

# 基于 GA-LSSVM 和近红外傅里叶变换的霉变板栗识别

周竹<sup>1</sup>, 李小昱<sup>1\*</sup>, 李培武<sup>2</sup>, 高云<sup>1</sup>, 展慧<sup>1</sup>, 刘洁<sup>1</sup>

(1. 华中农业大学工学院, 武汉 430070; 2. 中国农业科学院武汉油料作物研究所, 武汉 430062)

**摘要:** 为克服板栗近红外光谱变量多、共线性强等缺点, 该文对标准正态变量变换预处理后的板栗近红外光谱进行傅里叶变换, 并用不同方法建模, 提高识别精度。采用试探法提取近红外光谱傅里叶系数, 建立了基于最小二乘支持向量机分类器的霉变板栗识别模型。当提取前 35 点傅里叶系数时, 板栗的平均识别正确率为 93.56%; 构造 GA-LSSVM 算法, 建立的霉变板栗识别模型所用傅里叶系数减少为 13 点, 对测试集中合格板栗、表面霉变板栗和内部霉变板栗的平均识别正确率分别为 95.89%、100% 和 98.25%, 板栗的总体平均识别正确率提高到 97.54%。为霉变板栗的识别提供了快速鉴别分析方法。

**关键词:** 遗传算法 (GA), 识别, 傅里叶变换, 板栗, 近红外光谱, 最小二乘支持向量机 (LSSVM)

doi: 10.3969/j.issn.1002-6819.2011.03.060

中图分类号: O657.33

文献标志码: A

文章编号: 1002-6819(2011)-03-0331-05

周竹, 李小昱, 李培武, 等. 基于 GA-LSSVM 和近红外傅里叶变换的霉变板栗识别 [J]. 农业工程学报, 2011, 27(3): 331-335.

Zhou Zhu, Li Xiaoyu, Li Peiwu, et al. Near-infrared spectral detection of moldy chestnut based on GA-LSSVM and FFT[J]. Transactions of the CSAE, 2011, 27(3): 331-335. (in Chinese with English abstract)

## 0 引言

板栗霉变是造成板栗损失的重要原因。目前剔除霉变板栗主要采用人工分选或盐水浮选, 分选效率低。再加上板栗霉变可发生在栗壳内, 肉眼难以识别, 致使分选错误率高。

近红外光谱定性分析技术是利用近红外光谱区域的全谱或部分波段进行模式识别, 已广泛应用于农副产品成分鉴别、种类鉴定等方面的研究<sup>[1-6]</sup>。本课题组的前期研究表明: 近红外光谱可用于带壳板栗的品质检测<sup>[7-8]</sup>。但近红外光谱通常包含 1 000~2 000 多个波长点的光谱数据, 直接使用会影响识别的速度与效率, 因而需要对光谱数据进行降维。傅里叶变换可将光谱信号从波长域转换到频域, 使用较少的傅里叶系数就能够代表原始光谱信息<sup>[9]</sup>。相比于主成分分析法, 傅里叶变换具有只针对单条光谱产生作用、不受整个样本集光谱数据影响的优点<sup>[10-11]</sup>。傅里叶变换已成功应用于近红外光谱分析领域<sup>[11-14]</sup>。提取傅里叶系数的点数不同会导致所建模型精度的差异, 而目前对傅里叶系数的选择主要采用试探法<sup>[11-12]</sup>。如何优选近红外光谱傅里叶变换后的系数, 从而提高所建模型的精度, 仍是一个值得研究的问题。

最小二乘支持向量机 (least squares support vector

machine, LSSVM) 是建立在 VC 维理论和结构化风险最小原理基础上的一种非线性建模方法, 具有学习速度快, 泛化性好等特点, 能够较好地解决小样本、非线性、高维数和局部极小点等问题<sup>[15]</sup>。遗传算法 (genetic algorithms, GA) 借鉴生物界自然选择和自然遗传机制进行随机搜索, 可用于参数优选<sup>[16]</sup>。本文拟对板栗近红外光谱进行傅里叶变换, 首先采用试探法提取傅里叶系数建立霉变板栗识别模型, 其次使用 GA-LSSVM 算法同时优选傅里叶系数及分类器核参数, 以建立识别精度更高的霉变板栗识别模型。

