

基于小波变换与支持向量机回归的冬小麦叶面积指数估算

梁 栋^{1,2}, 杨勤英^{1,2,3}, 黄文江^{2,3}, 彭代亮³, 赵晋陵^{1,2}, 黄林生^{1,2}, 张东彦^{1,2}, 宋晓宇⁴

- (1. 安徽大学 计算机智能与信号处理教育部重点实验室, 安徽 合肥 230039;
2. 安徽大学 电子信息工程学院, 安徽 合肥 230039;
3. 中国科学院遥感与数字地球研究所 数字地球重点实验室, 北京 100094;
4. 北京农业信息技术研究中心, 北京 100097)

摘要: 叶面积指数(LAI)是作物长势诊断及产量预测的重要参数。通过对冬小麦采样点的高光谱曲线进行连续小波变换(CWT), 然后利用小波系数与 LAI 建立支持向量机回归(SVR)模型, 实现冬小麦不同生育时期的叶面积指数估算。通过对所研究方法与选取的植被指数、偏最小二乘(PLS)回归等5种方法的反演结果进行统计分析。结果表明: 利用连续小波变换确定的 LAI 的敏感波段为 680、739、802、895 nm, 对应尺度分别为 8、4、9 和 8, 对应小波系数的 LAI 回归确定系数(R^2)明显高于冠层反射率的回归确定系数; 利用小波系数与 LAI 建立的 SVR 模型的反演精度最高, 模型实测值与预测值的检验精度(R^2)为 0.86, 均方根误差(RMSE)为 0.43; 而常用植被指数(归一化植被指数, NDVI; 比值植被指数, RVI)建立的估测模型对冬小麦多个生育时期 LAI 反演精度最低($R^2 < 0.76$, $RMSE > 0.56$)。因此利用连续小波变换进行数据预处理, 能更好地筛选出对叶面积指数敏感的信息, LAI 回归方法比较结果表明, SVR 比 PLS 更适合于 LAI 的估测, 通过将 CWT 与 SVR 结合(CWT-SVR)能实现不同生育时期冬小麦叶面积指数的遥感估算。

关键词: 叶面积指数(LAI); 高光谱; 连续小波变换(CWT); 支持向量机回归(SVR); 偏最小二乘(PLS)

中图分类号: TP391 文献标志码: A 文章编号: 1007-2276(2015)01-0335-06

Estimation of leaf area index based on wavelet transform and support vector machine regression in winter wheat

Liang Dong^{1,2}, Yang Qinying^{1,2,3}, Huang Wenjiang^{2,3}, Peng Dailiang³, Zhao Jinling^{1,2}, Huang Linsheng^{1,2}, Zhang Dongyan^{1,2}, Song Xiaoyu⁴

- (1. Key Laboratory of Intelligent Computer & Signal Processing, Ministry of Education, Anhui University, Hefei 230039, China;
2. School of Electronic and Information Engineering, Anhui University, Hefei 230039, China;
3. Key Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; 4. Beijing Agriculture Information Technology Research Center, Beijing 100097, China)

Abstract: Leaf area index (LAI) is an important parameter of crop diagnosis and yield prediction. The LAI of winter wheat obtained from Beijing city had been estimated successfully by support vector

收稿日期: 2014-05-18; 修订日期: 2014-06-24

基金项目: 国家自然科学基金(61172127, 41301471, 41271412); 国家 863 计划(2012AA12A307)

