using GA to search the entire band range (460 ~ 1 800 nm), the 9 best characteristic spectral ranges (318 wavelengths) were used as the input variables of SVM model and the accuracy of the prediction set classification was 100%. Then GA method was used continually and 71 wavelengths were extracted, the corresponding SVM model was built with 99.57% accuracy. Although the classification accuracy rate declined, the complexity of the model was greatly optimized by reducing the input variables to 71. The experiment results showed that the application of Vis/NIRs combined with SVM can effectively detect the navel oranges contaminated with different pesticides.

Key words: pesticide contamination; Vis/NIRs; non - destructive detection; support vector machine; genetic algorithm

收稿日期: 2010 - 05 - 26 **修回日期:** 2010 - 06 - 17
基金项目: 国家自然科学基金资助项目(30760101)和江西省教育厅科学技术研究项目(GJJ08513)
作者简介: 黎静(1977 -),女,讲师,硕士,主要从事农业机械设计及农产品品质无损检测研究,E - mail: lijing3815@163.com; * 通讯作者: 刘木华,博士,教授,E - mail: suikelmh@sohu.com。

随着人们生活水平的提高,农产品品质与安全问题已经受到人们越来越多的重视。在我国,检测农产品的农药残留主要由食品安全检验检疫局来完成,其采用的分析方法主要是气相色谱和液相色谱。这些常规的分析方法不仅费时、费力,对被检测样品具有破坏性,而且也不能满足现代大规模的工业生产。近红外光谱具有很强的穿透能力,在生产流水线上配置近红外装置能直接检测农产品,便可实现现场实时在线快速的无损检测。

近20年来,在分析化学领域中,光谱技术已经作为一种快速、无损、可多组分同时分析的技术,在工、农业领域中被越来越广泛的应用[1-5]。Sirinnapa Saranwong等[6-7]应用近红外光谱技术检测马铃薯表面的农药残留,而且还对此方法检测农药残留的可靠性进行了进一步的研究。Sergio Armenta等[8]应用傅立叶红外光谱仪建立了自动检测稻虱净的实验装置和数学模型。但是这些方法并不是直接从样品表面获得近红外光谱,而是先用化学溶液(如丙酮和乙腈)来清洗样品表面得到溶解了农药的溶液,然后采集这些溶液的近红外光谱,并对此光谱进行分析处理。胡淑芬等利用了激光图像技术识别脐橙表面的农药残留[9-10]。本文主要利用直接在水果表面采集的近红外光谱,应用遗传算法[11-17](Genetic Algorithm,GA)对整个光谱区间进行优选,确定最佳的特征光谱区间及波长,并将其作为非线性函数估计的支持向量机器法[16](A Library for Support Vector Machines,SVM)的输入变量,建立预测模型,对未知样本的脐橙是否被农药污染以及被何种农药污染进行了识别与预测。

1 材料与方法

1.1 仪器设备

实验所使用的近红外光谱仪为Quality Spec Pro光谱仪(Analytical Spectral Devices,Inc.,USA),其波长范围350~1 800 nm,光谱采样间隔1 nm,扫描30次。应用Indico 4.0(Analytical Spectral Devices,Inc.,USA)软件和与之配套的标准白板采集归一化后的光谱数据,利用Matlab 7.1(Mathworks,Inc.,USA)软件进行数据分析与处理。

1.2 实验材料

实验使用的3种农药分别是:φ=24%灭多威水剂(methomyl)、φ=20%氰戊菊酯(fenvzlerate)乳油、φ=40%氧乐果(omethoate)乳油,分别用蒸馏水配置1∶500倍的农药溶液,其浓度为果农常用浓度。

实验样品来源于中国江西省赣州市宁都县某脐橙果园。选择没有表面缺陷、碰伤的脐橙共345个,清洗并自然风干,然后采集脐橙最大横径处的光谱数据,将这些未喷施任何农药的脐橙数据作为第4组。完成数据采集后把脐橙随机分成3组,每组脐橙个数分别为114、114和117个,分别喷施配置好的3种不同农药溶液,第1组为喷施灭多威后的脐橙,第2组为喷施氰戊菊酯后的脐橙,第3组为喷施氧乐果后的脐橙,每一组为一个类别。在环境温度为10 ℃和相对湿度为60%的实验室条件下,把上述前3组脐橙放置168 h后采集脐橙最大横径处的光谱数据。按每组脐橙总数2∶1的比例随机抽取建模集与预测集样品。表1列出了4个组分别用于建模集和预测集的脐橙光谱数据个数,其中第1组到第4组中,每个组中用于建模集和预测集的脐橙数目分别为76、76、79、230和38、38、39、115,建模集和预测集的样品数目总数分别为460和230。