## 1 GA-LSSVM 算法的实现

在模式识别领域, 特别是对高维数据分类时, 分类器的参数与待分类数据相互影响, 两者共同决定分类效率和精度。因此, 为了获得较好的分类效果, 需要获取待分类数据子集和优化分类器参数。Huang<sup>[17]</sup>等人的研究表明, 当分类器性能最佳时, 待分类数据最优子集和分类器最优参数可同时获得。遗传算法可将待分类数据和分类器参数同时编码, 在合适的适应度函数约束下, 通过随机选择、交叉和变异等遗传操作寻找最优参数。LSSVM 分类器原理可参考文献<sup>[18]</sup>, 其常用的 4 个核函数中, 径向基核函数 (RBF) 应用最广泛, 它有正则化参数  $\gamma$  和核函数参数  $\sigma^2$  等两个参数。本文使用 RBF 核函数作为 LSSVM 分类器的核函数。

采用 GA 算法对特征子集或相关参数进行优化, 一般包括染色体的编码与解码、适应度函数的构建、遗传操作算子的选择、精英保留策略、终止条件设定等 5 个方面。为了实现 GA 算法同时对板栗近红外光谱傅里叶系数及 LSSVM 分类器核参数  $\gamma$  和  $\sigma^2$  进行优化, 需要将它们

收稿日期: 2010-05-07 修订日期: 2010-07-07

基金项目: 高等学校博士学科点专项科研基金 (20090146110018)

作者简介: 周竹 (1984—), 男, 湖北汉川人, 研究方向为智能化检测技术。武汉 华中农业大学工学院, 430070。Email: zhouzhuniba@163.com

\*通信作者: 李小昱 (1953—), 女, 河南叶县人, 教授, 博士生导师, 中国农业工程学会高级会员 (E041200068S), 研究方向为智能化检测技术。

武汉 华中农业大学工学院, 430070。Email: lixiaoyu@mail.hzau.edu.cn

编码在同一染色体中。因此,染色体的编码与解码以及构建合适的适应度函数是实现该算法的核心。

### 1.1 染色体的编码与解码

将板栗近红外光谱傅里叶系数、LSSVM 分类器参数  $\gamma$  和  $\sigma^2$  同时包含在如图 1 所示的染色体中。染色体的第 1 部分为待优化的板栗近红外光谱傅里叶系数,其长度为  $k_1$ ,染色体基因位上的二进制值  $b_i(i=1\cdots k_1)$  反映了傅里叶系数的选择状态。其值为 0 表示该傅里叶系数未被选择,为 1 时表示该位置上的傅里叶系数被选中。第 2、3 部分分别对应分类器参数  $\gamma$ 、 $\sigma^2$  的编码,长度分别为  $k_2$ 、 $k_3$ 。在计算染色体适应度或返回最优解时需要将  $\gamma$  和  $\sigma^2$  对应的基因型转换为表现型,即将其二进制编码解码为实数值。解码函数可表述为

$$P(R, k) = R_{\min} + \left( \sum_{i=1}^k b_i \cdot 2^{k-i} \right) \cdot \frac{R_{\max} - R_{\min}}{2^k - 1} \quad (1)$$

式中,  $P$  为对应参数  $R$  具有  $k$  位基因染色体的实际值,  $R \in \{\gamma, \sigma^2\}$ ,  $k \in \{k_2, k_3\}$ ;  $R_{\min}$ ,  $R_{\max}$  分别对应参数  $R$  的最小值和最大值。

染色体		
傅里叶系数	正则化参数 $\gamma$	核函数参数 $\sigma^2$
$b_1 \dots b_{k_1}$	$c_1 \dots c_{k_2}$	$d_1 \dots d_{k_3}$

注:  $b_i(i=1\cdots k_1)$ 、 $c_j(j=1\cdots k_2)$ 、 $d_m(m=1\cdots k_3)$  分别为板栗近红外光谱傅里叶系数、LSSVM 分类器参数  $\gamma$  和  $\sigma^2$  二进制编码基因位,其值均为 0 或 1