作者简介: 梁栋(1963-), 男, 教授, 博士, 主要从事信号与信息处理方面的研究。Email: dliang@ahu.edu.cn

通讯作者: 黄文江(1972-), 男, 研究员, 博士生导师, 博士, 主要从事植被定量遥感方面的研究。Email: huangwenjiang@ceode.ac.cn

machine regression (SVR) model built with LAI and wavelet coefficients of hyperspectral reflectance. The inversion results of this paper method and other five methods, such as selected vegetation indices and partial least-square (PLS) regression models, were analyzed. It was found that the sensitive bands to assess LAI were 680 nm, 739 nm, 802 nm, and 895 nm, and the corresponding wavelet decomposition scales were 8, 4, 9, and 8 determined by continuous wavelet transform(CWT), respectively. The decision coefficient (R^2) of regression equation between LAI and wavelet coefficient was significantly higher than that of between LAI and canopy reflectance. The SVR model based on wavelet coefficients performed best with R^2 of 0.86, and RMSE of 0.43, while the regression models based on two common spectral vegetation indices (NDVI and RVI) performed poor in estimating LAI of winter wheat's multiple birth period ($R^2 < 0.76$, $RMSE > 0.56$). It can conclude that the pretreatment method of CWT is better effective for selecting sensitive spectral characteristics to LAI. Meanwhile, SVR is more suitable for developing model in LAI estimation than PLS regression. The combination of CWT and SVR is feasible to realize remote sensing inversion of LAI in the whole growth period of winter wheat.

Key words: leaf area index(LAI); hyperspectral; continuous wavelet transform (CWT); support vector machine regression(SVR); partial least-square(PLS)

0 引言

叶面积指数(LAI)是指单位地表面上植物叶片单面面积的总和,表征植被与大气能量交换的面积大小,控制许多生物和物理过程,与植被冠层的截获、光合和呼吸作用等密切相关,同时也是许多生态系统生产率模型的重要驱动因子^[1]。此外,LAI还是反映作物的长势与产量信息的最重要的参数之一,因此实时动态监测作物 LAI 对作物长势监测和产量预测有着重要的意义。

传统的实验室测量 LAI 的方法费时费力,并且仅能获得地面有限点的 LAI 值,难以实现大面积作物长势的监测。大量实验表明,遥感技术为大面积实时估算 LAI 提供了一个强有力的工具。目前,运用遥感光谱数据进行叶面积指数反演的方法可以归为两类:经验模型法和物理模型法^[2-3]。虽然这些方法在一定程度上都能实现 LAI 的估测,但是仍有一定的限制,如光谱数据受到的干扰较多、经验模型法中的植被指数运用的波段较少及物理模型法的输入参数较多且不容易选择等。近年来,小波变换和支持向量机(SVM)作为新兴学科已经被广泛应用到农业遥感中,并取得了比其他常用方法较好的结果^[4-5]。连续小波变换能够将光谱分解成不同尺度或频率的组分进而检测微妙信号,以表征信号的局部频谱特征^[6]。

由于植被的各种理化成分对光谱的反射和吸收特征具有明显的局部性质,小波变换的局部分析特性在植被的生物生理参数反演方面能得到充分的应用。SVM 是基于最小化结构风险和置信范围的多变量回归,在运用多光谱数据反演 LAI 的过程中既可以输入多波段的信息,又可以避免过学习问题。由于高光谱数据具有波段多、数据量大、冗余度大等特点,在利用 SVM 进行多变量回归前,必须进行有效的筛选,连续小波变换则可以实现这个目的,如果将两者有机结合,则可以高效地反演 LAI。然而,目前的 LAI 遥感反演仅是单独利用小波变换或者 SVM 方法,尚未出现将两种方法结合起来预测 LAI 的研究。文中先将小波变换用于冬小麦的高光谱数据,筛选出对 LAI 最敏感的信号,然后用支持向量机回归(SVR)算法建立冬小麦 LAI 回归模型。将优选后的敏感波段信息融合到回归模型中,能够提高冬小麦多个生育时期的 LAI 反演精度。

1 材料与方法

1.1 试验概况

试验区位于北京市昌平区小汤山精准农业示范基地,地处北纬 $40^{\circ}11'$,东经 $116^{\circ}27'$,气候类型为典型的暖温带半湿润大陆季风气候,土壤类型为潮土。试验开展于 2003 年 9 月至 2004 年 6 月,在 2004

