

# 基于支持向量机模型和图像处理技术的甜椒叶面积测定

宰松梅, 温季<sup>\*</sup>, 郭冬冬, 韩启彪, 邓忠, 孙浩, 赵东彬

(中国农业科学院农田灌溉研究所, 新乡 453003)

**摘要:** 叶片是作物进行光合作用的重要器官, 是研究作物对光能吸收的一个主要的生物学指标。应用支持向量机理论, 建立了支持向量机叶面积模型, 输入参数为叶片长度、叶片最大宽度, 输出参数为叶面积。通过对应用计算机图像处理技术测量所得到的样本数据进行训练, 以叶片的长度、宽度作为输入参数对叶面积进行模拟及检验, 并将模拟结果与线性回归和人工神经网络模型进行了对比。结果表明, 支持向量机叶面积模型的最大误差为 6.09%, 平均误差为 2.73%, 模拟精度达到 0.996。该方法能较真实地反映甜椒叶面积的实际大小, 具有较好的实用价值和应用前景。

**关键词:** 支持向量机, 模型, 图像处理, 甜椒, 叶面积, 测定

doi: 10.3969/j.issn.1002-6819.2011.03.045

中图分类号: S641.3, O234

文献标志码: A

文章编号: 1002-6819(2011)-03-0237-05

宰松梅, 温季, 郭冬冬, 等. 基于支持向量机模型和图像处理技术的甜椒叶面积测定[J]. 农业工程学报, 2011, 27(3): 237-241.

Zai Songmei, Wen Ji, Guo Dongdong, et al. Determination of leaf area of sweet pepper based on support vector machine model and image processing[J]. Transactions of the CSAE, 2011, 27(3): 237-241. (in Chinese with English abstract)

## 0 引言

叶片作为植物进行光合和蒸腾作用的主要器官, 其发育状况对作物的生长发育及产量具有很大的影响<sup>[1]</sup>。叶面积大小直接影响作物的受光, 是表征作物对光能吸收的一个重要的生物学指标<sup>[2]</sup>, 是众多关于估算植物特征参数模型输入或输出的重要参数, 常用在植物生物生化、生理生态、作物栽培管理与育种等研究和应用方面, 因此, 精确、快速、经济地测定叶片面积具有重要意义<sup>[3]</sup>。多年来, 为探讨叶面积快捷、精确的测量方法, 人们对植物叶面积的测定方法进行了大量的研究, 提出了多种的方法: 叶面积仪器法虽然测量速度快, 但当植叶片颜色较深时, 测量值偏大, 反之, 偏小, 不规则的叶片边缘也会造成较大的测量误差; 求积仪方法操作简单, 使用方便, 该方法的误差主要来自机械本身误差和叶片形状, 叶子的长宽比越大, 误差就越大, 周长越大, 误差也越大<sup>[4]</sup>; 方格纸法不仅测定速度慢, 叶片需离体测量, 而且要求设置合理的取舍比例<sup>[5]</sup>; 打孔称重法的测量误差主要受叶片的大小、厚薄、叶龄、打孔方式以及叶片含水量的影响<sup>[6]</sup>; 纸重法虽然不受叶片含水率、厚薄、叶龄

的影响, 但工作量大, 且需要精确的称量设备; 抛物线法和系数法的误差大; 回归方程法在实际使用时, 只需测量叶宽和叶长, 但测量精度偏低; 图像处理法建立在计算机图像处理的基础之上, 原理为计算机中的平面图像是由若干个网状排列的像素组成的, 通过分辨率计算出每个像素的面积, 然后统计叶片图像所占的像素个数, 再乘以单个像素的面积就可以得到叶面积<sup>[7-8]</sup>, 但每次都需要叶片离体测量, 且计算工作量大; 神经网络常用来模拟、处理影响因素多、关系复杂的系统, 为高度非线性动态关系的时间序列预测提供了一条有效途径<sup>[9]</sup>, 近年来, 人工神经网络因其固有的学习能力和适应能力, 开始在作物叶面积测量中得以应用<sup>[10]</sup>。但在实际应用中, 人工神经网络需要大量的样本用于训练, 且容易出现过拟合现象, 模型的泛化能力也较差<sup>[11]</sup>。

