

网络出版时间:2013-06-20 16:22

网络出版地址: <http://www.cnki.net/kcms/detail/61.1390.S.20130620.1622.031.html>

基于 GA-LSSVM 的苹果糖度近红外光谱检测

刘燕德, 周延睿

(华东交通大学 机电工程学院, 江西 南昌 330013)

[摘要] **【目的】**结合遗传算法和最小二乘支持向量机(GA-LSSVM), 优化苹果糖度近红外光谱检测的数学模型, 提高模型的检测精度和稳定性。**【方法】**在 GA-LSSVM 模型建立过程中, 采用遗传算法自动获取最小二乘支持向量机的最优参数。**【结果】**相比于偏最小二乘法(PLS)、传统最小二乘支持向量机(LSSVM) 和遗传偏最小二乘法(GA-PLS) 数学模型, GA-LSSVM 法建立的模型预测效果最优, 模型的相关系数为 0.94, 预测均方根误差为 0.32 °Brix。**【结论】**GA 和 LSSVM 相结合的优化方法在提高苹果糖度近红外光谱检测精度和稳定性方面是可行的。

[关键词] 苹果; 糖度检测; 近红外光谱; 遗传算法; 最小二乘支持向量机

[中图分类号] S661.1; O657.33

[文献标志码] A

[文章编号] 1671-9387(2013)07-0229-06

GA-LSSVM based near infrared spectroscopy detection of apple sugar content

LIU Yan-de, ZHOU Yan-rui

(School of Mechatronics Engineering, East China Jiaotong University, Nanchang, Jiangxi 330013, China)

Abstract: **【Objective】**The objective of the present research was to optimize the detection of sugar content in apples for improving the detection precision and robustness using near infrared spectroscopy, combined with genetic algorithms and least squares support vector machine(GA-LSSVM). **【Method】**In the process of establishing GA-LSSVM model, GA method was used to select the optimal parameters of LSSVM automatically. **【Result】**Compared with partial least squares(PLS) model, GA-PLS model and LSSVM model, GA-LSSVM model was more accurate than others. The correlation of predictive model(R_p) was 0.94, and the root mean square error of prediction(RMSEP) was 0.32 °Brix. **【Conclusion】**It was feasible to improve the precision of near infrared spectroscopy detection of apple sugar content by the combination GA and LSSVM.

Key words: apple; sugar content detection; near infrared spectroscopy; genetic algorithms; least squares support vector machines

近红外光谱技术具有非破坏性、检测速度快、样品无需前处理等优点, 所以被广泛用于农产品的品质检测^[1-3]。但是在近红外光谱检测技术的应用过程中, 外界条件如温度、基线漂移、光源稳定性、样品状态等变化, 都会对检测精度产生一定程度的影响, 因此光谱信息变量的筛选和模型的优化, 对于建立

稳健的近红外光谱检测技术数学模型十分重要。目前, 在近红外光谱检测技术中使用的变量筛选方法, 主要有连续投影算法(Successive projections algorithm, SPA)^[4-5]、无信息变量的消除法(Uninformative variable elimination, UVE)^[6]、蒙特卡罗方法(Monte carlo methods, M-C)^[7-8]、间隔偏最小二乘

[收稿日期] 2012-09-11

[基金项目] 科技部农业科技成果转化项目(2011GB2C500008); 赣鄱英才 555 工程领军人才培养计划项目

[作者简介] 刘燕德(1967—), 女, 江西吉安人, 教授, 博士生导师, 主要从事智能无损检测研究。E-mail: jxliuyd@163.com

法(Interval PLS,iPLS)^[9]、遗传算法(Genetic algorithms,GA)^[10]等,其中 GA 是近年来国内外应用较为广泛的一种特征变量筛选方法^[11-12]。光谱检测技术的建模方法有偏最小二乘法(Partial least squares,PLS)、主成分回归法(Principal component regression,PCR)和多元线性回归法(Multiple linear regression,MLR)等,但以上方法都是基于线性回归方式的多元校正方法,对于处理非线性问题效果并不是很理想。非线性模型建立方法主要有神经网络(Artificial neural networks,ANN)、支持向量机(SVM)和核偏最小二乘法(Kernel partial least squares,KPLS)等^[13]。最小二乘支持向量机(Least squares support vector machine,LSSVM)^[14]是一种改进的 SVM 方法,其采用最小二乘线性系统作为损失函数,将求解过程变成了求解一次线性方程组,使求解速度得以加快,并且计算的复杂性得到了降低。