年冬小麦起身期(4月1日)、拔节期(4月15日)、孕穗期(4月28日)、开花期(5月10日)、灌浆期(5月19日、5月28日)分别采集实验点数据,共125个样本点数据,其中70%的样本点数据用于建立模型,30%的样本点数据用于模型的验证。试验点数据的采集内容包括冠层光谱和LAI。

1.2 冠层光谱和LAI测定

使用美国ASD公司生产FieldSpec Pro FR光谱仪,选择在少云无风的天气,于北京时间10:00~14:00采集冬小麦起身期、拔节期、孕穗期、开花期、灌浆期的冠层光谱。测量时,探头垂直向下并且与冠层相距1.3m的距离,在测量前后均用40cm×40cm BaSO₄校正板校正,每个采样点测定20次,取其平均值为最后测量值。冠层光谱测定后,于相应位置进行取样操作,取样面积为0.3m²,将取样样本放入黑色塑料袋中密封带回实验室,采用比叶重法测量叶面积指数。

1.3 小波变换

小波变换将不同频率组成的混合信号用分辨率不同的窗口分解成不同频率的块信号,可以表征信号的局部频谱特征。小波变换又可分为连续小波变换(CWT)和离散小波变换(DWT)。由于DWT在输出参数的解析上比较困难,文中采用CWT变换。

小波函数基是通过同一母函数经过伸缩和平移得到一组函数序列,如公式(1)^[7]:

$$\Psi_{a,b} = \frac{1}{\sqrt{a}} \Psi\left(\frac{\lambda-b}{a}\right) \quad (1)$$

$$\int_{-\infty}^{\infty} \Psi(\lambda) d\lambda = 0 \quad (2)$$

式中: $\Psi(\lambda)$ 为小波母函数,满足公式(2); λ 为光谱的波段数; a 为伸缩因子; b 为平移因子。通过连续小波变换分解后,反射光谱 $f(\lambda)$ 的连续小波变换定义为:

$$W_f(a,b) = \langle f(\lambda), \Psi_{a,b} \rangle = \int_{-\infty}^{\infty} f(\lambda) \Psi_{a,b}(\lambda) d\lambda \quad (3)$$

连续小波系数 $W_f(a_j, b_j)$ 构成一个二维($m \times n$)的量图,分别为分解尺度和波段($i=1, 2, \dots, m, j=1, 2, \dots, n$)。每个元素的值代表了一个小波系数,反映了对应位置下的光谱形状与特定小波基的相似程度^[6]。

1.4 支持向量机回归(SVR)

支持向量机(SVM)是Vapnik等提出的一种机器学习方法,被应用于各种分类和回归问题^[8-10]中。支

持向量机回归(SVR)首先用线性回归函数 $f(x)=w^*x+b$ 拟合数据 $\{x_i, y_i\}, i=1, \dots, n, x_i \in R^d, y_i \in R$ 的问题,依据SVM原理,引入 ε 不敏感系数及允许拟合误差的情况下的松弛因子 ξ 和 ξ^* ,且 ξ 和 ξ^* 均大于等于0,则问题可描述为^[11]:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (4)$$

$$\text{s.t.} \begin{cases} y_i - wx_i - b \leq \varepsilon + \xi_i \\ wx_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (5)$$

式中:第一项是为了提高泛化能力;第二项是为了减少误差;大于0的常数 C 被称为惩罚因子或正则化系数,控制对超出误差范围的样本的惩罚程度。引入Lagrange算法,对Lagrange因子 α_i, α_i^* 最大化目标函数,回归函数 $f(x)$ 则可表示为:

$$f(x) = \sum_{\text{支持向量}} (\alpha_i - \alpha_i^*) (x_i \cdot x) + b \quad (6)$$