Vapnik 提出的支持向量机 (support vector machine, SVM) 理论数学推导严谨, 理论基础坚实, 并具有全局最优和较好的泛化能力, 在模式分类、回归分析、函数逼近、时间序列预测、数据挖掘等研究领域模式识别和回归建模中得到了广泛的应用<sup>[12]</sup>。利用支持向量机模型的植物叶面积测定技术, 目前还未见报导。针对上述测定方法的不足, 首先应用数字图像技术对待模拟叶片的长度、宽度和面积进行精确测定, 然后利用支持向量机理论, 分别以叶片的长度、宽度作为输入参数进行模拟, 并与线性回归和人工神经网络方法的测定结果进行对比。结果表明, 该方法具有较好的模拟效果。

## 1 样品采集及分析方法

### 1.1 样品采集

供试叶片样本采自中国农业科学院农田灌溉研究所节水灌溉试验场, 甜椒品种为江苏省农业科学院蔬菜研

收稿日期: 2010-10-18 修订日期: 2011-02-21

基金项目: 国家 863 计划课题 (2006AA100213); 农业科技成果转化资金项目 (2009GB23320465; 2010GB23260584)

作者简介: 宰松梅 (1970—), 女, 河南南阳人, 副研究员, 博士, 中国农业工程学会会员 (E041200421S), 主要从事节水灌溉与环境方面的研究。新乡 中国农业科学院农田灌溉研究所, 453003。

Email: ggszsm@yahoo.com.cn

\*通信作者: 温季 (1960—), 男, 研究员, 博士, 中国农业工程学会会员 (E041200416S), 主要从事节水灌溉技术与中低产田综合治理方面的研究。新乡 中国农业科学院农田灌溉研究所, 453003。

Email: wenji60@yahoo.com.cn

究所培育的早丰1号,株高70~75 cm,开展度75~80 cm,果实形状为不规则灯笼形。在每株甜椒的东西南北中5个部位的发育枝上,随机采取成熟叶片各2片,每株10片。共在13株甜椒上采摘了130片叶子。

## 1.2 分析方法

叶面积的正确测量是研究叶面积的前提,建立在计算机图像处理基础之上叶面积测定具有严密的科学性<sup>[13]</sup>,它利用计算机中的平面图像是由若干个网状排列的像素组成,通过分辨率计算每个像素的面积,统计叶片图像所占的像素个数,再乘以单个像素的面积就可以得到叶面积。为防止叶片因蒸腾水分变形造成测量误差,将采集的叶片装入保鲜袋带回实验室,并马上进行计算机图像扫描处理。

所需基本设备包括硬件(DELL755 台式电脑、HP2820 扫描仪)和软件(Photoshop cs2 等)。

### 1.2.1 叶片数字图像获取

选取青椒叶片若干,整齐地平放在扫描仪工作台面上,并确保各叶片相互无重叠和卷曲。设定扫描分辨率为300 dpi、缩放比例为100%、扫描方式为24位真彩图、RGB格式进行扫描,并以JPG格式存储于计算机中。利用“磁性套索工具”对图像中的单个叶片进行描边工作,目的是保证选中的叶片不包括多余的像素。在对叶片的选择工作完成后,新建一个图像文件,以白色作为背景色,将所选择的图像粘贴到新建的图像文件中,另外命名并存储到计算机中,重复这一步骤,直到所有的叶片都处理完成。

### 1.2.2 叶片长度和宽度的获取

用Photoshop打开分割后的叶片图像文件,然后利用“编辑”>“自由变换”命令对图像进行旋转,将叶片的长轴旋转至垂直方向,由于在扫描设置中选用的缩放比例为100%,所以此时图像的高度和宽度即为叶片的长度和宽度<sup>[14]</sup>。利用“图像”-“图像大小”命令即可取得叶片的长度和宽度。

