

文章编号: 0253-2239(2010)04-1192-05

基于近红外光谱技术检测除草剂胁迫下油菜叶片中脯氨酸含量的方法

孙光明¹ 刘 飞¹ 张 帆² 金宗来² 何 勇¹ 周伟军² 冯 雷¹

(¹ 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029; ² 浙江大学农业与生物技术学院, 浙江 杭州 310029)

摘要 应用近红外光谱技术实现除草剂胁迫下油菜叶片中脯氨酸含量的检测。对 248 个经过除草剂丙酯草醚处理后的油菜叶片, 经过烘干、磨碎后进行光谱扫描。经过 Savitzky-Golay 平滑、变量标准化(SNV)、二阶求导预处理后, 应用偏最小二乘法(PLS)建立脯氨酸含量的预测模型, 同时提取有效特征变量作为神经网络(BPNN)和最小二乘-支持向量机(LS-SVM)的输入值, 并建立相应的模型。用 186 个样本建模, 62 个样本预测。结果表明, 最小二乘-支持向量机能够获得最优的预测效果, 预测的相关系数(r)、预测标准差(RMSEP)和偏差分别为 0.995, 0.041 和 0.000。说明应用近红外光谱技术结合最小二乘-支持向量机能够定量获得油菜叶片中脯氨酸的含量。

关键词 光谱学; 近红外光谱; 偏最小二乘法; 反向传播神经网络; 最小二乘-支持向量机

中图分类号 O657.33 文献标识码 A doi: 10.3788/AOS20103004.1192

Fast Determination of Proline in Herbicide-Stressed Oilseed Rape Leaves Based on Near Infrared Spectroscopy

Sun Guangming¹ Liu Fei¹ Zhang Fan² Jin Zonglai²
He Yong¹ Zhou Weijun² Feng Lei¹

(¹ College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310029, China)
(² College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, Zhejiang 310029, China)

Abstract Near infrared (NIR) spectroscopy was applied for the fast determination of proline in oilseed rape leavers. The oilseed rape leaves were treated by herbicide, 248 samples were collected for NIR spectral scanning within the wavelength region of 1100~2500 nm. Smoothing way of Savitzky-Golay with 7 segments, standard normal variate (SNV) and second derivative were used as preprocessing methods of spectral data before the calibration stage. Partial least squares (PLS) analysis was applied as calibration method as well as a way to extract the new eigenvectors which could be used to represent the most useful information of original spectra and compress the spectral dimensionality. The selected new eigenvectors were used as the input data matrix of back propagation neural network (BPNN) and least squares-support vector machine (LS-SVM) to develop the BPNN and LS-SVM models. The calibration set was composed of 186 samples, whereas 62 samples in the validation set. The results indicated that LS-SVM model achieved the best prediction performance, and LS-SVM model outperformed PLS and BPNN models. The correlation coefficients (r), root mean square error of prediction (RMSEP) and bias by LS-SVM model were 0.995, 0.041 and 0.000, respectively. The overall results demonstrated that NIR spectroscopy combined with LS-SVM model could be successfully applied for the determination of proline in oilseed rape leaves treated by herbicide.

收稿日期: 2009-05-12; 收到修改稿日期: 2009-07-21

基金项目: 国家 863 计划(2007AA10Z210; 2006AA10Z234)、国家自然科学基金(30671213)、浙江省自然科学基金重点项目(Z3090295)、中央高校基本科研究业务费专项资金和浙江省研究生创新科研项目(YK2008014)资助课题。

作者简介: 孙光明(1986—), 硕士研究生, 主要从事数字农业和农产品光谱检测技术研究。

E-mail: guangming19861113@163.com

导师简介: 何 勇(1963—), 男, 教授, 博士生导师, 主要从事精细农业技术和农业装备自动化方面的研究。

E-mail: yhe@zju.edu.cn

Key words spectroscopy; near infrared spectroscopy; partial least squares analysis; back propagation neural network; least squares-support vector machine