对于非线性回归,支持向量机主要是将低维空间的非线性问题变为高维空间的线性问题,并在高维空间中进行线性回归。只需要用非线性核函数代替上式中的内积运算即可,则非线性拟合的表达式为:

$$f(x) = \sum_{\text{支持向量}} (\alpha_i - \alpha_i^*) K(x_i \cdot x_j) + b \quad (7)$$

常用的核函数有多项式核函数(POLY)、径像基核函数(RBF)和ANOVA核函数(ANOVA)。文中采用径像基核函数(RBF)。

1.5 光谱植被指数的选择

经验模型法中多种植被指数被用于叶面积指数反演,文中选取最常用和最基础的两种植被指数:归一化植被指数(NDVI)和比值植被指数(RVI)。NDVI和RVI的计算公式如下:

$$NDVI = (R_{\text{mir}} - R_{\text{red}}) / (R_{\text{mir}} + R_{\text{red}}) \quad (8)$$

$$RVI = R_{\text{mir}} / R_{\text{red}} \quad (9)$$

式中: R_{mir} 、 R_{red} 分别表示近红外(800nm)、红光(680nm)波段的反射率。

2 结果与分析

2.1 冬小麦光谱反射率与LAI的相关分析

从图1(a)中所有采集的各波段光谱反射率与对应冬小麦LAI的相关分析($n=125$)结果可以看出,冬小麦LAI与350~700nm之间的冠层反射率存在

明显的负相关关系,并在红光吸收谷处负相关系数(r)达到 0.68;760~920 nm 之间的冠层反射率与冬小麦 LAI 存在较好的正相关关系,相关系数保持在 0.7~0.8 之间;700~760 nm 之间的冠层反射率与冬小麦 LAI 的相关系数变化较大,由显著的负相关变为显著的正相关。图 1(b) 显示的确定系数(R^2)与相关系数的变化趋势基本一致,冬小麦 LAI 与 760~920 nm 光谱反射率的相关性最好, R^2 在 0.5~0.6 之间,其次是 585~690 nm 和 480~520 nm 的反射率光谱, R^2 在 0.44~0.5 之间,并在 502 nm、649 nm、778 nm 及 860 nm 4 个波段处 R^2 取最大值。

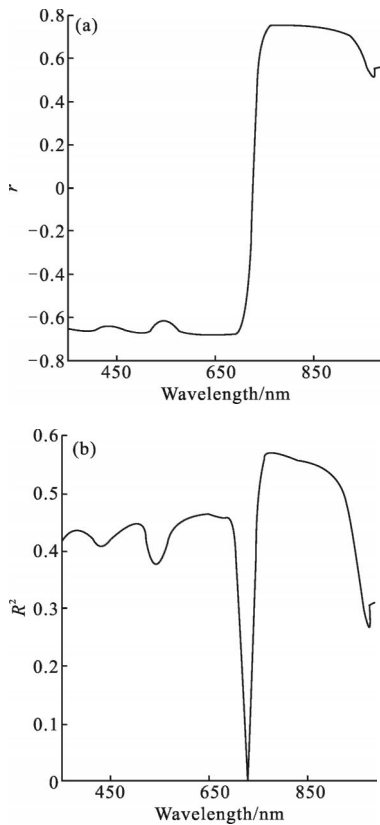


图 1 冬小麦叶面积指数与光谱反射率的相关分析及线性回归分析

Fig.1 Correlation and linear regression between LAI and reflectance on winter wheat

2.2 冬小麦冠层光谱经过 CWT 变换后的小波系数与 LAI 的相关分析

该研究通过前人的研究经验,选取 BiorSplines 小波系的“biro1.3”^[7],通过 Matlab 编程对 89 个样本点的高光谱曲线进行 CWT 变换(为了减少数据量同时保持小波伸缩及尺度的有效性,连续小波分解的

