

最小二乘支持向量机用于时间序列叶面积指数预测

梁 栋^{1,2}, 谢巧云^{1,2,3}, 黄文江^{2,3}, 彭代亮³, 杨晓华⁴, 黄林生^{1,2}, 胡 勇³

(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 遥感科学国家重点实验室, 北京 100094; 4. 气象水文空间天气总站, 北京 100094)

摘 要: 遥感反演的叶面积指数(LAI)时间序列被广泛应用于气候模拟、作物长势监测等研究。但遥感数据受天气等因素影响, 时间序列的 LAI 数据存在缺失。支持向量机(SVM)是一种有效的数据分类和回归预测工具, 而最小二乘支持向量机(LS-SVM)是对 SVM 的有效改进。以西藏那曲县为例, 使用 2003–2011 年 MODIS LAI 产品, 分别用 LS-SVM 和 SVM 两种方法对研究区域 2011 年 LAI 时间序列进行预测, 并用 MODIS 原始 LAI 以及部分地面实验样点值进行验证。结果表明, 基于 LS-SVM 的 LAI 时间序列预测算法的精度比基于 SVM 的算法高, 从而证明 LS-SVM 方法能够弥补遥感反演时间序列 LAI 数据的缺失问题, 对提高时间序列的 LAI 遥感产品质量具有重要意义。

关键词: 最小二乘支持向量机(LS-SVM); 支持向量机(SVM); 叶面积指数; 时间序列; MODIS
中图分类号: TP79 **文献标志码:** A **文章编号:** 1007-2276(2014)01-0243-06

Using least squares support vector machines to estimate time series leaf area index

Liang Dong^{1,2}, Xie Qiaoyun^{1,2,3}, Huang Wenjiang^{2,3}, Peng Dailiang³, Yang Xiaohua⁴,
Huang Linsheng^{1,2}, Hu Yong³

(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. The State Key Laboratory of Remote Sensing Science, Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; 4. Space Weather Center Meteorological and Hydrographic Department, Beijing 100094, China)

Abstract: The multi-temporal leaf area index (LAI) data retrieved from remote sensing images have been widely used in climate simulation, crop growth monitoring and etc. However, there might be some missing data owing to temporal resolution, weather and some other factors. The support vector machine (SVM) is a kind of machine learning algorithm that has excellent properties. The least squares support vector machine (LS-SVM) algorithm is an improved algorithm of SVM. In this paper, the LS-SVM and SVM models were used to predict the LAI time series products of MODIS data of Naqu in year 2011, based on

收稿日期: 2013-05-10; 修订日期: 2013-06-25

基金项目: 国家自然科学基金(61172127, 41201354); 国家 863 项目(2012AA12A307); 高等学校博士学科点科研基金(20113401110006)

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

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

the MODIS LAI from 2003 to 2011. The results show that LS-SVM method performs better than SVM method. Therefore the predicted LAI data is proved to be very supportive for making up for the loss of remote sensing LAI time-series data, the LS-SVM method proposed in this study is significant to improve the quality of the LAI time series remote sensing products.

Key words: least squares support vector machine (LS-SVM); support vector machine (SVM); leaf area index(LAI); time series; MODIS

0 引言

叶面积指数(LAI)是陆面过程中的一个十分重要的植被生物物理参数,是表征植被冠层结构最基本的参量之一,它控制着植被的许多生物、物理过程,如光合、呼吸、蒸腾、碳循环和降水截获等,其定义为单位地面面积上绿色叶子的单面面积之和^[1]。传统的 LAI 地面测量方法获得信息有限,遥感估算能提供地面测量所无法得到的大区域、高时相 LAI 产品,因此卫星遥感是大区域研究 LAI 的重要途径^[2]。目前已发布的全球 LAI 产品包括 VEGETATION, AVHRR, MISR, POLDER 和应用非常广泛的 MODIS LAI 产品。但是,这些 LAI 产品在时间和空间上存在一定的不连续性,此外不同来源的数据产品之间存在不一致,这些极大地限制了遥感反演的 LAI 在地表过程、气候模拟、以及全球变化研究中的重要应用。因此在时间序列上精确预测 LAI 具有非常重要的意义。