目前,LSSVM 模型的参数多依靠经验人为选取,这给模型的应用和推广带来了不便。本研究拟采用遗传算法对 LSSVM 模型进行参数优化,以提高水果糖度检测模型的预测精度和效率。

1 材料与方法

1.1 试验样品

试验所用 100 个苹果样品购自南昌某水果市

场,将苹果表皮灰尘等清理干净,依次编号,然后在苹果样品的赤道位置上等间距设 3 个不同的标记,作为该样品采集光谱和测定糖度含量标准值的位置。将苹果样品在温度为 15 ℃、相对湿度为 60% 的环境下储藏 24 h。选取前 75 个样品作为校正集,剩余的 25 个样品作为验证集。

1.2 试验装置及光谱采集

近红外光谱采集试验装置如图 1 和图 2 所示。样品透射光谱的采集采用美国海洋公司的 QE65000 光谱仪,光谱数据的储存和转化采用此光谱仪自身携带的软件 SpectraSuite(Ocean Optics INC.,USA)。光源采用 4 盏欧司朗直流卤素光源(12 V,100 W),每边 2 盏对称布置在被测样品的两侧,苹果样品放置在 4 盏光源的交汇处,保持苹果样品的果梗与水平面平行。水果样品下面放置 1 个厚 3 mm、直径 50 mm 的圆形软橡胶垫,软橡胶垫圆心处有 1 个直径为 10 mm 的圆形孔,透过苹果样品的光通过这个孔到达光纤探头。在光纤上靠近苹果样品一端加装光纤探头(74-UV,直径 5 mm,量程 200~2 000 nm,Ocean Optics INC.,USA)。试验中采用空气作为参比。光谱仪积分时间设置为 100 ms,平滑度设置为 15,平均数设置为 1。分别采集苹果样品赤道位置上 3 个不同标记处的光谱,然后求其平均值作为该样品的光谱值。

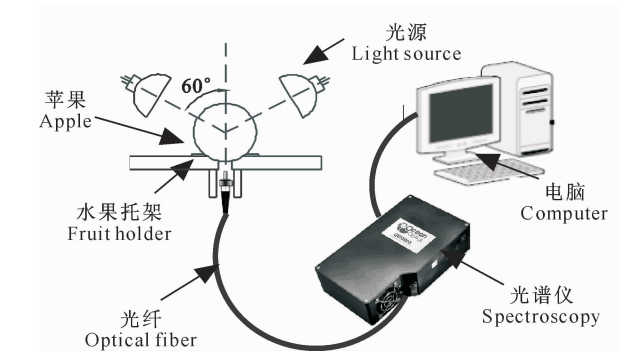


图 1 近红外光谱采集试验装置示意图
Fig.1 Schematic diagram of NIRs acquisition device

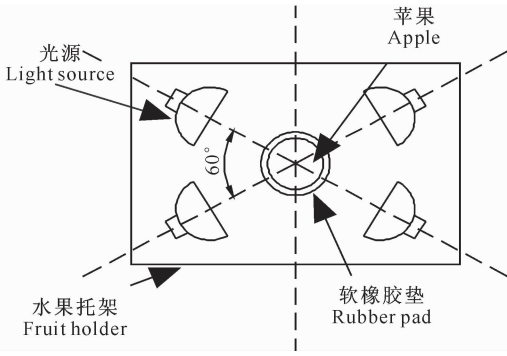


图 2 光源和苹果样品的俯视图
Fig.2 Top view of light source and apples

1.3 苹果糖度标准值的测定

采集苹果样品的透射光谱值后,立即测定苹果样品的糖度。将苹果样品上标好位置的果肉切下并榨汁,将 2~3 滴榨取的果汁,滴于数字式折射仪(PR-101 Cat. No3442,ATGO,Japan)的折光棱镜表面中央,对苹果样品进行糖度的测定,取苹果样品赤道位置上 3 个不同标记处糖度的平均值作为标准