图 1 染色体编码结构

Fig.1 Coding of chromosome

### 1.2 适应度函数的确定

适应度函数是遗传算法中种群进化的依据,对遗传算法的收敛速度、能否找到最优解至关重要。本文采用的适应度函数包括 2 个部分,其表达式如下

$$fitness = c_1 \times accu + c_2 \times zeros \quad (2)$$

式中,  $fitness$  为适应度值;  $accu$  为样本分类正确率;  $zeros$  为未被选择傅里叶系数的个数;  $c_1$ 、 $c_2$  分别表示分类正确率和未参与建模的傅里叶系数个数的权重。 $c_1$ 、 $c_2$  的大小对应了分类正确率与所选特征数目的重要程度。为了使分类正确率最高,同时所选傅里叶系数最少,通常可设置较大的  $c_1$  以及相对小的  $c_2$ 。

## 2 材料与方法

### 2.1 试验材料

试验用板栗为湖北麻城油栗,栗质量范围为 8.10~20.70 g, 购买后按中国商业行业标准 SB/T 10192-1993 规定贮藏。样品分为合格板栗和霉变板栗两类,其中合格板栗 73 个,不同程度的外部霉变、内部霉变板栗分别为 73 个、57 个。外部霉变板栗依 GH/T 1029-2002 判定,内部霉变板栗在光谱采集后去壳确定。

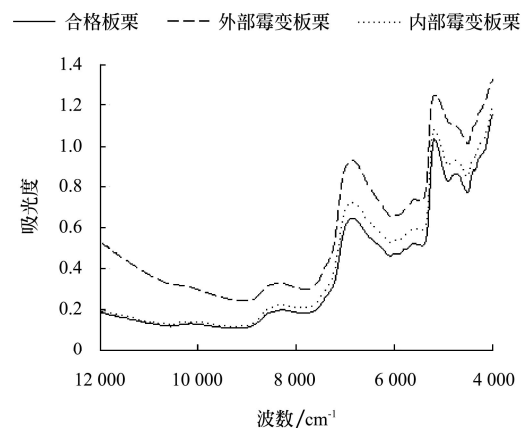
### 2.2 试验装置

板栗近红外光谱采集采用德国 Bruker 公司 FT-2 NIR 近红外光谱仪 (VECTOR 33 型), 配备镀金积分球、样品旋转器、12 mm 石英杯、PbS 检测器和 OPUS 分析软

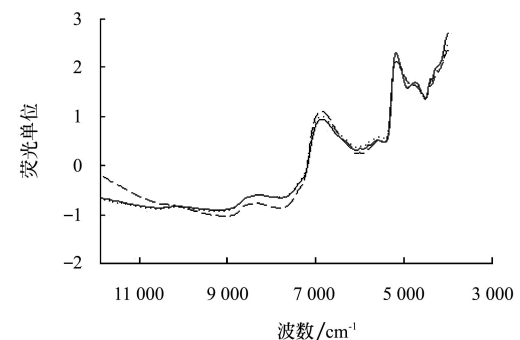
件。

### 2.3 光谱采集及预处理

将完整带壳板栗果实平置于光源上方并完全遮蔽光源进行板栗光谱数据采集。光谱扫描范围: 12 000~4 000  $\text{cm}^{-1}$ ; 扫描次数: 64 次; 采样间隔为 3.856  $\text{cm}^{-1}$ 。73 个合格板栗、73 个外部霉变板栗以及 57 个内部霉变板栗的原始平均光谱如图 2a 所示。为了便于快速傅里叶变换 (光谱点数满足  $2^n$ ), 截取 11 896~4 000  $\text{cm}^{-1}$  内板栗近红外光谱, 共有 2 048 个变量。对光谱进行标准正态变量变换 (SNV), 可消除板栗大小及光谱散射的影响。经 SNV 预处理后的板栗光谱如图 2b 所示。



a. 原始平均光谱



b. SNV 预处理平均光谱

图 2 板栗的原始平均光谱和 SNV 预处理后平均光谱

Fig.2 Chestnut mean spectra from raw data and SNV preprocessed data

### 2.4 板栗光谱的傅里叶变换

McClure<sup>[9]</sup>等人的研究指出,近红外光谱经傅里叶变换后,原始光谱信息主要集中在前 50 点傅里叶系数中,因此较少的傅里叶系数能够代替原始光谱较多波长点数据,从而实现光谱的降维。试验对 203 个板栗的 SNV 预处理光谱进行傅里叶变换,提取前 50 点傅里叶系数。