1 引言

油菜是我国四大油料作物之一,占全国油料作物面积的 40%以上,占全国油料作物总产量的 30%以上^[1]。脯氨酸(Proline)作为一种有机溶剂存在于油菜植物体内。当油菜处于逆境胁迫下,由植物体内脯氨酸生物合成激活和脯氨酸降解的抑制共同导致的脯氨酸的积累^[2],并且在逆境压力解除后迅速被分解利用^[3]。普遍认为,逆境胁迫下脯氨酸大量积累可能对植物在压力解除后的恢复有很大帮助,因为脯氨酸可以作为碳源和氮源,并且能够为细胞恢复提供大量的能量^[4]。因此,其含量可以作为植物抗逆性的生理指标之一,脯氨酸量高,植物抗性强^[5]。传统的脯氨酸检测方法为磺基水杨酸法^[6]、酸性茚三酮显色法和比色法^[7]等。这些方法费时、费力、消耗品昂贵,且污染环境。高速氨基酸自动分析仪和高效液相色谱等各种分析仪器需要对样品进行很复杂的前期处理,因此,急需一种简单有效的检测方法来实现除草剂胁迫下油菜叶片中脯氨酸的检测。近红外光谱技术是指利用近红外谱区包含的物质信息,进行有机物质定性和定量分析的一种分析技术^[8~10]。相对于周期长、操作复杂的传统化学测量方法,这是一种快速的物质检测方法。刘飞等^[11,12]应用可见/近红外光谱技术实现了油菜叶片中乙酰乳酸合成酶(ALS)和蛋白质含量的快速无损检测。苗玉奎等^[13]应用近红外光谱技术测定了转基因油菜籽中的芥酸和硫苷。Montes等^[14]研究了在联合收割机上使用近红外光谱仪进行油菜籽品质分析的问题。冯雷等^[15]应用多光谱视觉传感技术检测了油菜中的氮含量。方慧等^[16]应用可见/近红外光谱检测了油菜叶片中叶绿素含量。吴建国等^[17]构建了整粒油菜籽脂肪酸成分近红外反射光谱分析模型。裘正军等^[18]应用 SPAD 和光谱技术研究了油菜生长期间的氮素变化规律。而应用光谱技术对除草剂丙酯草醚胁迫下油菜叶片中脯氨酸含量检测的研究还少有报道。

本文通过对丙酯草醚除草剂处理后的油菜叶片的近红外漫反射光谱进行分析,结合现代化学计量学方法,得到一种光谱技术检测油菜叶脯氨酸含量的方法,为进一步研究油菜在除草剂胁迫下的其它生理指标的变化奠定了基础。

2 材料与方法

2.1 仪器设备

试验使用 Foss NIR Systems 5000(丹麦 Foss 公司)近红外光谱仪,在室温 20 ℃下用小样品杯对所有粉样进行光谱扫描。光谱测定范围为 1100~2500 nm,采样间隔为 2 nm,扫描光谱以 $\lg(1/R)$ 的形式存储;光谱采集软件为 WINISI II V1.5,分析软件为 The Unscrambler V9.6 (CAMO AS, Oslo, Norway), Matlab V7.0 (The Math Works, Natick, USA), DPS(Data Processing System)。

2.2 油菜样本制备及脯氨酸含量的测定

试验所用油菜采自大田,采用一般的田间管理方式。在油菜五叶一心期时进行叶面喷施不同浓度丙酯草醚(0, 100, 500 和 1000 mg/L),每个小区面积为 2.5 m²,共 4 个小区。分别于处理后 7, 14 和 28 d,在各处理小区随机选取 10 株长势均匀的油菜植株,展开叶片按从顶端到底部不同叶位分别选取 5 片叶子(28 d 时对照和 100 mg/L 处理已具有 6 片展开叶,故选取 6 片),分别在 105 ℃烘箱中杀青 30 min 后于 80 ℃下烘干至恒重。烘干后的样品用研钵磨成粉末并过 60 目(245 μm)筛,然后于 4 ℃干燥条件下保存备用。同一处理下相同叶位的叶片混合均匀后分成 4 个样本分别测定。共得到 248 个样本,其中 186 个样本作为建模集样本,62 个样本作为预测集样本。