值。表 1 列出了苹果样品糖度的变化范围、平均值、标准差及变异系数,由此可知苹果样品校正集所建立的模型能较好地适用于预测集样品。

1.4 GA-LSSVM 算法的实现

LSSVM 算法是对经典 SVM 算法的改进,其采用最小二乘线性系统作为损失函数,将不等式约束变成等式约束,并将经验风险由误差的一范数改为

二范数,也就将求解二次优化问题转化为求解一次求解速度^[15]。
线性方程组问题,从而降低了计算的复杂性,加快了

表 1 苹果样品糖度实测值的统计

Table 1 The statistics of sugar content for calibration and prediction sets

样品集 Sample set	样品数量 Quantity	糖度/(°Brix) Sugar content				变异系数/% Coefficient of variation
		最大值 Max	最小值 Min	平均值 Mean	标准差 Standard deviation	
校正集 Calibration set	75	15.8	10.27	12.89	1.06	8.26
验证集 Prediction set	25	14.6	11.13	12.83	0.92	7.21

在 LSSVM 中,核函数参数 σ^2 、正则化参数 γ 的选取对模型的精度有重要影响,因而是要提高 LSSVM 模型的精度和稳定性,必须正确优化调整这些参数。目前,LSSVM 的参数选取通常采用参数空间穷尽搜索法,这种方法的缺点是经验性较强,在较大范围内得到最优参数的效率低。遗传算法不需要明确目标函数的数学方程和倒数表达式,并克服了传统优化方法较易陷入局部极值的缺点,是一种全局寻优算法,寻优效率较高。使用遗传算法对 LSSVM 模型进行参数优化的基本步骤如下;

Step 1:设置初始参数,即设置初始种群规模、进化代数、交叉概率和变异概率等。

Step 2:群体初始化,对参数的二进制代码进行优化,随机产生一个给定大小的初始群体。

Step 3:对种群的每个个体进行解码,利用 LSSVM 算法计算预测值,并计算种群中每个个体的适应度。

Step 4:根据 GA 适应度,对解群体进行复制、变异、交叉操作,生成下一代群体。

Step 5:若满足 GA 训练停止条件(训练误差 ≤ 0.001 或达到迭代次数),则停止训练并转向步骤 6,否则转向步骤 3。

Step 6:结束遗传训练,得到最优的 LSSVM 参数(γ, σ^2),并建立 GA-LSSVM 模型。

2 结果与分析

2.1 有效光谱波段的选择

首先使用比尔-朗伯定律将苹果样品的透射光谱全部转换为吸收光谱。为提高模型建立过程中数据处理的效率和所选择的光谱波段的有效性,首先应去除具有明显噪声及不含可溶性固形物特征光谱的波段。然后再结合有关研究,采用光谱范围为 500~1 050 nm 的波段进行数据处理和建模^[16]。苹果样品的吸收光谱如图 3 所示。由图 3 可以看出,

光谱波段 500~1 050 nm 中包含了有关苹果样品中可溶性固形物的重要信息,并且含有相对较少的光谱噪声。

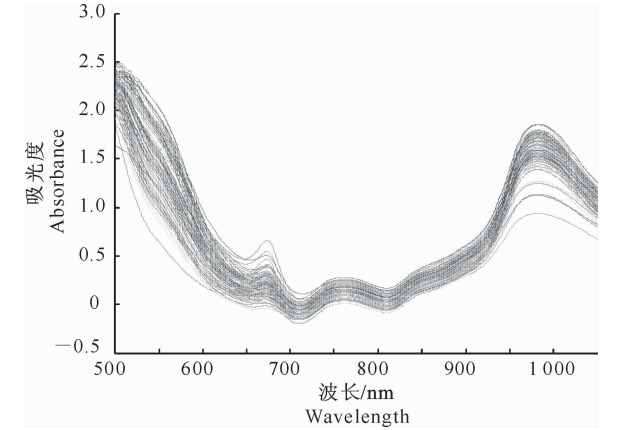


图 3 光谱范围为 500~1 050 nm 的苹果样品的吸收光谱
Fig. 3 Absorbance spectra plot based on the selective transmittance wavelength(500—1 050 nm)