## 3 结果与分析

### 3.1 基于近红外光谱傅里叶系数的霉变板栗识别模型

以提取的前 50 点傅里叶系数作为输入,板栗是否霉变作为输出,建立基于 LSSVM 分类器的霉变板栗识别模型,并采用 GA 算法优化 LSSVM 分类器参数。

GA-LSSVM 算法的参数设置如下:种群大小 40,染

色体长度 20，其中  $\gamma$ 、 $\sigma^2$  的编码长度均为 10，设置  $\gamma_{\min}=0.1$ ， $\gamma_{\max}=100$ ， $\sigma^2_{\min}=0.01$ ， $\sigma^2_{\max}=20$ ，交叉概率  $P_c=0.6$ ，变异概率  $P_m=0.05$ ，最大进化代数 为 500。采用 5 折交叉验证法来估计 LSSVM 分类器的正确分类率。即将 73 个合格板栗和 130 个霉变板栗平均分成 5 组，每组样本由原始样本中随机抽取形成。试验分 5 次进行，每次将其中的 1 组样本作为测试集，将其余的 4 组样本作为训练集，最后计算出两类板栗的平均正确分类率。由于只需要优选 LSSVM 参数，则设置式 (2) 中  $c_1=1$ ， $c_2=0$ ，即只以分类正确率作为适应度函数。

采用试探法提取前 5、10、15、20、25、30、35、40、45、50 点傅里叶系数并分别建立识别模型。表 1 给出了使用不同点数傅里叶系数建模时霉变板栗的平均识别正确率及对应的最优核参数。从表 1 中可以看出，霉变板栗的平均识别正确率随建模傅里叶系数的点数不同存在差异。使用前 35 点傅里叶系数建模时，模型的平均识别正确率达到 93.55%，为试探法所建模型中识别率最高。当采用的傅里叶系数少于 35 点时，由于提取的傅里叶系数较少，不足以充分反映板栗差异的信息，从而导致识别率不高。而傅里叶系数多于 35 点时，所含板栗差异的信息又过多，引入了干扰信息，降低了识别率。

表 1 不同点数傅里叶系数所建 LSSVM 模型的识别率及参数  
Table 1 Mean discriminating results and parameters based on LSSVM model of different Fourier coefficients

提取点数	$\gamma_{opt}$	$\sigma^2_{opt}$	平均识别正确率/%
5	6.0569	1.202	88.63
10	89.1604	17.831	90.61
15	18.6543	3.7227	89.62
20	65.6258	13.1217	90.1
25	42.6771	8.5297	90.57
30	84.0824	16.8149	92.59
35	44.7279	8.94	93.56
40	53.4191	10.6791	92.57
45	42.9701	8.5883	92.55
50	56.3487	11.2654	92.55

注： $\gamma_{opt}$ 、 $\sigma^2_{opt}$  分别为最优的正则化参数  $\gamma$ 、核函数参数  $\sigma^2$ 。

### 3.2 基于 GA-LSSVM 霉变板栗识别模型的傅里叶系数及分类器参数优选

为了进一步提高模型的精度，对前 50 点傅里叶系数进行筛选，优化傅里叶系数的组合，以解决试探法中傅里叶系数过多或过少所带来的识别率降低问题。

以 50 点傅里叶系数为输入，板栗是否霉变作为输出，建立基于 LSSVM 分类器的霉变板栗识别模型。采用 GA-LSSVM 算法同时优选傅里叶系数及 LSSVM 中的核参数  $\gamma$  和  $\sigma^2$ 。GA-LSSVM 算法的参数设置如下：染色体长度 69，其中傅里叶系数编码长度  $k_1$  为 49， $\gamma$ 、 $\sigma^2$  的编码长度  $k_2$ 、 $k_3$  均为 10，最大进化代数为 500。以 5 折交叉验证平均识别率及未选中傅里叶系数的个数作为适应度值，则设置式 (2) 中  $c_1=10\ 000$ ， $c_2=2$ 。其余参数同 3.1 中所述。