表1 每组脐橙样本数目及农药类型

Tab.1 Navel orange sample amounts and pesticide type of each group

组别 Group	农药种类(1∶500) Pesticide type	数量 Amount		总数 Total
		建模集 Prediction set	预测值 Colibration set	
1	φ=24%灭多威水剂 φ=24% Methomyl AS	76	38	114
2	φ=20%氰戊菊酯乳油 φ=20% Fenvzlerate EC	76	38	114
3	φ=40%氧乐果乳油 φ=40% Omethoate EC	78	39	117
4	未喷施农药 Non-sprayed	230	115	345

1.3 数据处理方法

近红外光谱往往包含一些与待测样品性质无关的因素带来的干扰,导致了近红外光谱的基线漂移

和光谱的不重复[17]，因此必须对原始光谱进行预处理。本文采用标准正交变量变换对原始光谱进行预处理。

遗传算法是模仿自然界生物进化机制发展起来的优化方法，作为一种实用、高效、鲁棒性强的优化技术，已为广大学者重视。在近红外光谱检测应用中，遗传算法被用来对整个光谱区间进行优化，选择出最有效的特征区间或特征波长[6-8]。图 1 为建模集脐橙 460 个脐橙样本的原始光谱图，光谱范围是 350 ~ 1 800 nm。从图 2 中可以看出在波段 350 ~ 459 nm 区间的光谱曲线所包含的噪音较多，因此本文中所选用的波段范围为 460 ~ 1 800 nm，共计 1 341 个波长数据。把整个区间分成 40 个子区间，除第 40 个子区间包含 54 个波长数据外，其余子区间均包含 33 个波长数据。用 1 个含有 0/1 且长度为 40 个字符的字符串 S 来表示 40 个子区间的选取，其中 1 和 0 分别表示其所对应的子区间是否被选取。

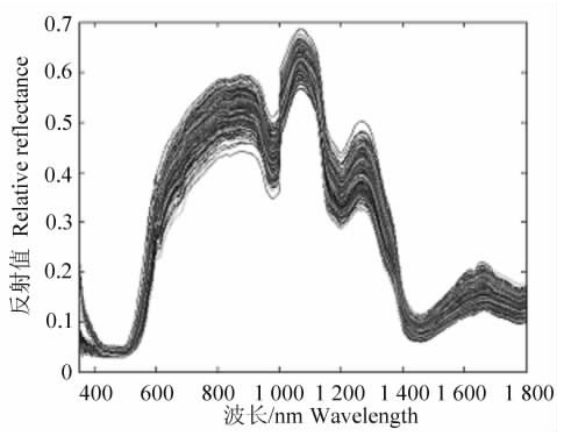

图 1　建模集样本的原始光谱图

Fig. 1　Vis/NIRs of navel orange

本文应用支持向量机(SVM) 通过建模集数据建立模型，然后应用此模型对预测集数据进行识别，以分类的准确率(Accuracy) 为遗传算法的目标函数，终止条件为达到最大的迭代次数。最佳的特征光谱区间为遗传迭代后识别准确率为最大值的区间组合。支持向量机的模型选择问题就是给定一个核函数，通过调节核参数和误差惩罚参数 C 来提高支持向量机训练精度，同时降低错误率，因此支持向量机的参数选择直接影响着 SVM 的性能。本研究建立模型所采用的核函数为高斯径向基核函数，即公式(1) 所示:

$$K(x_i, x_j) = \exp(-\gamma \cdot \| x_i - x_j \|^2), \gamma > 0 \quad (1)$$

(1) 式中，(x_i, x_j) 为样本数据。应用 *SVM* 进行识别时，核函数参数 γ 和 C 的选择是一个重要问题，网格搜索法简单直接，因为每一个参数对(γ, C) 是独立的，可以并行地进行网格搜索，因此本文采用网络搜索方法来选择最优的独立参数对(γ, C)，以得到的最高准确率为评判依据，通过对全光谱的分析得到 γ 和 C 的取值分别为 232 与 3。

2　结果与讨论

2.1　波段范围及特征波长的选择

本文中 GA 的设定参量: 区间数 40，初始群体 30，变量的二进制位数 40，代沟 0.9，交叉概率 0.7，遗传迭代次数 45。图 2 是经过 45 次迭代后的结果，实线表示每一代中最大的识别准确率，虚线表示每一代中所有个体识别准确率的平均值，可以看出在第 40 到第 42 代识别准确率最大且值为 100%，说明最佳特征波长组合对预测集的识别准确率达到的 100%。经过对这三代中最优个体的分析发现，其最佳特征区间均相同，其原因是在第 40 代中的最佳个体作为父辈被完整的遗传到了下一代中，因此最佳迭代数为第 40 代。图 3 是第 40 代的最佳特征区间组合，图中灰色的条带区域表示此处的波段为被选中的 9 个特征区间，对应的波长数目为 318 个。