2.2 光谱数据的预处理

在采集近红外光谱时,有许多高频随机噪声、基线漂移、样本颗粒大小和光散射等噪声信息的干扰,这将影响近红外光谱与样品内有效成分含量间的关系,从而影响所建立模型的预测精度和稳定性^[17]。为了提高模型的预测精度和稳定性,在建模前使用 Unscramble 8. 0(CAMO AS, Trondheim, Norway)对光谱数据进行 2 种预处理:(1)为了去除光谱信息中叠加的随机误差对模型精度的影响,采用 Savitsky-Golay 的卷积平滑法对光谱进行平滑处理,平滑点数选择 5。(2)为了有效消除基线和其他背景的干扰,提高分辨率和灵敏度,采用 Savitsky-Golay 的卷积求导法,对光谱分别进行二阶微分,微分点数选择 15。

2.3 GA-LSSVM 参数优化

LSSVM 中核函数参数 σ^2 和正则化参数 γ 选择得是否合理将直接决定模型性能的优劣,如果 σ^2 过小则会对样本数据产生过学习现象;如果 σ^2 过大则会对样本数据造成欠学习现象。本研究采用遗传算

法对 LSSVM 模型的参数进行优化选择,设置种群规模为 30,进化代数 为 100,交叉概率为 0.5,变异概率为 0.01,独立运行 100 次。经过 5 次重复遗传训练后,选取每次遗传训练后频率较高的波长点作为 LSSVM 模型的输入变量。

将 GA 分析中得到的 102 个波长下的光谱值作为 LSSVM 模型的输入变量,选用高斯径向基函数(RBF)为核函数,对核函数参数 σ^2 和正规化参数 γ ,

将二步格点搜索法与 L-折交互验证法($L=10$)相结合,通过粗选和精选 2 个步骤进行优化,第 1 次参数寻优过程如图 4 所示,其中网格点“·”反映了第 1 步格点搜索的范围及步长,曲线反映了误差等高线,网络点“×”反映了第 2 步格点搜索的范围及步长^[18],参数优化后的建模结果如表 2 所示。经比较,得最优参数 $\gamma=1\ 451.599\ 5, \sigma^2=40\ 921.717\ 9$ 。