### 1.2.3 叶面积的获取

执行Photoshop中的“选择”>“色彩范围”命令,设定所要选取的颜色为“高光”,然后执行“选择”>“反向”命令,这样选择的区域就变为叶片<sup>[15]</sup>。执行“窗口>直方图”命令。点击“直方圈”面板右上的小箭头,勾选“扩展视图”和“显示统计数据”,然后点按“不使用高速缓存的刷新”按钮,记下“像素”后的数据,即可通过简单计算得到精确的叶面积。

## 2 叶面积 SVM 模型

神经网络虽然能较好地应用于高度非线性对象建模<sup>[16]</sup>。但神经网络技术存在一些不易解决的难题,如难以确定神经网络的隐层节点数、存在过学习现象、训练过程中存在局部极小问题等<sup>[17]</sup>。Vapnik提出的支持向量机是依据统计学习理论和结构风险最小化原理的新型学习机<sup>[18]</sup>,与神经网络相比,它的优点是训练算法中不存在局部极小问题,可以自动设计模型复杂度,不存在维数灾难问题,泛化能力强。SVM近来已被成功地用来解决

函数估计问题。尤其适合于样本有限,且高维模式问题。支持向量机方法的几个主要优点有:1)针对有限样本情况,得到现有信息下的最优解;2)算法为一个二次型寻优问题,得到的结果是全局最优点,避免了局部极值问题;3)算法复杂度与样本维数无关。

### 2.1 输入参数

影响叶面积的参数主要为:叶片长度、叶片宽度。研究表明,二者与叶面积均有较高的相关度。文中用叶长和叶宽作为输入参数,叶面积作为输出参数,并对上述两输入参数的测量精度进行分析对比。

### 2.2 模型算法

训练样本集假定为 $\{(x_i, y_i), i=1, 2, \dots, l\}$ ,其中输入值 $x_i \in R^n, y_i \in R$ 为对应的目标值, $l$ 为样本数。 $\varepsilon$ 为不敏感损失函数

$$\varepsilon(f(x_i) - y_i) = \begin{cases} 0, & |f(x_i) - y_i| \leq \varepsilon \\ |f(x_i) - y_i| - \varepsilon, & |f(x_i) - y_i| > \varepsilon \end{cases} \quad (1)$$

式中, $f(x)$ 为回归估计函数, $y$ 为目标值, $\varepsilon$ 为设计参数, $\varepsilon > 0$ 。学习的目的是构造 $f(x)$ ,使 $f(x)$ 与目标值之间的距离小于 $\varepsilon$ ,同时函数的VC维最小,这样对于未知样本 $x$ ,可最优地估计出对应的目标值。

对于训练集为非线性情况时,支持向量机通过某一非线性函数 $\Phi(\cdot)$ ,将训练数据 $x$ 映射到一个高维(可能为无穷维)线性特征空间中,构造回归估计函数,这种非线性变换通过定义适当的核函数 $K(x, x')$ 来实现,其中 $K(x, x') = \Phi(x)\Phi(x')$ 。估计函数 $f(x)$ 为

$$f(x) = w\phi(x) + b \quad (2)$$

式中, $w$ 为超平面的法向量; $\phi(x)$ 为核空间映射函数; $b$ 为偏置量。最优化问题为

$$\begin{aligned} \min & \left[ \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) \right] \\ \text{s.t.} & y_i - w\phi(x_i) - b \leq \varepsilon + \xi_i \end{aligned} \quad (3)$$

或

$$\begin{aligned} w\phi(x_i) + b - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i & \geq 0, \xi_i^* \geq 0, i=1, 2, \dots, l \end{aligned}$$

式中, $c$ 为惩罚系数( $c > 0$ ), $c$ 越大表示对超出 $\varepsilon$ 管道数据点的惩罚越大; $\xi_i, \xi_i^*$ 为松弛变量。采用拉格朗日乘子法求解这个具有线性不等式约束优化问题。

$$\begin{aligned} \max & \left[ -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i - a_i^*)(a_j - a_j^*)k(x_i \cdot x_j) - \varepsilon \sum_{i=1}^l (a_i + a_i^*) + \sum_{i=1}^l y_i (a_i - a_i^*) \right] \\ \text{s.t.} & \sum_{i=1}^l (a_i - a_i^*) = 0 \\ & 0 \leq a_i \leq c \\ & 0 \leq a_i^* \leq c \end{aligned} \quad (4)$$