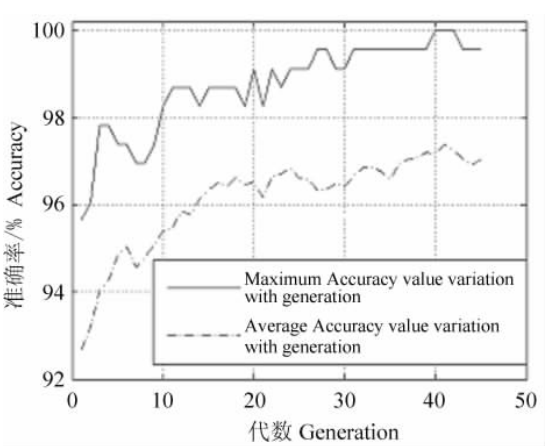

图 2　每代中最大值 Accuracy 与 Accuracy 的均值的变化

Fig. 2　Maximum Accuracy and average Accuracy value variation with generation

在被选中的9个特征区间的基础上，继续应用GA－SVM算法进行优化。此时GA的设定参量为：区间数318，初始群体40，变量的二进制位数318，代沟0.9，交叉概率0.7，遗传迭代次数100。图4为经过100次迭代后，每个变量按照其被选取的次数所绘制的频率图。根据每个变量被选取的频率，由大到小重新排序，然后依次累加作为SVM的输入变量。图5为输入变量数与识别准确率变化曲线图，由图可以看出在变量数为296时，模型的识别准确率也达到了100%，虽然其模型的输入变量数减少了32.07%，但是其变量数仍然较多。因此，根据图5识别准确率的变化趋势，最终确定输入变量数为71，而此时的识别准确率为99.57%。

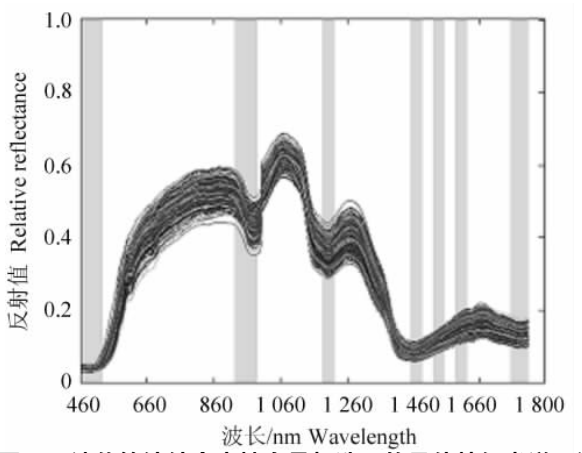

图3 遗传算法结合支持向量机选取的最佳特征光谱区间

Fig. 3 Characteristic spectral regions selected by genetic algorithm and support vector machine