尺度定为 21,22,23,24,25,26,27,28,29,210),并将小波变换系数与对应的冬小麦 LAI 做相关分析,生成相关系数图,如图 2 所示。从图中可以看出,与冬小麦 LAI 相关性较好的波段集中在 440~490 nm、529~585 nm、648~834 nm、870~952 nm,对应尺度分别为 1~6、1~8、1~10、7~8, R^2 的最小值为 0.55,并在第 8 尺度的 680 nm、第 4 尺度的 739 nm、及第 9 尺度的 802 nm、及第 8 尺度的 895 nm 达到最大值($R^2 > 0.66$),明显高于冠层光谱反射率与 LAI 的确定系数。该结果表明 LAI 敏感波段主要集中在红光和近红外处,利用连续小波变换进行数据预处理,能更好地筛选出对叶面积指数敏感的信息。

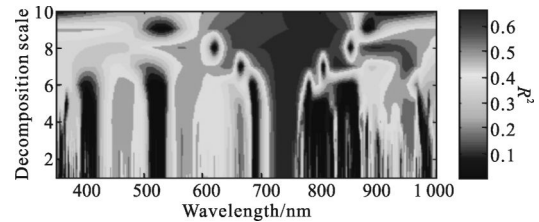


图 2 小波系数与叶面积指数的相关系数

Fig.2 Correlation scalogram between wavelet coefficient and leaf area index

2.3 建模方法的优化选择

为择取较好的建模方法,将采用小波系数与支持向量机算法建立的回归模型(CWT-SVR)、采用小波系数与偏最小二乘算法建立的回归模型(CWT-PLS)、冠层光谱经过支持向量机算法建立的回归模型(R-SVR)、冠层光谱经过偏最小二乘算法建立的回归模型(R-PLS)及植被指数回归模型(NDVI-LAI, RVI-LAI)进行比较分析以实现模型选优。

通过文中分析,冠层光谱反射率取 502 nm、649 nm、778 nm 及 860 nm 4 个波段的反射率作为输入自变量;小波系数取第 8 尺度的 680 nm、第 4 尺度的 739 nm、第 9 尺度的 802 nm 及第 8 尺度的 895 nm 作为输入自变量,以校正集 89 个样本在 Matlab 中建立 PLS 模型和 SVR 模型(使用 LIBSVM)。通过尝试发现 NDVI 和 RVI 与 LAI 的拟合方程以指数函数为最佳。

为比较模型的精度,将 6 种模型对校正集与预测集的拟合结果进行了 R^2 与均方根误差(RMSE)的分析检验,其结果见表 1。由表可知,植被指数回归模型由于应用的波段信息较少,在校正集和预测集上反演精度都最低。在预测集上,以冠层反射率为自变

量,SVR 模型得到的 RMSE 比 PLS 模型得到的 RMSE 降低了 11.7%;以小波系数为自变量,SVR 模型的 RMSE 比 PLS 模型的 RMSE 降低了 9.6%,说明基于最小化结构风险和置信范围的 SVR 算法比 PLS 算法更适合于 LAI 的估测。同时由表 1 还可以看出,以小波系数为自变量的 SVR 模型对校正集和预测集均有最大的 R^2 (0.8145、0.875)和最小的 RMSE (0.4662,0.3812) 对 LAI 回归具有最好的拟合效果,在校正集和预测集上其 RMSE 比以冠层反射率为自变量的 SVR 模型分别降低了 18.6%和 11.7%,表明利用连续小波变换进行数据预处理能更好地筛选出对叶面积指数敏感的信息。此外,文中校正集和预测集的数据均选用了 5 个生育时期的数据,说明以小波系数为自变量的 SVR 模型能较好地实现多个生育期 LAI 的反演。

表 1 LAI 预测模型在校正集与预测集的决定系数 (R^2)与均方根误差(RMSE)

Tab.1 Determination coefficient (R^2) and root mean square error (RMSE) of LAI prediction models in calibration and validation set