近年来,很多学者致力于 LAI 反演算法的研究和分析。存在的算法主要分为两类:基于统计的方法^[3-4]和基于机理模型的方法^[5-6]。基于统计的方法,是根据研究对象多年累积的数据及其对应时刻,结合气温、降水等气象因素,进行相关性分析,实现预测和估算。但是这种方法在适用性和灵活性上有所欠缺。基于模型的方法,主要根据时间序列数据本身的特点,建立数学模型,实现拟合和预测等功能。这种方法具有很强的适用性,但是由于计算复杂,目前没有得到广泛应用。

支持向量机(support vector machines, SVM)是一种基于结构风险最小化原理的机器学习技术,具有较好的泛化性能和精度,减小了对经验的依赖。最小二乘支持向量机 (least squares support vector machines, LS-SVM) 是对 SVM 的有效改进。由于其

优越的学习能力,在国内外学术界受到广泛重视,成为当前混合反演方法中的一种重要手段,并已在电力、金融等领域得到了越来越多的应用^[7-8]。但是目前尚未出现 SVM 以及 LS-SVM 在 LAI 预测方面的研究。文中以西藏那曲县的 LAI 时间序列为研究对象,比较 LS-SVM 和 SVM 模型对遥感反演时间序列 LAI 的预测精度,最终达到填补时间序列上缺失数据,提高 MODIS LAI 产品质量的效果。

1 研究方法

1.1 支持向量机和最小二乘支持向量机原理

支持向量机是 VAPNIK 等在统计学理论上建立起来的一种新的机器学习方法,其基本思想是通过非线性映射 $\Phi(\cdot)$ 将 n 维输入、1 维输出向量:

$$(x_k, y_k), x_k \in R^n, y_k \in R, k=1, \dots, N$$

从原空间映射到高维特征空间 F , 在 F 中构造最优线性回归函数:

$$f(x) = \omega \cdot \Phi(x) + b \quad (1)$$

同时利用了结构风险最小化原则,引入了间隔的概念,并巧妙地利用原空间的核函数取代高维特征空间的点积运算,简化了复杂计算。

根据结构风险最小化原则,寻求公式(1)中的 ω 和 b 就是优化。最小二乘支持向量机在传统支持向量机的基础上采用不同的损失函数,同时将支持向量机算法中的不等式约束条件转化为等式约束条件,在此过程中求解的是线性方程组,从而使问题得到简化,大幅提高了运行效率。由此得到预测输出为:

$$y(x) = \sum_{i=1}^l \alpha_i \Phi^T(x) \Phi(x_i) + b \quad (2)$$

令 $K(x, x_i) = \Phi^T(x) \Phi(x_i)$ 。 $K(x, x_i)$ 称为核函数,它必须满足 Mercer 定理^[9]。常用的核函数主要有多项式核函数、RBF 核函数和 Sigmoid 核函数等。

1.2 LS-SVM 模型的建立

研究使用 MATLAB 2010a 软件,结合 LS-SVM 工具箱进行建模。模型的建立包括核函数与参数的选取。在线性不可分的情况下,LS-SVM 中核函数的选定非常关键,它的选择直接影响到算法的实现与效果。选取 RBF 核函数,因为它具有优秀的局部逼近性,应用非常广泛。核函数遵循 Mercer 条件,这保证了 LS-SVM 训练的二次规划问题是凸优化问题,即得到的最优解为全局最优解^[10-11]。在选定核函数之后,LS-SVM 需要完成参数的选择,包括超参数 γ 和核参数 σ 。 γ 对应于标准向量机中的惩罚因子 C ,它影响到训练误差的大小和泛化能力的强弱; σ 反映了训练样本数据的分布或范围特性,它决定了局部领域的宽度, σ 越大意味着方差越小。交叉验证法选择参数可以有效避免过拟合现象,因此文中采用这种方法来寻找最优参数。利用 LS-SVM 预测 LAI 的步骤如图 1 所示。