式中, $a_i, a_i^*$ 为拉格朗日乘子。由最优化理论知, $a_i \times a_i^* = 0, 0 \leq a_i \leq c, 0 \leq a_i^* \leq c$ ,则 $a_i, a_i^*$ 非同时为0时所对应的 $x_i$

为支持向量 (support vector, 简称 SV); 而  $0 < a_i < c$ ,  $a_i^* = 0$ ;  $a_i = 0$ ,  $0 < a_i^* < c$  所对应的  $x_i$  为标准支持向量。

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) k(x_i \cdot x) + b \quad i=1, \dots, l \quad (5)$$

式中,

$$b = \frac{1}{N_{NSV}} \left\{ \sum_{0 < a_i < c} \left[ y_i - \sum_{x_j \in SV} (a_j - a_j^*) k(x_j \cdot x_i) - \varepsilon \right] + \sum_{0 < a_i^* < c} \left[ y_i - \sum_{x_j \in SV} (a_j - a_j^*) k(x_j \cdot x_i) + \varepsilon \right] \right\} \quad (6)$$

由式 (5) 和 (6) 可知, 通过非线性函数将样本数据映射到具有高维的特征空间, 只需要计算核函数, 不需要计算非线性函数, 避免了高维特征空间造成的维数灾的情况。

SVM 的解最终归结为求解一个凸二次规划问题。目前, 数学模型运用计算机求解的常用算法有: 梯度投影法、共轭梯度法、内点法、快速下降法等有效算法。

### 2.3 核函数的确定

根据泛函理论, 只要核函数  $k(x, x_i)$  满足 Mercer 条件, 就可以对变换空间中的内积。SVM 中可以采用不同的核函数构造实现输入空间不同类型的非线性决策面的学习机器。

由于核函数对算法的影响较大, 故选择一个最好的核函数极为重要。这里以径向基函数作为 SVM 方法中的核函数。径向基函数的形式为

$$k(x, x_i) = \exp \left( - \frac{\|x - x_i\|^2}{2\sigma^2} \right) \quad (7)$$

式中,  $x_i$  为支持向量的样本因子向量;  $x$  为预报因子向量;  $\sigma$  为核参数。

### 2.4 样本数据集的构建

模型输入参数由输入量 (叶片长度、叶片宽度) 和输出量 (叶片面积) 组成。为了消除各个因子由于量纲单位不同的影响, 对样本的输入和输出参数进行归一化处理, 输入输出数据被归一化至 [0, 1] 范围内, 归一化处理采用式 (8), 即

$$X_i = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}} \quad (8)$$

式中,  $x_{i\max}$ ,  $x_{i\min}$  分别为第  $i$  样本因子向量的最大值和最小值;  $X_i$  为标准化后的第  $i$  个输入向量。  $x_i$  为第  $i$  个输入因子向量。

## 3 实例分析

### 3.1 参数的确定

对模型中的参数赋予初值, 然后利用 SVM 进行学习训练, 最后根据训练后获得的参数进行预测。给定惩罚参数初值为 500, 循环步长为 1, 不敏感损失参数的上界为 500, 回归模型的择优标准采用误差绝对差方法, 迭代次数 10 000。

通过上述的参数, 对模型进行学习训练, 最终得出的最优模型参数: 惩罚参数为 500, 不敏感损失参数为

147.92, 核参数为 0.001。绝对差为 2.56, 相关系数为 0.996, 均方差为 2.73。

### 3.2 测定结果与对比分析

将 SVM 算法应用植物叶片测量, 首先选择观测资料构成样本数据集, 然后进行学习训练, 最后根据训练后获得的最优模型参数, 并对预测样本进行测量。为了对比 SVM 模型在叶片测量上的效果, 分别用线性回归和神经网络模型对相同的数据进行叶面积测定, 其结果见表 1。