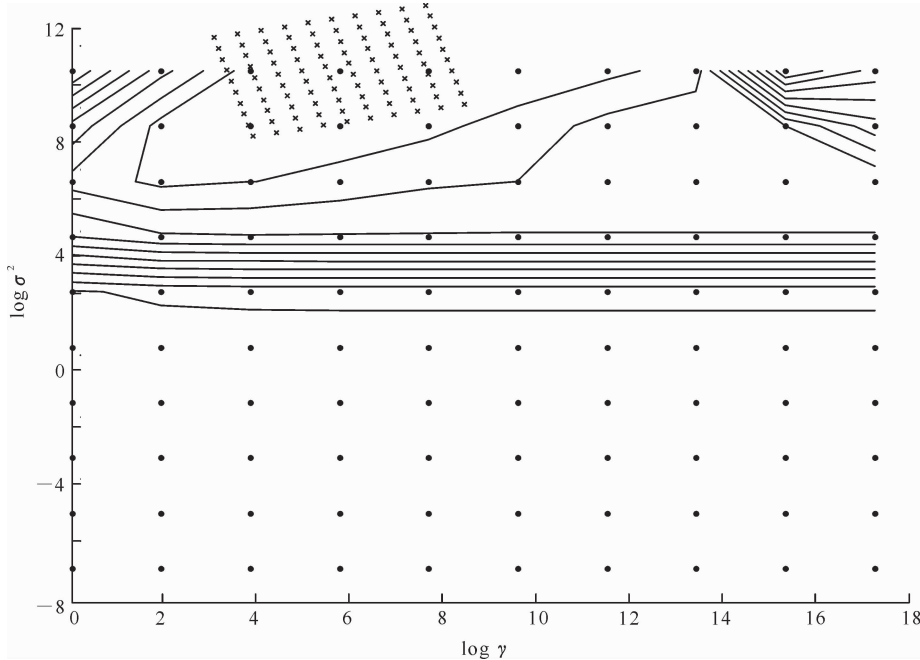


图 4 GA-LSSVM 模型参数 γ 和 σ^2 的第 1 次优化过程

Fig. 4 The first optimization of σ^2 and γ for GA-LSSVM model

表 2 GA-LSSVM 模型各次参数优化后的建模结果

Table 2 Results of GA-LSSVM model for each parameters optimization

优化次数 Optimization time	γ	σ^2	校正集 相关系数 Correlation of calibration model	校正集均方根 误差/(°Brix) Root mean square error of calibration	验证集 相关系数 Correlation of prediction model	验证集均方根 误差/(°Brix) Root mean square error of prediction
1	1 451.599 5	40 921.717 9	0.97	0.26	0.94	0.32
2	1 037.494 1	58 625.336 8	0.97	0.27	0.93	0.34
3	1 443.611 1	78 389.623 0	0.97	0.25	0.93	0.34
4	1 786.991 9	57 572.401 5	0.98	0.24	0.93	0.33
5	1 557.482 8	62 853.063 7	0.97	0.24	0.93	0.35

2.4 多元校正方法的比较

采用经过预处理后的光谱分别建立 PLS、GA-PLS、LSSVM 与 GA-LSSVM 模型。其中在 GA-PLS 模型中,GA 的参数设置为:初始种群规模为 30,进化代数为 100,交叉概率为 0.5,变异概率为 0.01,独立运行 100 次。

经过 5 次重复遗传训练后,建立 PLS 模型;在

LSSVM 模型中,采用人工多次搜索,人为选取核函数参数和正则化参数,根据误差结果选择最优参数。以上 4 种模型的性能对比如表 3 所示。从表 3 可以看出,应用 GA-LSSVM 建模得到的结果为 $R_c=0.97, RMSEC=0.26, R_p=0.94, RMSEP=0.32$,校正集和验证集的相关系数均较高,并且差值较小;校正集和验证集的均方根误差均较小,并且二者差

值非常接近。由此可知,在对苹果糖度进行预测的过程中,GA-LSSVM 模型优于 PLS、GA-PLS 与 LSSVM 等模型。图 5 为利用 GA-LSSVM 模型得

到的苹果糖度预测值与标准值的相关关系图。由图 5 可见,模型的预测能力较好。

表 3 4 种建模方法所建立模型的性能比较
Table 3 Results of PLS,GA-PLS,LSSVM and GA-LSSVM

建模方法 Methods	校正集 相关系数 Correlation of calibration model	校正集均方根 误差/(°Brix) Root mean square error of calibration	验证集 相关系数 Correlation of prediction model	验证集均方根 误差/(°Brix) Root mean square error of prediction
PLS	0.96	0.31	0.90	0.47
GA-PLS	0.91	0.45	0.93	0.39
LSSVM	0.98	0.59	0.91	0.39
GA-LSSVM	0.97	0.26	0.94	0.32

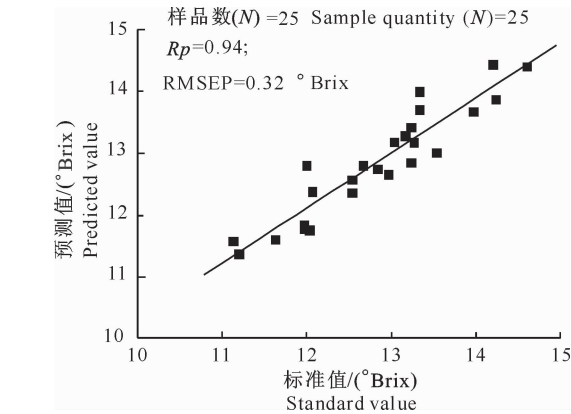


图 5 苹果糖度 GA-LSSVM 模型预测值与标准值的相关关系

Fig.5 Scatter plot of GA-LSSVM prediction model

3 结 论

遗传算法在组合优化问题上具有很大的搜索优势,适应能力较强。本研究利用遗传算法的全局寻优功能,对 LSSVM 进行参数的寻优,并建立了基于 GA-LSSVM 算法的苹果糖度预测模型。模型校正集相关系数为 0.97,均方根误差为 0.26;验证集相关系数为 0.94,均方根误差为 0.32。与 PLS、GA-PLS、LSSVM 模型相比,在采用相同试验样本和相同预处理方法的条件下,GA-LSSVM 模型提高了预测精度,降低了预测误差率,减少了建模时间。该结果表明,在使用近红外光谱分析技术检测苹果糖度的过程中,利用遗传算法对 LSSVM 模型进行参数优化,可以有效提高模型的可靠性和建模效率,具有很好的实用价值和应用前景。