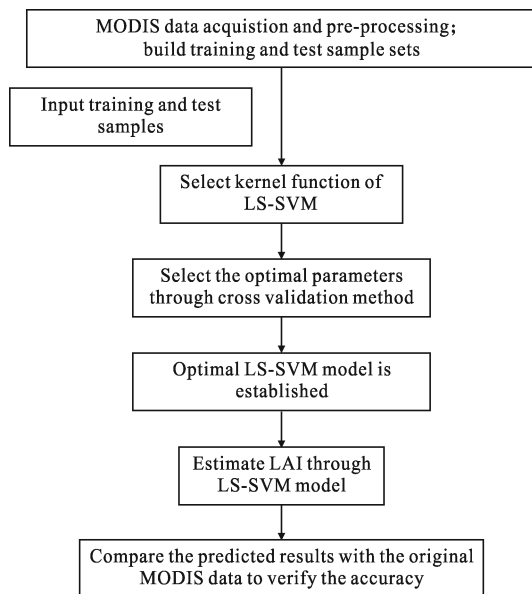


图 1 研究方法流程图

Fig.1 Flow chart of this research

1.3 研究区域与数据准备

文中以中国西藏那曲县为示范区,以 2003-2011 年 MODIS 的 LAI 产品(时空分辨率)为研究数据。对研究区域每个像元点,用 2003-2010 年原始 MODIS LAI 数据,构成该像元的 LAI 时间序列。取 2003-2009 年的序列作为训练样本,2010 年的序列作为训练目标;取 2004-2010 年的序列作为测试样本,2011 年的序列作

为测试目标。首先根据训练数据集建立最优 LS-SVM 模型,再用最优的 LS-SVM 模型对测试样本进行预测,得到每个像元 2011 年的 LAI 时间序列。

2 结果分析与讨论

文中采用相对百分比误差 (relative percentage error, RPE)、平均相对百分比误差 (mean relative percentage error, MRPE)、绝对百分比误差 (absolute percentage error, APE) 和平均绝对百分比误差 (mean absolute percentage error, MAPE) 来对预测值进行评价。计算公式如下:

$$RPE = \frac{P-R}{R} \times 100\% \quad (3)$$

$$MRPE = \frac{1}{N} \sum RPE \quad (4)$$

$$APE = \frac{|P-R|}{R} \times 100\% \quad (5)$$

$$MAPE = \frac{1}{N} \sum APE \quad (6)$$

式中: R 是实际 MODIS 的 LAI 值; P 是预测 LAI 值; N 是样本数,研究中 N 为 46,即每个点的预测输出为 2011 年 46 景的 LAI 值。

文中用支持向量机方法做了对比实验,得出预测结果。为了确定 LS-SVM 及 SVM 对时间序列 LAI 的预测能力,采用 2011 年 MODIS LAI 原始数据对整个研究区域预测值进行初步验证,再用部分地面实测数据做进一步验证。

2.1 基于 MODIS 原始 LAI 产品的误差分析

(1) 计算研究区域所有像元预测 LAI、MODIS 原始 LAI、多年 LAI 的均值,四种均值随时间变化的曲线如图 2 所示。

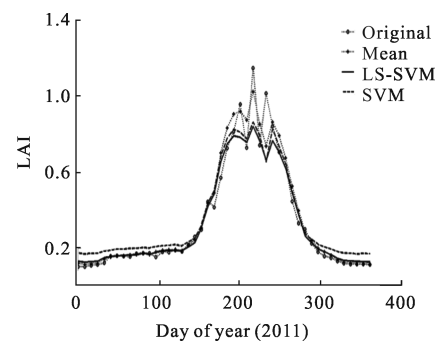


图 2 LS-SVM 和 SVM 方法预测结果与原始 LAI 以及多年均值对比
Fig.2 Comparison between prediction results of LS-SVM and SVM method, the original LAI and the mean LAI in 2003-2011

从图 2 中可以看出,在 2011 年年末和年初,LS-SVM 预测与原始值非常接近。中期是植被生长旺盛的时期,原始 LAI 随时间的推移波动较大,而 LS-SVM 预测的 LAI 时间序列表现的较为平缓,相邻两景数据之间相差不会特别大,而原始 LAI 第 201、217、233 天的 3 景数据,分别较前一期 LAI 增长了 17.58%、57.47%、36.98%,因此这 3 景数据 LS-SVM 预测值偏低。但是 LS-SVM 预测曲线更加连续和稳定,与多年均值更接近,说明预测值更符合植被实际生长规律。分别计算两种方法预测结果相对于原始

LAI 的误差,LS-SVM 方法的 MAPE 为 10.01%,SVM 方法的 MAPE 为 24.46%。图 2 的分析和误差结果均证明 LS-SVM 对研究区域整体时间序列 LAI 的预测效果较好。