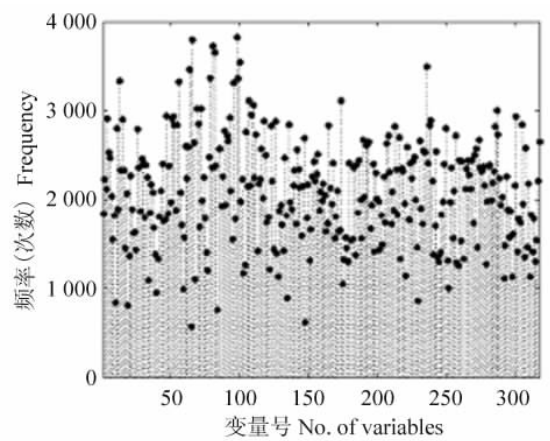

图4 变量选取的频率图

Fig. 4 Selected frequency of each variable

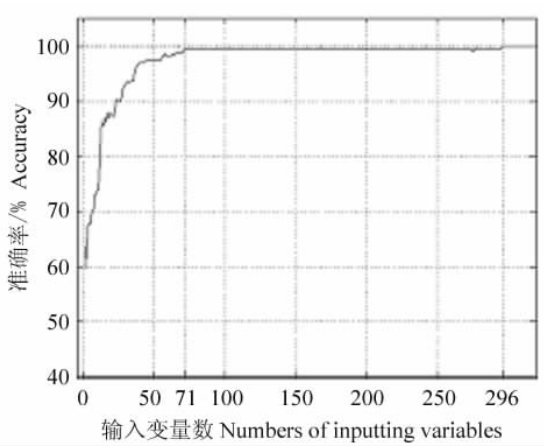

图5 准确率随输入变量数的变化趋势

Fig. 5 Accuracy versus inputting variables

2.2 GA－SVM选取特征波长与全光谱识别准确率的比较

表2列出应用GA－SVM所建立的模型结果与仅应用SVM方法所建立的全光谱模型的比较结果。

从表2中可以看出SVM的全光谱预测准确率较低，且SVM模型的输入参数为1 341个，这使得模型的复杂度很高。使用GA－SVM所得到的模型建立在9个最佳特征光谱区间的基础之上(共318个光谱数据点)，无论是模型的预测精度还是模型的简洁度都优于全光谱模型，而且对预测集的预测结果的准确率为100%。因此该方法不仅在降低模型复杂度(模型的输入参数为318个)上效果明显，而且其模型的预测精度也得到大幅度的提高。在选定的9个最佳特征光谱基础上，应用GA－SVM方法对模型的输入变量进行进一步的筛选，得到特征波长为71个，虽然其预测精度略有降低，但是其模型的预测能力却没有明显的改变，说明应用GA－SVM方法所建立的识别模型比全光谱模型更加稳定、简洁。

3 结 论

本文主要应用GA－SVM方法，对整个脐橙的近红外光谱(光谱范围460～1 800 nm)进行了特征光谱区间及波长的选取，并建立了识别被不同农药污染脐橙的识别模型。结果表明，与全光谱区间相比，应用GA－SVM算法选取最佳特征波长，使得模型的输入参数从1 341个减少到71个，不仅大大的缩短了运算时间，而且也剔除了光谱曲线特征差异不明显的区段，并且模型的预测精度也有显著的提高，本文中的预测精度达到了99.56%。因此，结合GA和SVM两种方法选取特定波长一方面可以简化模型的复杂程度，提高预测能力，另一方面可以为设计近红外快速检测仪提供一种可行的波长选取方法。

表 2　不同选择方法下应用 SVM 方法的检测结果

Tab. 2　Summary of SVM results by different methods

选择方法 Select method	所选择的光谱区间或波长/nm Selected spectral region or wavelength	SVM 模型中变量数 Variables in SVM model	预测集识别准确率/% Accuracy of prediction set
支持向量机(全光谱) SVM (Full – spectrum)	460 ~ 1 800	1 341	83.91
遗传算法 – 支持向量机(9 个特征波段) GA – SVM (9 characteristic spectral regions)	460 ~ 525, 922 ~ 987, 1 186 ~ 1 218, 1 450 ~ 1 482, 1 516 ~ 1 548, 1 582 ~ 1 614, 1 747 ~ 1 800	318	100
遗传算法 – 支持向量机(71 个特征波长) GA – SVM (71 characteristic wavelengths)	1 584, 1 195, 940, 1 754, 971, 941, 1 790, 522, 1 539, 520, 1 476, 1 477, 1 524, 922, 1 478, 1 800, 947, 1 548, 1 756, 1 474, 1 200, 1 533, 1 607, 946, 1 525, 967, 1 770, 960, 944, 485, 470, 1 600, 1 768, 1 530, 1 218, 1 588, 978, 1 189, 514, 512, 1 787, 925, 973, 981, 1 589, 474, 463, 510, 948, 1 783, 511, 506, 964, 1 769, 927, 924, 965, 1 458, 962, 951, 515, 472, 954, 934, 523, 1 586, 955, 937, 936, 525, 953	71	99.56

参考文献:

[1] V A McGlone, C J Clark, R B Jordan. Comparing density and VNIR methods for predicting quality parameters of yellow – fleshed kiwifruit (Actinidia chinensis) [J]. Postharvest Biol Technol, 2007, 46(1): 1 – 9.

[2] 赵杰文, 张海东, 刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法 [J]. 光学学报, 2006, 26(1): 136 – 140.

[3] 王多加, 周向阳, 金同铭, 等. 近红外光谱检测技术在农业和食品分析上的应用 [J]. 光谱学与光谱分析, 2004, 24(4): 447 – 450.

[4] 赵杰文, 郭志明, 陈全胜, 等. 近红外光谱法快速检测绿茶中儿茶素的含量 [J]. 光学学报, 2008, 28(12): 2302 – 2306.