表 2 给出了使用 GA-LSSVM 算法同时优选傅里叶系数及 LSSVM 分类器后所建模型对板栗识别的正确率及核参数。对比表 1 和表 2：使用 GA-LSSVM 同时进行傅里叶系数及 LSSVM 分类器参数优选后，只需要 13 点傅里叶系数，所建模型的平均识别正确率就达 97.54%，与试探法中采用前 35 点傅里叶系数建立的模型相比，建模所用傅里叶系数数目减少了 22 点，总体识别率提高了 3.98%。

表 2 GA-LSSVM 方法获得的平均识别率及参数  
Table 2 Mean discriminating results and parameters based on GA-LSSVM

建模方法	点数	平均识别正确率/%	$\gamma_{opt}$	$\sigma^2_{opt}$
GA-LSSVM	13	97.54	39.4545	7.8848

注： $\gamma_{opt}$ 、 $\sigma^2_{opt}$  分别为最优的正则化参数  $\gamma$ 、核函数参数  $\sigma^2$ 。

### 3.3 2 种模型的比较

以 35 点傅里叶系数作为输入建立的模型为模型 1，以经过 GA-LSSVM 同时优选板栗近红外光谱傅里叶系数和 LSSVM 分类器核参数建立的模型为模型 2。表 3 给出了两个模型在获得最优核参数后，采用 5 折交叉验证法计算出板栗的平均识别正确率。

表 3 两种模型对板栗识别的比较  
Table 3 Detection results of two models

模型	建模时间/h	测试集			训练集		
		合格板栗/%	霉变板栗/%		合格板栗/%	霉变板栗/%	
			表面霉变	内部霉变		表面霉变	内部霉变
1	1.4	90.41	100	89.47	100	100	100
2	9.7	95.89	100	98.25	98.35	100	98.60

从图 2 中可以看出，合格板栗光谱与内部霉变板栗光谱形态相似，会增加识别它们的难度。表 3 中 2 个模型对板栗样本总体识别率的差异也正是体现在对合格板栗及内部霉变板栗的识别上。与模型 1 相比，模型 2 对测试集中合格板栗、内部霉变板栗的识别率分别提高了 5.48%、8.78%。由于模型 2 需要同时对傅里叶系数及 LSSVM 核参数进行优化，其建模所用时间 (9.7 h) 比模型 1 所用时间 (1.4 h) 长。然而所建模型对相同的测试集进行识别时，发现模型 2 的识别时间比模型 1 的识别时间短约 40 ms，这主要是因为模型 2 的输入 (13 点) 要少于模型 1 的输入 (35 点)。

### 3.4 讨论

主成分分析法是一种常用的数据降维方法，本课题组前期对板栗近红外光谱进行主成分分析，提取若干主成分，建立了基于神经网络的霉变板栗识别模型<sup>[7]</sup>。板栗近红外光谱经矢量归一化预处理所建模型对测试集中合格板栗、表面霉变板栗和内部霉变板栗的识别正确率可达 94.74%、94.44% 和 92.31%。本文采用 5 折交叉验证法，相当于测试集的样本增加了 5 倍，对测试集中合格板栗、表面霉变板栗和内部霉变板栗的平均识别正确率分别为 95.89%、100% 和 98.25%。因此，模型的识别正确率和适应性均有提高。此外，主成分分析法获取的主成分数依

赖于整个样本集,而傅里叶变换独立于样本集,只作用于单条板栗光谱。相比于主成分分析法,对板栗近红外光谱进行傅里叶变换,用 GA-LSSVM 算法同时优选板栗近红外光谱傅里叶系数及分类器核参数进而建立模型并用来识别霉变板栗是一种较好的方法。

#### 4 结 论

1) 采用近红外光谱技术判别霉变板栗,针对板栗近红外光谱变量多、共线性强的特点,对板栗近红外光谱进行傅里叶变换。采用试探法提取前 50 点傅里叶系数建立霉变板栗的识别模型。当提取的傅里叶系数点数为 35 时,使用遗传算法只优选 LSSVM 分类器参数,所建模型的平均识别率达到 93.56%,可用于霉变板栗的识别。