[参考文献]

[1] 陆婉珍. 现代近红外光谱分析技术 [M]. 北京:中国石化出版社,2007:9-11.
Lu W Z. Modern near infrared spectroscopy analytical technolo-

gy [M]. Beijing:China Petrochemical Press,2007:9-11. (in Chinese)

[2] 韩东海,王加华. 水果内部品质近红外光谱无损检测研究进展 [J]. 中国激光,2008,35(8):1123-1131.
Han D H, Wang J H. Review of nondestructive measurement of fruit quality by means of near infrared spectroscopy [J]. Chinese Journal of Lasers,2008,35(8):1123-1131. (in Chinese)

[3] 付兴虎,付广伟,毕卫红. 近红外光谱技术在水果品质无损检测中应用的研究与现状 [J]. 红外,2006,27(2):33-37,48.
Fu X H, Fu G W, Bi W H. Research of application of NIRS technology in nondestructive detection of quality of fruits [J]. Infrared,2006,27(2):33-37,48(in Chinese)

[4] Fernandes D D S, Gomes A A, Costa G B, et al. Determination of biodiesel content in biodiesel/diesel blends using NIR and visible spectroscopy with variable selection [J]. Talant, 2011, 87:30-34.

[5] Moreira E D, Pontes M J, Galvão R K, et al. Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection [J]. Talanta, 2009, 79(5):1260-1264.

[6] Chen D, Cai W S, Shao X G. Removing uncertain variables based on ensemble partial least squares [J]. Analytica Chimica Acta, 2007, 598(1):19-26.

[7] Xu H, Liu Z C, Cai W S, et al. A wavelengths selection method based on randomization test for near-infrared spectral analysis [J]. Chemometrics and Intelligent Laboratory Systems, 2009, 97(2):189-193.

[8] 洪明坚,温志渝. 一种多模型融合的近红外波长选择算法 [J]. 光谱学与光谱分析, 2010, 30(8):2088-2092.
Hong M J, Wen Z Y. A new wavelength selection algorithm based on the fusion of multiple models [J]. Spectroscopy and Spectral Analysis, 2010, 30(8):2088-2092. (in Chinese)

[9] Zou X B, Zhao J W, Huang X Y, et al. Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of ‘Fuji’ apple based on different PLS models [J]. Chemometrics and Intelligent Laboratory Systems, 2007, 87(1):43-51.

[10] Jiang H, Liu G H, Mei C L. Rapid determination of pH in solid-state fermentation of wheat straw by FT-NIR spectroscopy

- and efficient wavelengths selection [J]. Analytical and Bioanalytical Chemistry, 2012, 404(2): 603-611.
- [11] 祝诗平, 王一鸣, 张小超, 等. 基于遗传算法的近红外光谱谱区选择方法 [J]. 农业机械学报, 2004, 35(5): 152-156.
Zhu S P, Wang Y M, Zhang X C, et al. Region selection method of near infrared spectrum based on genetic algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery, 2004, 35(5): 152-156. (in Chinese)
- [12] 褚小立, 袁洪福, 王艳斌, 等. 遗传算法用于偏最小二乘方法建模中的变量筛选 [J]. 分析化学, 2001, 29(4): 437-442.
Chu X L, Yuan H F, Wang Y B, et al. Variable selection for partial least squares modeling by genetic algorithms [J]. Chinese Journal of Analytical Chemistry, 2001, 29(4): 437-442. (in Chinese)
- [13] 褚小立. 化学计量学方法与分子光谱分析技术 [M]. 北京: 化学工业出版社, 2011: 61-83.
Chu X L. Molecular spectroscopy analytical technology combined with chemometrics and its application [M]. Beijing: Chemical Industry Press, 2011: 61-83. (in Chinese)
- [14] Liu L J, He D Y, Yang S M, et al. Applying chemometrics approaches to model and predict the binding affinities between the human amphiphsin SH3 domain and its peptide ligands [J]. Protein and Peptide Letters, 2010, 17(2): 246-253.
- [15] 孙 俊, 毛罕平, 羊一清. 基于 GA-LS-SVM 的水稻叶片含氮率预测 [J]. 江苏大学学报: 自然科学版, 2010, 31(1): 6-10.
Sun J, Mao H P, Yang Y Q. Prediction of nitrogen content rate of paddy rice leaf based on GA-LS-SVM [J]. Journal of Jiangsu University: Natural Science Edition, 2010, 31(1): 6-10. (in Chinese)
- [16] Fan G Q, Zha J W, Du R, et al. Determination of soluble solids and firmness of apples by Vis/NIR transmittance [J]. Journal of Food Engineering, 2009, 93(4): 416-420.
- [17] 郭志明, 陈立平, 黄文倩, 等. 近红外光谱结合 GA-LSSVR 分析烟草尼古丁含量 [J]. 激光与光电子学进展, 2012, 49(2): 63-67.
Guo Z M, Chen L P, Huang W Q, et al. Application of genetic algorithm-least squares support vector regression with near infrared spectroscopy for prediction of nicotine content in tobacco [J]. Laser & Optoelectronics Progress, 2012, 49(2): 63-67. (in Chinese)
- [18] 徐惠荣, 陈晓伟, 应义斌. 基于多元校正法的香梨糖度可见/近红外光谱检测 [J]. 农业机械学报, 2010, 41(12): 126-129.
Xu H R, Chen X W, Ying Y B. Multivariate approach to determine sugar content of fragrant pears with temperature variation by Visible/NIR spectroscopy [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(12): 126-129. (in Chinese)