脯氨酸含量测量方法是用万分之一天平准确称取粉样 0.05 g 置于 10 ml 的安培瓶中,加入 5 ml (6 mol/L) HCl,用循环真空泵抽真空并在酒精喷灯上封口后置于 110 ℃烘箱中消解 24 h;将冷却后的消解液完全转移到 50 ml 容量瓶中定容、摇匀;吸取 5 ml 上清液置于圆底磨口锥形瓶中,在旋转蒸发仪上(60 ℃~70 ℃)抽真空 15 min 左右蒸干盐酸;重新加入 0.02 mol/L 的 HCl(PH 2.2)10 ml,超声波充分溶解后用一次性针筒吸出,经 0.22 μm 的一次性滤头过滤至氨基酸分析专用样品瓶中,4 ℃黑暗保存,采用高速氨基酸自动分析仪(日立 L-8900)分析测定,每个样本重复测定 3 次。

2.3 光谱数据预处理

为了消除原始光谱数据中基线漂移、粉末颗粒不均匀和其他随机噪音的影响,需要对光谱数据进行平滑(Smoothing)、变量标准化(SNV)、一阶及二阶导数处理等预处理。通过对比后,采用 Savitzky-

Golay 平滑法(SG)平滑,平滑点数为 7,然后进行 SNV 处理,再进行二阶导数处理。

3 试验结果与分析

3.1 样品的光谱图及脯氨酸含量统计分析

油菜样品的原始近红外光谱图如图 1,处理后的光谱曲线更加紧凑,如图 2 所示。选取了 1100~2498 nm 波段数据作为分析对象。图中横坐标为光谱波长,纵坐标为吸光度值。1300 nm 和 1900 nm 两波段处的吸光度值存在一定的梯度,这可能是由于油菜叶粉末中脯氨酸含量的不同,造成因物质电子跃迁,转动和振动不同而对近红外区光谱吸收度的不同。这在理论上提供了能够进行定量分析的依据。样品中脯氨酸含量(干重测量方式)的统计值如表 1 所示,单位 mg/100mg DW 表示在干重状态下每

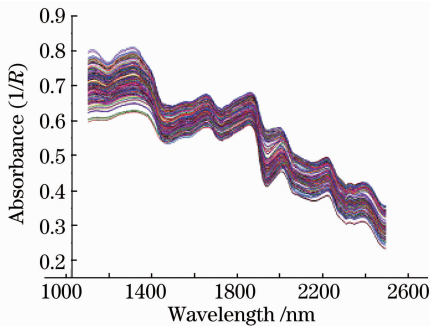


图 1 油菜叶片的近红外光谱原始光谱图

Fig. 1 Original NIR spectra of oilseed rape leaves

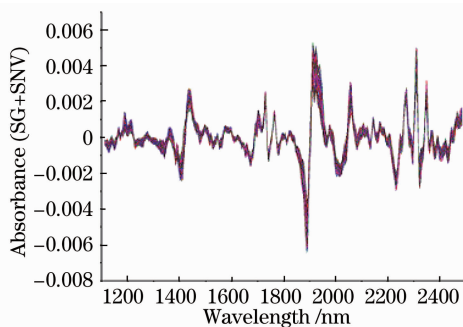


图 2 经过 SG,SNV 和二阶微分后的光谱图

Fig. 2 Preprocessed spectra by the smoothing way of SG, SNV and second derivative

表 1 油菜叶片中脯氨酸含量的统计值/(mg/100 mg DW)
Table 1 Statistical values of in proline oilseed rape leaves
/(mg/100 mg DW)