表 1 实际叶面积和模型计算值比较

实测值/ cm <sup>2</sup>	SVM		线性回归		BP 算法	
	预测值/ cm <sup>2</sup>	相对 误差/%	预测值/ cm <sup>2</sup>	相对 误差/%	预测值/ cm <sup>2</sup>	相对 误差/%
82.33	86.40	-4.94	86.77	-5.39	94.01	-14.19
65.52	61.53	6.09	64.23	1.97	66.35	-1.27
125.34	127.32	-1.58	117.44	6.30	119.62	4.57
91.97	96.35	-4.76	94.77	-3.04	102.09	-11.01
65.25	66.79	-2.36	69.68	-6.79	73.20	-12.20
109.49	108.34	1.05	106.17	3.03	110.76	-1.16
135.15	135.15	0	125.97	6.79	123.42	8.68
45.84	46.32	-1.05	47.18	-2.92	47.23	-3.05

在 SVM 和神经网络叶面积模型中, 将 130 组叶片参数随机划分为: 模型训练为 100 组, 模型试验为 22 组, 模型检验为 8 组; 在线性回归模型中, 以叶片的长度和宽度作为变量, 用 122 组叶片参数进行线性拟合, 模型检验的数据与支持向量机和神经网络相同。不同模型对叶面积检验数据的计算结果见表 1。从表 1 中可以看出, 支持向量机模型计算叶面积的最大误差为 6.09%, 平均误差为 2.73%; 线性回归模型计算叶面积的最大误差为 6.79%, 平均误差为 4.53%; 神经网络模型计算叶面积的最大误差为 -14.19%, 平均误差为 7.02%。

神经网络虽然具有良好的函数逼近能力, 通过机器学习可以逼近任意线性、非线性函数, 但由于它自身的不足, 如为了达到一定的模型精度, 要用大量的样本用于训练, 在实际应用中存在一定的局限, 同时, 神经网络在应用中还容易出现过拟合现象, 模型的泛化能力差。在本试验中, 神经网络对叶面积的测定结果不及 SVM 和线性回归方法。与神经网络相比, SVM 对有限样本的情况, 可以得到最优解, 最优解为全局最优解, 避免了神经网络会出现的局部极值问题, 样本的维数与算法的难易和复杂程度无关。与线性回归方法相比, 采用线性回归方法计算的叶面积的最大误差略高于 SVM, 但其平均误差明显高于 SVM, 说明其整体测定误差较大。由此可见, SVM 信息综合能力强、结构风险小、推广能力好、预测精度高, 在本试验中, 对叶面积的测量优于线性回归方法和神经网络技术。

## 4 模型精度评价

模型精度是指模型计算的模拟值与实际值的接近程度, 取决于模型选取的合理性、所建模型的正确性、资

料的准确和全面性。模型精度是评价所选方法是否适用的一个主要指标。

模型精度评价常用误差评价函数表示, 即: 均方根误差、平均相对误差和测量精度。

1) 均方根误差 (RMSE) - 标准误差: 考核模型的预测值与实测值的离散程度。 $E_{RMSE} \geq 0$ , 当预测值和实际值完全相同时,  $E_{RMSE} = 0$ 。

$$E_{RMSE} = \sqrt{\frac{\sum_{i=1}^n (g_i - h_i)^2}{n-1}} \quad (9)$$

式中,  $n$  为样本数;  $g$  为计算值;  $h$  为实测值;  $i$  为样本序号。

2) 平均相对误差 (MAPE): 考核模型的计算值与实测值的平均相对误差大小, 当计算值和实际值完全相同时,  $E_{MAPE} = 0$ 。

$$E_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{h_i - g_i}{h_i} \right| \quad (10)$$

3) 测量精度 (MA): 考核在均值附近, 计算值和实际值偏差之间的相关性。 $-1 \leq E_{MA} \leq 1$ , 当模型预测无误差时,  $E_{MA} = 1$ 。

$$E_{MA} = \frac{\sum_{i=1}^n (g_i - \bar{g})(h_i - \bar{h})}{(n-1)\sigma_g \sigma_h} \quad (11)$$

标准差

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (12)$$

式中,  $\bar{g}$  为计算值的平均值;  $\bar{h}$  为实测值的平均值;  $\sigma_g$  为计算值标准差;  $\sigma_h$  为实测值标准差。