Model	Calibration set(n=89)		Validation set(n=36)	
	R^2	RMSE	R^2	RMSE
NDVI-LAI	0.696 3	0.714 0	0.700 4	0.669 1
RVI-LAI	0.678 9	0.640 0	0.758	0.568 6
R-SVR	0.770 4	0.502 0	0.822 2	0.483 3
R-PLS	0.719 9	0.554 0	0.770 2	0.547 3
CWT-SVR	0.848 2	0.408 5	0.860 8	0.426 9
CWT-PLS	0.744 4	0.529 2	0.834 3	0.472 4

为进一步直观地观察模型对未知样本的适应性,将 6 种模型预测的结果分别与实测值进行回归拟合,如图 3 所示。由图可以更加清晰地看出,植被指数回归模型反演精度最低,特别是 NDVI 对较大

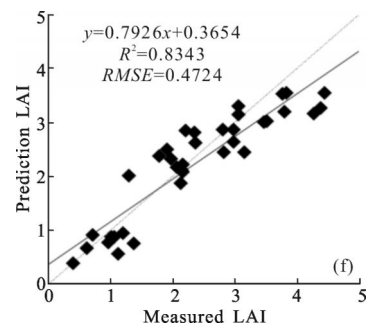
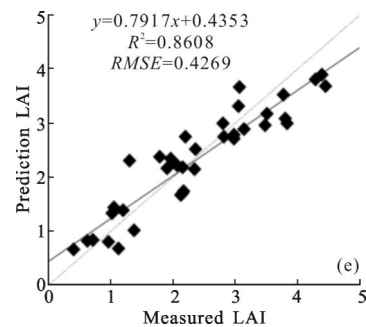
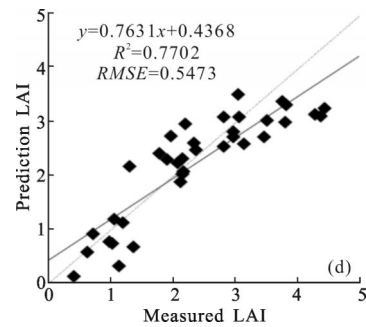
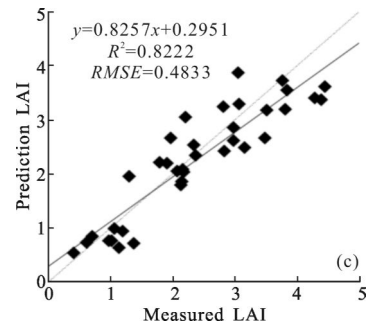
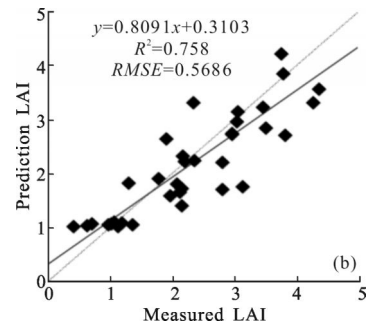
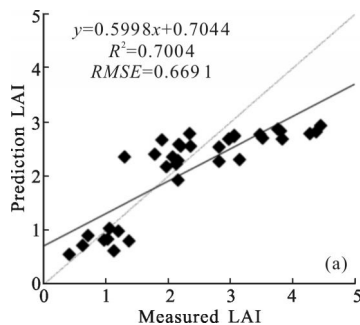


图 3 6 种模型预测 LAI 与实测 LAI 的散点图

Fig.3 Scatter plots between prediction LAI value from six models and measured LAI value

的 LAI 不能很好地预测, 这与 NDVI 在 LAI 较大时容易饱和是一致的; 以小波系数为自变量的 SVR 模型的实测值与预测值沿直线 $y=x$ 紧凑且均匀的分布, 能更加有效地实现 LAI 的预测。