Data set	Sample No.	Range	Mean	Standard deviation
Calibration	186	0.480~2.667	1.238	0.413
Prediction	62	0.485~2.246	1.225	0.397
All	248	0.480~2.667	1.235	0.409

100 mg 油菜叶片中脯氨酸的含量。可见脯氨酸含量分布均匀,对建立的模型的稳定性和代表性很有帮助。

3.2 偏最小二乘法模型的建立

偏最小二乘法(PLS)可以提取油菜叶片光谱数据的有效信息,结合脯氨酸含量的标准值建立光谱数据和脯氨酸含量的相关关系模型,对预测集样进行预测。建模时采用交互验证法(Cross-validation)来防止模型产生过拟合。模型的相关系数越高,预测标准差和偏差越小,模型的精度越高。采用预处理后的光谱数据,建立了快速检测油菜叶中脯氨酸含量的 PLS 模型。

把预处理后的光谱数据作为输入变量,油菜叶的脯氨酸含量为预测值,建立预测油菜叶脯氨酸含量的 PLS 模型。该模型对预测样本进行预测的相关系数 r 、预测标准差(RMSEP) R 和偏差 B 分别为 0.971,0.097 和 -0.141,说明模型预测结果是可靠的。图 3 为预测集样本的真实值与预测值的散点分布图。横坐标为真实值,纵坐标为 PLS 模型的预测值,预测样本分布于回归直线的两侧。结果表明,应用近红外光谱结合偏最小二乘法,对油菜叶中脯氨酸含量进行预测是可行的。

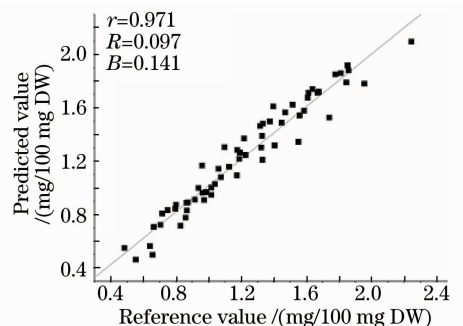


图 3 应用 PLS 模型对油菜叶脯氨酸含量的预测散点分布图

Fig. 3 Predicted value versus reference values of proline in oilseed rape by PLS model

3.3 神经网络模型的建立

PLS 分析计算所有 248 个样本的主成分,由于前 6 个主成分的累计贡献率达到 97%,得到模型的最佳主成分数为 6,选择前 6 个主成分作为神经网络的输入值。神经网络的输出层为油菜叶脯氨酸含量值。该网络采用 Sigmoid 传递函数,通过调整网络参数得到最优的网络结构,减少模型的拟合残差,从而得到最优的预测模型:输入层为 6 个输入单元,隐含层为 4 个隐含单元和 1 个输出层单元的三层反向传播神经网络(BPNN)模型。通过对 186 个建模

样本进行训练,限定最大迭代次数为 2000,允许误差为 0.0001。对预测集 62 个样本进行预测,拟合残差为 1.9×10^{-3} ,对预测样本进行预测的相关系数 $r=0.972$, $R=0.095$, $B=0.007$,预测结果和 PLS 模型相当,但其偏差值小于 PLS 模型。预测集样本的真实值与预测值的散点分布图,如图 4 所示。

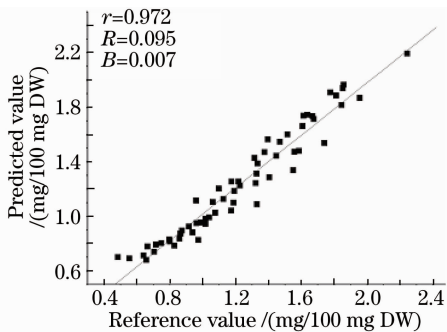


图 4 应用 BPNN 模型对油菜叶脯氨酸含量的预测散点分布图

Fig. 4 Predicted value versus reference value of proline in oilseed rape by BPNN model