以上 3 种预测方法的预测精度评价见表 2。

表 2 各预测方法的预测精度评价

Table 2 Evaluation for the prediction accuracy of various forecasting methods

方法	平均相对误差	均方根误差	模拟精度
支持向量机	0.027	2.916	0.996
二元线性回归	0.045	5.456	0.992
BP 神经网络	0.070	8.251	0.972

表 2 对 3 种叶面积测定模型进行了模型精度评价, 是对模型是否适用的综合反映。支持向量机叶面积模型的平均相对误差和均方根误差分别为 0.027 和 2.196, 均小于二元线性回归方法和人工神经网络; 虽然 3 个叶面积模型对所研究问题的模拟精度都达到了 0.9 以上, 但支持向量机叶面积模型的模拟精度为 0.996, 具有较强的模拟精度, 该方法可以较好地应用于叶面积的测定。

## 5 结 论

原始样本的精确程度直接关系到模型测定叶面积的准确程度。文中详细地论述了叶面积图像处理的原理和

方法, 该方法不受叶片形状和厚薄等因素的影响, 适用于各种叶面积的测量, 但需对叶面积进行采摘后测量, 这是它仅有的不便之处。基于计算机图像处理技术具有严密的科学基础, 精度较高, 使用该方法作为模型的样本数据, 建立的叶面积测量模型精度较高。

文中建立的 3 种叶面积模型由于只需在野外测量叶片的长度和宽度, 均可实现叶面积的快速、无损测定。相比之下, 支持向量机方法由于其在支持有限样本情况下, 实现全局最优解, 且不会出现局部极值问题。该方法适用性广, 可以对任何品种、任何形状的植物叶面积建立测量模型, 是叶面积测量中较好的方法。从文中的计算和模型精度评价结果可以看出, 支持向量机叶面积模型的最大误差为 6.09%, 平均误差为 2.73%, 模拟精度达到 0.996。支持向量机叶面积模型的测试结果优于线性回归方法和人工神经网络技术。

## [参 考 文 献]

- 冯冬霞. 便携式叶面积仪的研制[D]. 北京: 中国农业大学, 2005.  
Feng Dongxia. Design and Development of Portable Leaf Area Meter[D]. China Agricultural University, 2005. (in Chinese with English abstract)
- 于浩. 便携式活体叶面积测量仪的研制[D]. 哈尔滨: 哈尔滨工业大学, 2009.  
Yu Hao. Development of Portable Meter for Live Leaf Area [D]. Harbin Institute of Technology, 2009. (in Chinese with English abstract)
- 涂育合. 雷公藤叶面积的测算方法[J]. 北华大学学报: 自然科学版, 2007, 8(6): 553-557.  
Tu Yuhe. On measurement method of leaf area for tripterygium wilfordii[J]. Journal of Beihua University: Natural Science, 2007, 8(6): 553-557. (in Chinese with English abstract)
- 王俊河, 李艳彩. 叶形影响求积仪测算叶面积精度的分析研究[J]. 山西林业科技, 1993, (3): 43-44.
- 冯冬霞, 施生锦. 叶面积测定方法的研究效果初报[J]. 中国农学通报, 2005, 21(6): 150-152.  
Feng Dongxia, Shi Shengjin. Research on night measurement methods of leaf area[J]. Chinese Agricultural Science Bulletin, 2005, 21(6): 150-152. (in Chinese with English abstract)
- 缪鹏程. 作物生长需水信息采集系统的设计与研究[D]. 合肥: 安徽农业大学, 2009.  
Liao Pengcheng. Design and Study of Water Requirement Information System for Crop Growth[D]. Anhui Agricultural University, 2009. (in Chinese with English abstract)
- 杨劲峰, 陈清, 韩晓日, 等. 数字图像处理技术在蔬菜叶面积测量中的应用[J]. 农业工程学报, 2002, 18(4): 155-157.  
Yang Jinfeng, Chen Qing, Han Xiaori, et al. Measurement of vegetable leaf area using digital image processing techniques [J]. Transactions of the CSAE, 2002, 18(4): 155-157. (in Chinese with English abstract)
- 程鸿, 吕军芬. CAD 图形处理技术在植物叶面积测量中的