2) 构造 GA-LSSVM 算法,该算法中染色体编码由板栗近红外光谱待优化傅里叶系数、LSSVM 分类器参数  $\gamma$ ,  $\sigma^2$  共同组成。使用该算法对板栗近红外光谱傅里叶系数及 LSSVM 分类器参数进行同时优化。从板栗近红外光谱前 50 点傅里叶系数中优选 13 点傅里叶系数用于 LSSVM 分类器建模,模型的平均识别率达到 97.54%。与未优选傅里叶系数建立的模型相比,不仅建模所需傅里叶系数减少,而且对测试集中合格板栗、内部霉变板栗的识别率分别提高了 5.48%、8.78%。

3) 板栗近红外光谱傅里叶变换不依赖于整个样本数据集,只作用于单条板栗光谱曲线。采用 GA-LSSVM 算法同时优选板栗近红外光谱傅里叶系数及分类器核参数并建立霉变板栗识别模型,其识别正确率高、适应性好,是一种较好的霉变板栗识别方法。

4) 使用 GA-LSSVM 算法对板栗近红外光谱傅里叶系数及 LSSVM 分类器参数进行同时优化建模耗时较长,下一步可通过优化算法代码,提高计算机性能从而减少建模时间。

#### [参 考 文 献]

- [1] Xie Lijuan, Ying Yibing, Ying Yitie. Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-squares support vector machines and other chemometrics[J]. *Journal of Food Engineering*, 2009, 94(1): 34—39.
- [2] 周子立,张瑜,何勇,等.基于近红外光谱技术的大米品种快速鉴别方法[J].*农业工程学报*,2009,25(8):131—135.  
Zhou Zili, Zhang Yu, He Yong, et al. Method for rapid discrimination of varieties of rice using visible NIR spectroscopy[J]. *Transactions of the CSAE*, 2009, 25(8): 131—135. (in Chinese with English abstract)
- [3] 杨燕,聂鹏程,杨海清,等.基于可见-近红外光谱技术的蜜源快速识别方法[J].*农业工程学报*,2010,26(3):238~242.  
Yang Yan, Nie Pengcheng, Yang Haiqing, et al. Rapid recognition method of nectar plant based on visible-near infrared spectroscopy[J]. *Transactions of the CSAE*, 2010, 26(3):238—242. (in Chinese with English abstract)
- [4] 郝勇,孙旭东,高荣杰,等.基于可见/近红外光谱与 SIMCA 和 PLS-DA 的脐橙品种识别[J].*农业工程学报*,2010,26(12):373—377.  
Hao Yong, Sun Xudong, Gao Rongjie, et al. Application of visible and near infrared spectroscopy to identification of navel orange varieties using SIMCA and PLS-DA methods[J]. *Transactions of the CSAE*, 2010, 26(12): 373—377. (in Chinese with English abstract)
- [5] Guillén A, Moral F G del, Herrera L J, et al. Using near-infrared spectroscopy in the classification of white and iberian pork with neural networks[J]. *Neural Computing & Applications*, 2010, 19(3): 465—470.
- [6] Sundaram J, Kandala C V K, Butts C L. Classification of in-shell peanut kernels nondestructively using VIS/NIR reflectance spectroscopy[J]. *Sensing and Instrumentation for Food Quality and Safety*, 2010, 4(2): 82—94.
- [7] 周竹,刘洁,李小昱,等.霉变板栗的近红外光谱和神经网络方法判别[J].*农业机械学报*,2009,40(增刊1):109—112.  
Zhou Zhu, Liu Jie, Li Xiaoyu, et al. Discrimination of moldy Chinese chestnut based on artificial neural network and near infrared spectra[J]. *Transactions of the CSAM*, 2009, 40(supp1): 109—112. (in Chinese with English abstract)
- [8] 刘洁,李小昱,李培武,等.基于近红外光谱的板栗水分检测方法[J].*农业工程学报*,2010,26(2):338—341.  
Liu Jie, Li Xiaoyu, Li Peiwu, et al. Determination of moisture in chestnuts using near infrared spectroscopy[J]. *Transactions of the CSAE*, 2010, 26(2): 338—341. (in Chinese with English abstract)
- [9] Burns D A, Ciurczak E W. *Handbook of Near-Infrared Analysis, Third Edition*[M]. New York: CRC Press, 2007: 97—100.
- [10] Kim J, Mowat A, Poole P, et al. Linear and non-linear pattern recognition models for classification of fruit from visible-near infrared spectra[J]. *Chemometrics and Intelligent Laboratory Systems*, 2000, 51(2): 201—216.
- [11] Wu W, Walczak B, Penninckx W, et al. Feature reduction by Fourier transform in pattern recognition of NIR data[J]. *Analytica Chimica Acta*, 1996, 331(1/2):75—83.
- [12] 肖武,李小昱,李培武,等.基于近红外傅里叶特征提取方法的土壤含水率检测[J].*农业机械学报*,2009,40(5):64—67.  
Xiao Wu, Li Xiaoyu, Li Peiwu, et al. Near-infrared spectral detection of soil moisture based on feature extraction of FFT[J]. *Transactions of the CSAM*, 2009, 40(5): 64—67. (in Chinese with English abstract)
- [13] Pasti L, Jouan-Rimbaud D, Massarta D, et al. Application of Fourier transform to multivariate calibration of near-infrared data[J]. *Analytica Chimica Acta*, 1998, 364(1/2/3): 253—263.
- [14] M Hana, W F McClure and T B Whitaker. Applying artificial neural networks. I. Estimating nicotine in tobacco from near infrared data[J]. *Journal of Near Infrared Spectroscopy*, 1995, 3(3): 133—142.
- [15] 陈孝敬,吴迪,虞佳佳,等.一种用于可见-近红外光谱特征波长选择的新方法[J].*光学学报*,2008,28(11):2153—2160.