3.4 最小二乘-支持向量机模型的建立

最小二乘-支持向量机(LS-SVM) [19~21]通过结构风险最小化原理来提高泛化能力,较好地解决了小样本、非线性、高维数以及局部极小等实际问题,是解决“维数灾”和“过学习”等传统困难的一种有效的手段。

和 BPNN 模型的输入一样,PLS 分析提取原始数据的前 6 个主成分作为最小二乘-支持向量机(LS-SVM)模型的输入值。采用径向基函数(RBF)作为核函数,待定参数:超参数 γ 和 RBF 核函数参数 σ^2 ,这两个参数在很大程度上决定了模型的学习泛化能力和预测能力。采用带有交互验证的网格搜索法(grid-search)得到这两个参数的最优组合。寻优过程由粗选和精选两个步骤完成。依据最小二乘-支持向量机原理和经验, γ 和 σ^2 的取值范围是 $10^{-3} \sim 10^3$ 。模型参数的最优组合为 $\gamma=720.9$, $\sigma^2=428.3$ 。该模型的预测结果如图 5 所示。该模型对预测集样本进行预测,其相关系数 $r=0.995$, $R=0.041$, $B=0.000$,预测精度优于 PLS 模型和 BPNN 模型。三种建模方法的比较,BPNN 和 LS-SVM 模型的预测结果都优于 PLS 模型,PLS 方法只能处理光谱数据中的线性相关关系,而 BPNN 和 LS-SVM 方法可以利用光谱数据中的非线性信息,进而提高了模型的预测精度。而 LS-SVM 模型的预测效果最好,原因是 LS-SVM 能够解决 BPNN 在建模过程中过拟合的缺点而又具有处理非线性和小样本问题

的优点。由预测集样本的真实值与预测值的散点分布图(图 5)可以看到样本在回归直线两侧的分布比另外两种模型更为紧密。

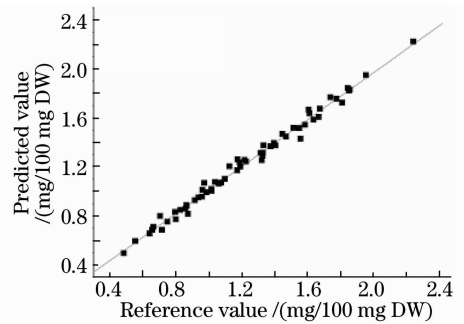


图 5 应用 LS-SVM 模型对油菜叶脯氨酸含量的预测散点分布图

Fig. 5 Predicted value versus reference value of proline in oilseed rape by LS-SVM model

4 结 论

近红外光谱技术结合化学计量学方法(PLS, BPNN 和 LS-SVM),可以实现除草剂胁迫下油菜叶片脯氨酸含量的快速有效的检测。通过对光谱数据进行预处理,得到预测脯氨酸值的 PLS 模型。同时得到近红外光谱数据的主成分,建立预测脯氨酸值的 BPNN 和 LS-SVM 模型。通过对三种模型的对比,最优模型为 LS-SVM 模型,其预测的相关系数 r , R 和 B 分别为 0.995, 0.041 和 0.000,说明应用近红外光谱技术检测油菜叶中脯氨酸含量是可行的,能获得满意的预测精度。