(2) 计算每个像元的 RPE,形成 RPE 分布直方图(图 3)。将图 3 中的 RPE 的绝对值(即 APE)进行分段统计,如图 4 所示。

从图 3 中可以看出,两种方法预测的大部分值都比原始值偏高。对照图 2,可以分析出预测偏高的值分布在年初和年末,预测偏低的值分布在年中植

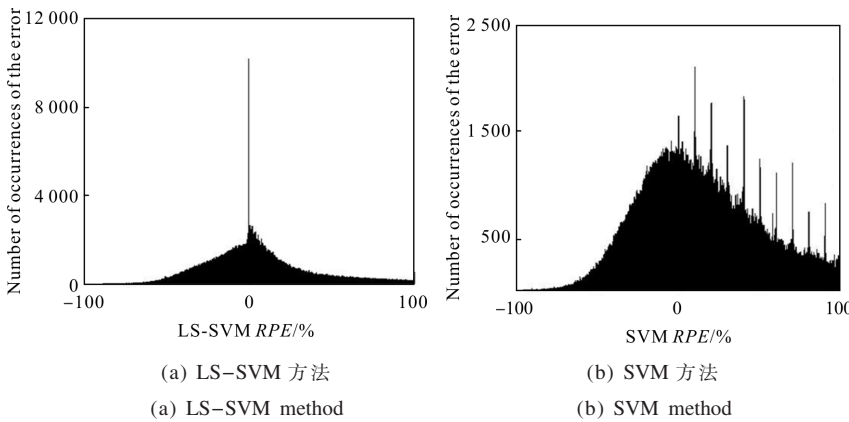


图 3 LS-SVM 和 SVM 预测结果的 RPE 统计直方图

Fig.3 Histograms of RPE based on prediction results of LS-SVM and SVM method

被生长旺期。从图 4 的分段统计中可以得出,LS-SVM 有 55.31%的 APE 在 20%以下,而 SVM APE 在 20%以下的比例只占 42.38%。由此说明 LS-SVM