- Chen Xiaojing, Wu Di, Yu Jiajia, et al. A new choice method of characteristic wavelength of visible/near infrared spectroscopy[J]. *Acta Optica Sinica*, 2008, 28(11): 2153—2160. (in Chinese with English abstract)
- [16] 张豪, 罗亦泳, 张立亭, 等. 基于遗传算法最小二乘支持向量机的耕地变化预测[J]. *农业工程学报*, 2009, 25(7): 226—231.
- Zhang Hao, Luo Yiyong, Zhang Liting, et al. Cultivated land change forecast on genetic algorithm and least squares support vector machines[J]. *Transactions of the CSAE*, 2009, 25(7): 226—231. (in Chinese with English abstract)
- [17] Cheng-Lung Huang, Chieh-Jen Wang. A GA-based feature selection and parameters optimization for support vector machines[J]. *Expert Systems with Applications*, 2006, 31(2): 231—240.
- [18] Suykens J A K, Vandewalle J. Least squares support vector machines classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293—230.

## Near-infrared spectral detection of moldy chestnut based on GA-LSSVM and FFT

Zhou Zhu<sup>1</sup>, Li Xiaoyu<sup>1\*</sup>, Li Peiwu<sup>2</sup>, Gao Yun<sup>1</sup>, Zhan Hui<sup>1</sup>, Liu Jie<sup>1</sup>

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;

2. Oil Crop Research Institute, Chinese Academy of Agricultural Science, Wuhan 430062, China)

**Abstract:** To overcome the shortage of massive data and multi colinearity in near infrared spectroscopy analysis, Fourier transform was employed on chestnut NIR spectra which was preprocessed with standard normal variate (SNV), and different modeling methods were utilized to improve recognition accuracy. Fourier coefficients were extracted from NIR spectra through trial method, and then discrimination models of moldy chestnuts were established based on least squares support vector machine (LSSVM). The highest mean accurate recognition rate of 93.56% was obtained when the first 35 Fourier coefficients were selected, and a hybrid algorithm, GA-LSSVM was developed and used to optimize the number of Fourier coefficients. As a result, the number of Fourier coefficients used for building recognition models was successfully reduced to 13, while the mean accurate recognition rate was raised to 97.54% and the discriminating rates of qualified chestnut, surface moldy chestnut and internal moldy chestnut were 95.89%, 100% and 98.25% for prediction, respectively. It is suggested that near infrared spectroscopy coupled with Fourier transform and GA-LSSVM methods can be used for rapid detection of moldy chestnut.

**Key words:** genetic algorithms (GA), identification, Fourier Transforms, chestnut, near infrared spectroscopy, least squares support vector machine (LSSVM)