(上接第 228 页)

- [12] 张丽萍, 张兴昌, 孙 强. SSA 土壤固化剂对黄土击实、抗剪及渗透特性的影响 [J]. 农业工程学报, 2009, 25(7): 45-49.
Zhang L P, Zhang X C, Sun Q. Effects of SSA soil stabilizer on compaction, shear strength and permeability characteristics of loess [J]. Transactions of the Chinese Society of Agricultural Engineering, 2009, 25(7): 45-49. (in Chinese)
- [13] 樊恒辉, 吴普特, 高建恩, 等. 固化土集流面无侧限抗压强度影响因素研究 [J]. 农业工程学报, 2006, 22(9): 11-15.
Fan H H, Wu P T, Gao J E, et al. Principal factors in unconfined strength of the solidified soil for the catchments area [J]. Transactions of the Chinese Society of Agricultural Engineering, 2006, 22(9): 11-15. (in Chinese)
- [14] 樊恒辉, 高建恩, 吴普特, 等. 土壤固化剂集流面不同施工工艺比较 [J]. 农业工程学报, 2006, 22(10): 73-77.
Fan H H, Gao J E, Wu P T, et al. Comparison of different construction techniques for the catchments area with the soil stabilizer [J]. Transactions of the Chinese Society of Agricultural Engineering, 2006, 22(10): 73-77. (in Chinese)
- [15] 高建恩, 孙胜利, 吴普特. 一种新型土壤固化剂: 中国, CN 200410073273. 5 [P]. 2005-06-29.
Gao J E, Sun S L, Wu P T. A new kind of soil stabilizer: Chinese, CN200410073273. 5 [P]. 2005-06-29. (in Chinese)
- [16] 中国建筑科学研究院. CECS 03—2007 钻芯法检测混凝土强度技术规程 [S]. 北京: 中国工程建设标准化协会, 2008.
China Academy of Building Research. CECS 03—2007 Technical specification for testing concrete strength with drilled core [S]. Beijing: China Committee for Engineering Construction Standardization, 2008. (in Chinese)
- [17] 交通部公路科学研究院. JTGE 60—2008 公路路基路面现场测试规程 [S]. 北京: 人民交通出版社, 2008.
Research Institute of Highway Ministry of Transport. JTGE 60—2008 Field test methods of subgrade and pavement for highway engineering [S]. Beijing: China Communication Press, 2008. (in Chinese)
- [18] 南京水利科学研究院. DL/T 5151—2001 水工混凝土砂石骨料试验规程 [S]. 北京: 中国电力出版社, 2002.
Nanjing Hydraulic Research Institute. DL/T 5151—2001 Test code for aggregates of hydraulic concrete [S]. Beijing: China Electric Power Press, 2002. (in Chinese)