的精度高于 SVM。

研究区 MAPE 空间分布如图 5(a)和(b)所示。空间误差分布图可以体现 LS-SVM 方法对单个像元

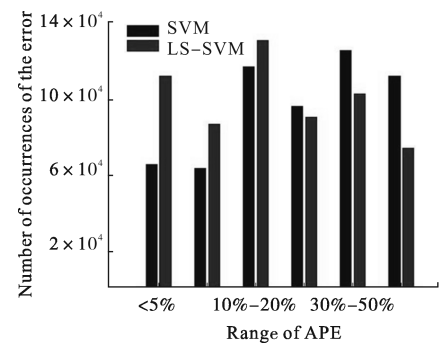


图 4 预测结果的 APE 分段统计柱状图

Fig.4 Histograms of APE segmented statistics based on prediction results

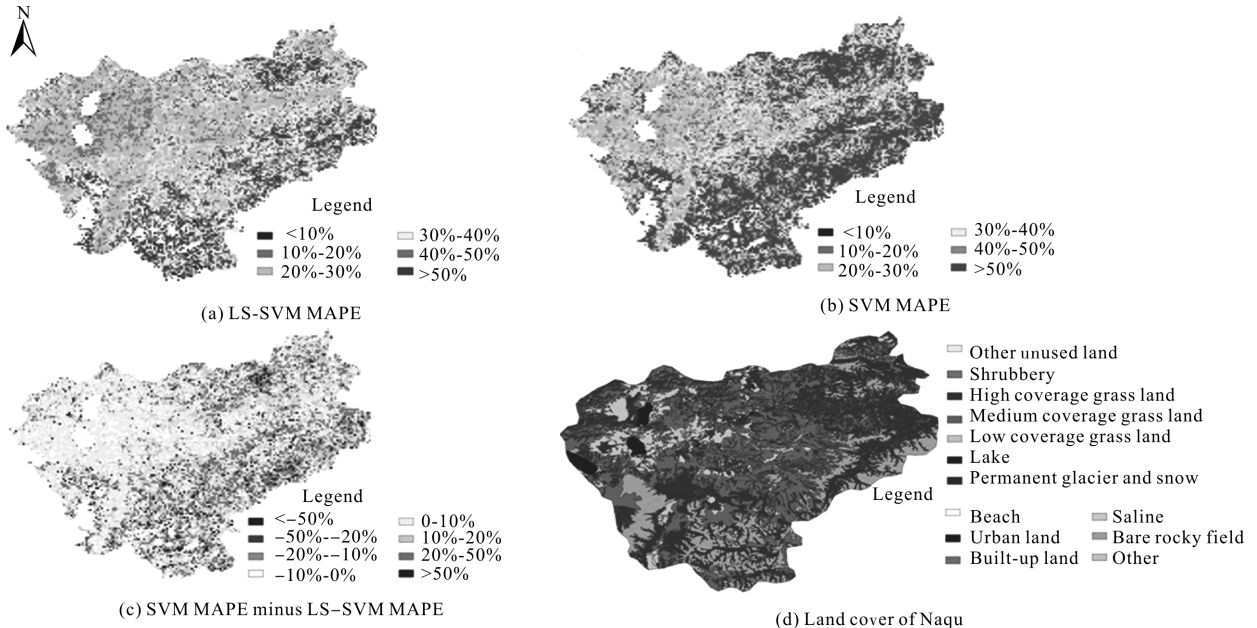


图 5 研究区域空间误差分布情况

Fig.5 Spatial error distribution in study area

的预测精度。为了更直观地比较两种方法,将每个像元 LS-SVM MAPE 对应减去 SVM MAPE,如图 5(c) 所示。

通过图 5(a)和(b)的直观比较,可以看出 SVM 预测误差比 LS-SVM 误差大,且由 5(c)的进一步量化比较得出,MAPE 相减的差值主要分布在 0~20%。将图 5(a)和(b)与研究区域土地覆盖情况(图 5(d))对照,可以看出,预测误差偏大的区域主要在裸岩石砾地。裸岩石砾地指地表为岩石或石砾,植被覆盖度小于 5%的土地。因此该地类不足以反映 LS-SVM 预测 LAI 的能力。

LS-SVM 对草地 LAI 预测精度为:低覆盖度草地>中覆盖度草地>高覆盖度草地。草地覆盖度越

高,LAI 变化则越剧烈,而 LS-SVM 预测的 LAI 曲线非常平滑,因此出现草地覆盖度越高,预测精度越低的现象。通过对单个像元点的依次统计分析,可以确定 LS-SVM 可以对低覆盖度草地时间序列 LAI 进行较好的预测。

2.2 基于地面实测 LAI 的误差分析

地面实验样点共 28 个。分别将 MODIS 原始 LAI 值、LS-SVM 预测的 LAI 值、SVM 预测的 LAI 值与地面实验数据对比,进行验证,结果如图 6,其中(b)和(c)均达到 0.01 水平的极显著相关。三种 LAI 与地面实验数据的 MRPE 分别为 41.19%,32.07%,33.71%。结合图 6 和 MRPE 可以看出,LS-SVM 的预测值比 MODIS 原始值以及 SVM 预测值更接近地面真值。

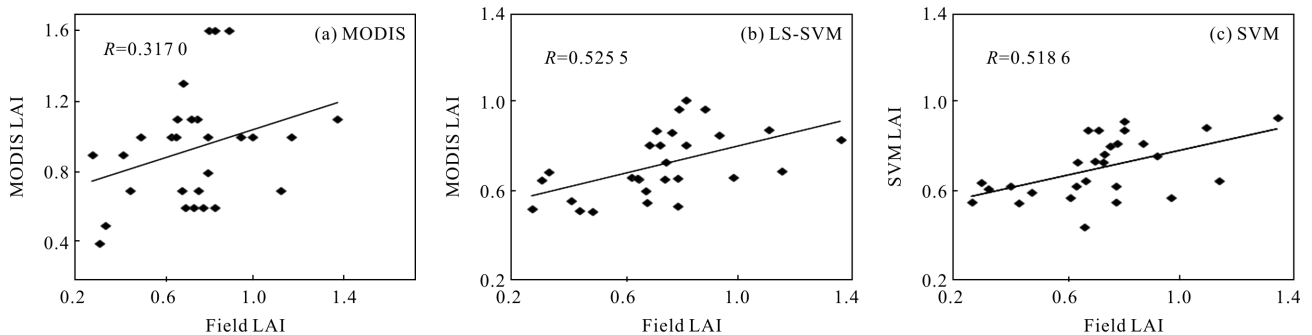


图 6 地面实测样点验证结果

Fig.6 Field calibration result of LAI

2.3 不确定性分析

LS-SVM 对时间序列 LAI 的预测精度还受很多因素的影响,如由于卫星数据获取本身的原因,如云等因素的影响,使得到的遥感卫星数据产品往往会受噪音影响很大^[12]。通过使用 2003-2011 年的产品,剔除所有值都为 0 的像元(异常点),和所有值都大于 250 的像元(MODIS 自身填充的数据)。但为了对研究区域整体进行分析,文中算法对所有像元点依次进行预测,并没有根据 MODIS 产品自身的质量控制文件剔除质量不合格像元。故预测结果中因产品本身质量造成的误差没有进行深入探讨。此外,算法中尚存在一些问题,有待进一步改进,如在植被生长旺盛时期,LS-SVM 的预测结果比原始 LAI 值偏低。模型参数的选择还需继续深入,还需对不同参数设定下的预测结果进行对比分析。