3 结 论

文中介绍了一种基于连续小波变换与支持向量机回归的 LAI 反演方法。以北京冬小麦为例, 将冬小麦的高光谱曲线进行连续小波变换, 然后利用小波系数与 LAI 建立 SVR 模型, 估算该地区冬小麦不同生育时期的 LAI。结果表明, 文中方法的拟合效果比其他 5 种常用方法好 ($R^2=0.8608$, $RMSE=0.4269$)。进一步说明多波段同单波段相比能更好地反映作物的信息, 提高 LAI 的反演精度; 冠层光谱反射率经 CWT 变换后得到的小波系数与 LAI 的相关性更好, 连续小波变换进行数据预处理, 能更好地筛选出对叶面积指数敏感的信息; 径向基核函数(RBF)的 SVM 模型与 PLS 模型相比, 对叶面积指数具有较高的估算精度。小波变换作为一种光谱处理工具, 能够很好地进行微妙信号的检测, 在作物定量遥感分析中具有重要的意义。SVM 算法具有更加严格的数学理论基础, 能够更加巧妙地解决多元非线性回归的问题, 而且具有良好的鲁棒性。文中将这两种方法结合起来, 为更加准确地估算 LAI 提供了基础。

参考文献:

- [1] Viana A, Gitelson A A, Nguy-Robertson A L, et al. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops [J]. *Remote Sensing of Environment*, 2011, 115: 3468–3478.
- [2] Knyazikhin Y, Martonchik J V, Myneni R B, et al. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data [J]. *Journal of Geophysical Research*, 1998, 103(24): 32257–32275.
- [3] Thenkabail P S, Smith R B, Pauw E D. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics [J]. *Remote Sensing of Environment*, 2000, 71: 158–182.
- [4] Blackburn G A, Ferwerda J G. Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis [J]. *Remote Sensing of Environment*, 2008, 112: 1614–1632.
- [5] Liang Dong, Guan Qingsong, Huang Wenjiang, et al. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(7): 117–123. (in Chinese)
梁栋, 管青松, 黄文江, 等. 基于支持向量机回归的冬小麦叶面积指数遥感反演 [J]. *农业工程学报*, 2013, 29(7): 117–123.
- [6] Cheng T, Rivard B, Sánchez-Azofeifa G A, et al. Continuous wavelet analysis for the detection of green attack damage due to mountain pine beetle infestation [J]. *Remote sensing of Environment*, 2010, 114: 899–910.
- [7] Bruce L M, Morgan C, Larsen S. Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2001, 39: 2217–2229.
- [8] Vapnik V N. *The Nature of Statistical Learning Theory*[M]. New York: Springer-Verlag, 1995.
- [9] Li Ming, Yang Jie, Wang Hui, et al. Novel color image restoration method: sliding window based support vector regression algorithm [J]. *Infrared and Laser Engineering*, 2006, 35(1): 79–82. (in Chinese)
黎明, 杨杰, 王辉, 等. 一种彩色图像复原新方法: 基于滑动窗口的支持向量回归算法 [J]. *红外与激光工程*, 2006, 35(1): 79–82.
- [10] Yao Yuan, Hu Gensheng, Liang Dong. Remote sensing image fusion based on wavelet support vector machine [J]. *Computer Engineering*, 2011, 37(3): 218–221. (in Chinese)
姚媛, 胡根生, 梁栋. 基于小波支持向量机的遥感影像融合 [J]. *计算机工程*, 2011, 37(3): 218–221.
- [11] Chu Xiaoli, Yuan Hongfu, Luo Xianhui, et al. Developing near infrared spectroscopy calibration model of molar ratio between methanol and isobutylene by support vector regression [J]. *Spectroscopy and Spectral Analysis*, 2008, 28(6): 1227–1231. (in Chinese)
褚小立, 袁洪福, 骆献辉, 等. 支持向量回归建立测定醇烯比的近红外光谱校正模型 [J]. *光谱学与光谱分析*, 2008, 28(6): 1227–1231.