- 应用[J]. 甘肃农业大学学报, 2003, 38(4): 467—470.  
Cheng Hong, Lu Jufen. Measurement of plant leaf area using CAD image processing techniques[J]. Journal of Gansu Agricultural University, 2003, 38(4): 467—470. (in Chinese with English abstract)
- [9] 杨叔子, 吴雅. 时间序列分析的工程应用[M]. 武汉: 华中理工大学出版社, 1994.
- [10] 吴春胤, 张文昭, 欧阳庆, 等. 基于 BP 神经网络模型的荔枝树叶面积测定方法[J]. 农业工程学报, 2007, 23(7): 166—169.  
Wu Chunyin, Zhang Wenzhao, Ouyang Qing, et al. BP neural network model for the measurement of the leaf area of litchi [J]. Transactions of the CSAE, 2007, 23(7): 166—169. (in Chinese with English abstract)
- [11] 陈广洲, 汪家权, 胡淑恒. 基于支持向量机的水资源开发利用评价模型[J]. 中国农村水利水电, 2008(11): 1—7.  
Chen Guangzhou, Wang Jiaquan, Hu Shuheng. Evaluation of development level of regional water resources based on support vector machines[J]. China Rural Water and Hydropower, 2008(11): 1—7. (in Chinese with English abstract)
- [12] Vapnik V. The Nature of Statistical Learning Theory[M]. Springer-Verlag, New York, 1995.
- [13] 李秀启, 李晓林, 李颜, 等. 黄瓜叶面积测量方法的评价[J]. 长江蔬菜, 2008, (6b): 71—73.
- [14] 甘露萍. 基于机器视觉技术的鲜烟叶含水量模型研究[D]. 重庆: 西南大学, 2009.  
Gan Luping. Study on Model of Fresh Tobacco Leaf's Water Contents Based on Machine Vision[D]. Southwest University, Chongqing, China, 2009. (in Chinese with English abstract)
- [15] 张冰. Photoshop CS2 中如何选你所选[J]. 科技信息, 2009: 574—575.
- [16] Joberg J, Zhang Q, Ljung L, et al. Nonlinear black-box modeling in system identification: a unified overview[J]. Automatic, 1995, 31(12): 691—724.
- [17] 张浩然, 韩正之, 李昌刚. 基于支持向量机的非线性模型预测控制[J]. 系统工程与电子技术, 2003, 25(3): 330—334.  
Zhang Haoran, Han Zhengzhi, Li Changgang. Support vector machine based nonlinear model predictive control[J]. Systems Engineering and Electronics, 2003, 25(3): 330—334. (in Chinese with English abstract)
- [18] 高昇. 基于支持向量机的非线性系统建模与预测控制[D]. 西安: 西安理工大学, 2006.  
Gao Yi. Nonlinear System Identification and Control Based on Support Vector Machine[D]. Xi'an: Xi'an University of Technology, 2006. (in Chinese with English abstract)

## Determination of leaf area of sweet pepper based on support vector machine model and image processing

Zai Songmei, Wen Ji<sup>\*</sup>, Guo Dongdong, Han Qibiao, Deng Zhong, Sun Hao, Zhao Dongbin  
(Farmland Irrigation Research Institute, Chinese Academy of Agricultural Sciences, Xinxiang 453003, China)

**Abstract:** As a vital organ for crop photosynthesis, leaf is one of the major biological indicators in the study on light absorption by crops. Support Vector Machine (SVM) theory was used to set up a SVM model for determination of leaf area of sweet pepper, the input parameters were the leaf length, maximum width of the leaf, and the output parameters were the leaf areas. Data measured by computer image processing technology were trained as samples, the length, maximum width of the leaf were used as input parameters to simulate and test the leaf area. The results were compared with those of linear regression and artificial neural network model. The results showed that the maximum error of leaf area determined by support vector machine model was 6.09%, and the average error was 2.73%, the simulation accuracy was 0.996. This method can well reflect the actual size of leaf area of sweet pepper, and has good practical value and application prospect.

**Key words:** support vector machine, models, image processing, sweet pepper, leaf area, determination