3 结 论

在介绍一种基于最小二乘支持向量机的遥感反

演时间序列 LAI 预测方法的基础上,以西藏那曲县为研究区域,利用 MODIS 历史 LAI 数据构建训练和测试样本,用 LS-SVM 和 SVM 两种方法对研究区域 2011 年 LAI 时间序列进行了预测。将两种预测结果与 MODIS 及实测值比较,结果显示,不论是在单个像元还是整个研究区域,LS-SVM 方法均能比 SVM 方法更好地对遥感反演时间序列 LAI 进行预测,且预测结果与植被实际生长曲线更接近。对于算法中存在的一些不足,在结果分析中进行了讨论,有待进一步改进。

参考文献:

- [1] My neni R B, Hoffman S, Knyazikhin Y, et al. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data [J]. *Remote Sensing of Environment*, 2002, 83(1-2): 214-231.
- [2] Peng D L, Zhang B, Liu L L, et al. Characteristics and

- drivers of global NDVI-based FPAR from 1982 to 2006[S]. *Global Biogeochemical Cycles*, 2012, 26, GB3015.
- [3] Hüttich C, Herold M, Schmullius C, et al. Indicator of Northern Eurasia's land-cover change trends from SPOT-Vegetation time-series analysis 1998-2005 [J]. *International Journal of Remote Sensing*, 2007, 28(18): 4199-4206.
- [4] Röder A, Hill J, Duguay B, et al. Using long time series of Landsat data to monitor fire events and post-fire dynamics and identify driving factors. A case study in the Ayora region (eastern Spain)[J]. *Remote Sensing of Environment*, 2008, 112(1): 259-273.
- [5] Kuusk A. A two-layer canopy reflectance model [J]. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 2001, 71: 1-9.
- [6] Alhamad M N, Stuth J, Vannucci M. Biophysical modeling and NDVI time series to project near-term forage supply: spectral analysis aided by wavelet denoising and ARIMA modelling [J]. *International Journal of Remote Sensing*, 2007, 28: 2513-2548.
- [7] Xin Zhiyun, Gu Ming. Complicated financial data time series forecasting analysis based on least square support vector machine [J]. *Journal of Tsinghua University (Science & Technology)*, 2008, 48(7): 1147-1149. (in Chinese)
辛治运, 顾明. 基于最小二乘支持向量机的复杂金融时间序列预测[J]. *清华大学学报(自然科学版)*, 2008, 48(7): 1147-1149.
- [8] Hou Qingjian, Wang Hongli. Random error coefficient prediction of laser gyro based on LSSVM [J]. *Infrared and Laser Engineering*, 2008, 37(5): 802-805. (in Chinese)
侯青剑, 王宏力. 基于 LSSVM 的激光陀螺随机误差系数预测[J]. *红外与激光工程*, 2008, 37(5): 802-805.
- [9] Ling Xufeng, Yang Jie, Ye Chenzhou. Support vector machine-based human face recognition method[J]. *Infrared and Laser Engineering*, 2001, 30(5): 318-327. (in Chinese)
凌旭峰, 杨杰, 叶晨洲. 基于支撑向量机的人脸识别技术[J]. *红外与激光工程*, 2001, 30(5): 318-327.
- [10] Wang Qunming, Wang Liguang, Liu Danfeng, et al. Sub-pixel mapping for land class with linear least square support vector machine [J]. *Infrared and Laser Engineering*, 2012, 41(6): 1669-1675.
- [11] Wang Peng, Wang Zhicheng, Zhang Jun, et al. Infrared weak small target detection based on multi-scale wavelet kernel LS-SVM [J]. *Infrared and Laser Engineering*, 2006, 35(5): 251-257. (in Chinese)
王鹏, 王志成, 张钧, 等. 基于多尺度小波核 LS-SVM 的红外弱小目标检测[J]. *红外与激光工程*, 2006, 35(5): 251-257.
- [12] Peng D L, Alfredo R. Huete, Huang J F, et al. Detection and estimation of mixed paddy rice cropping patterns with MODIS data [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2001, 13(1): 13-23.