[5] 王玉田, 崔立超, 王冬生, 等. 基于同步 – 导数荧光光谱法的多组分农药残留测定的研究 [J]. 光谱学与光谱分析, 2006, 26(11): 2085 – 2088.

[6] S Saranwong, S Kawano. Rapid determination of fungicide contaminated on tomato surfaces using the DESIR – NIR: a system for ppm – order concentration [J]. J Near Infrared Spectrosc, 2005, 13(3): 169 – 175.

[7] S Saranwong, S Kawano. The reliability of pesticide determinations using near infrared spectroscopy and the dry – extract system for infrared (DESIR) technique [J]. J Near Infrared Spectrosc, 2007, 15(4): 227 – 236.

[8] S Armenta. Automated Fourier transform near infrared determination of buprofezin in pesticide formulations [J]. J Near Infrared Spectrosc, 2005, 13(3): 161 – 168.

[9] 胡淑芬, 药林桃, 刘木华. 脐橙表面农药残留的计算机视觉检测方法研究 [J]. 江西农业大学学报, 2007, 29(6): 1031 – 1034.

[10] 胡淑芬, 刘木华, 林怀蔚. 基于激光图像的水果表面农药残留检测试验研究 [J]. 江西农业大学学报, 2006, 28(6): 872 – 876.

[11] T A Lestander. Selection of near infrared wavelengths using genetic algorithms for the determination of seed moisture content [J]. J Near Infrared Spectrosc, 2003, 11(6): 433 – 446.

[12] R Leardi, A Lupiáez, González. Genetic algorithms applied to feature selection in PLS regression: how and when to use them [J]. Chemometrics and Intelligent Laboratory Systems, 1998, 41(2): 195 – 207.

[13] R Leardi, L N. rgaard. Sequential application of backward interval PLS and genetic algorithms for the selection of relevant spectral regions [J]. J Chemomelrics, 2004, 18(11): 486 – 497.

[14] 皱小波,赵杰文. 用遗传算法快速提取近红外光谱特征区域和特征波长[J]. 光学学报,2007,27(7):1316-1321.
[15] R Leardi. Application of genetic algorithm-PLS for feature selection in spectral data sets[J]. J Chemometrics and Intelligent Laboratory Systems,2000,14(5):643-655.
[16] Rong-en Fan, Pai-hsuen Chen, Chih-jen Lin. Working set selection using second order information for training SVM[J]. Journal of Machine Learning Research,2005,6(11):1889-1918.
[17] 陆婉珍. 现代近红外光谱分析技术[M]. 2版. 北京:中国石化出版社,2006:35.

(上接第709页)

[8] Ju Z G, Ju Z G. Relationship among phenylanine ammonia-lyase activety[J]. Scidntia Horticulture,1995(61):215-226.
[9] 高大同. 套袋对梨、苹果果实生长发育及性状影响的研究[D]. 南京:南京农业大学,2006.
[10] 冯少菲. 黄金梨果皮发育、锈斑形成及套袋对其影响的研究[D]. 保定:河北农业大学,2006.
[11] 孟焕文,程智慧,杨玉梅,等. 套袋及遮光对黄色果实发育及品质的影响[J]. 西北农林科技大学学报:自然科学版,2004,32(12):44-45,51.
[12] 张建光,刘玉芳,孙建设,等. 光照强度对果实表面温度变化的影响[J]. 生态学报,2004,24(6):1306-1310.
[13] 吕英民,张大鹏. 果实发育过程中糖的积累[J]. 植物生理学通讯,2000,36(3):258-265.
[14] 胡红菊,陈启亮,王友平,等. 4个砂梨品种果实发育过程中主要糖酸含量的变化[J]. 华中农业大学学报,2007,26(2):251-255.
[15] 王涛,林媚,王海琴,等. 设施条件下4个中熟砂梨品种果实发育及糖酸含量的变化[J]. 中国农学通报,2008,24(8):350-354.
[16] 王少敏,高华君,王永志,等. 不同纸袋对丰水梨套袋效果比较试验[J]. 中国果树,2001(2):12-14
[17] Arakova, Uematsu N, Na Kajima H. Efect of bagging On fruit quality in apples[J]. Bulletin of the Faculty of Aculture Hirosaki University,1994,57:25-32.
[18] Li S H, Genard M, Bussi C. Fruit quality and leaf photosynthesis in response to microenvironment modification around individual fruit by covering the fruit with plastic in nectarine and peach trees[J]. Hort Sci & Biotechnology,2001,76(1):61-69.
[19] 郝燕燕,李妙玲,张惠荣,等. 套袋微环境对果实品质的影响及其机理分析[J]. 山西农业大学学报,2003(3):238-242.