参 考 文 献

- Zhang Guoping, Zhou Weijun. Crop Production[M]. Hangzhou: Zhejiang University Press, 2001
- 张国平,周伟军. 作物栽培学[M]. 杭州:浙江大学出版社, 2001
- K. Nakashima, R. Satoh, T. Kiyosue *et al.*. A gene encoding proline dehydrogenase is not only induced by proline and hypoosmolarity, but is also developmentally regulated in the reproductive organs of arabidopsis[J]. *Plant Physiology*, 2008, **118**(4): 1233~1241
- P. Trotef-Aziz, M. F. Niogret, C. Deleu *et al.*. The control of proline consumption by abscisic acid during osmotic stress recovery of canola leaf discs[J]. *Physiologia Plantarum*, 2003, **117**(2): 213
- P. D. Hare, W. A. Cress *et al.*. A regulatory role for proline metabolism in stimulating arabidopsis thaliana seed germination [J]. *Plant Growth Regulation*, 2003, **39**(1): 41~50
- Wei Dongping, Liu Peng, Xu Gendi *et al.*. Response of protectase activity, MDA and proline contents to aluminum-stress in leaves of rape seedlings[J]. *J. the Chinese Cereals and Oils Association*, 2008, **23**(5): 107~111
- 韦冬萍,刘鹏,徐根娣等. 油菜叶片保护酶活性、丙二醛及脯氨酸对铝胁迫的响应[J]. 中国粮油学报, 2008, **23**(5):

107~111

- 6 Zhang Guangli, Nie Rongbang. Study on judgement of maturity degree by free proline content in flue-cured tobacco leaves[J]. *Crop Reseach*, 2008, **22**(1): 31~33
张光利, 聂荣邦. 以烟叶脯氨酸含量判断田间成熟度的研究[J]. 作物研究, 2008, **22**(1): 31~33
- 7 Zhi Mingxing, Li Xiuju. Improvement on the method for measuring proline content[J]. *Plant Physiology*, 2005, **41**(3): 355~357
职明星, 李秀菊. 脯氨酸测定方法的改进[J]. 植物生理学通讯, 2005, **41**(3): 355~357
- 8 Yan Yanlu, Zhao Longlian, Han Donghai *et al.*. The Foundation and Application of Near Infrared Spectroscopy Analysis[M]. Beijing: China Light Industry Press, 2005
严衍禄, 赵龙莲, 韩东海等. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005
- 9 Liu Fei, He Yong, Wang Li. Methods for the prediction of sugar content of rice wine using visible-near infrared spectroscopy [J]. *Acta Optica Sinica*, 2007, **27**(11): 2054~2058
刘飞, 何勇, 王莉. 黄酒糖度预测的可见-近红外光谱方法研究[J]. 光学学报, 2007, **27**(11): 2054~2058
- 10 Zhao Jiewen, Guo Zhiming, Chen Quansheng *et al.*. Feasibility study on use of near-infrared spectroscopy in quantitative analysis of catechins in green tea [J]. *Acta Optica Sinica*, 2008, **28**(12): 2302~2307
赵杰文, 郭志明, 陈全胜. 近红外光谱法检测绿茶中儿茶素的含量[J]. 光学学报, 2008, **28**(12): 2302~2307
- 11 Liu Fei, Fang Hui, Zhang Fan *et al.*. Non-destructive determination of acetolactate synthase in oilseed Rape leaves using visible and near infrared spectroscopy[J]. *Chinese J. Analytical Chemistry*, 2009, **37**(1): 67~71
刘飞, 方慧, 张帆等. 应用光谱技术无损检测油菜叶片中乙酰乳酸合成酶[J]. 分析化学, 2009, **37**(1): 67~71
- 12 F. Liu, F. Zhang, Z. L. Jin *et al.*. Determination of acetolactate synthase activity and protein content of oilseed rape (*brassica napus* L.) leaves using visible/near-infrared spectroscopy [J]. *Analytica Chimica Acta*, 2008, **629**(1,2): 56~65
- 13 Rui Yukui, Huang Kunlun, Wang Weimin *et al.*. Spectroscopy and Spectral Analysis, 2006, **26**(12): 2190~2192
芮玉奎, 黄昆仑, 王为民等. 近红外光谱技术在检测转基因油菜籽中芥酸和硫甙上的应用研究[J]. 光谱学与光谱分析, 2006, **26**(12): 2190~2192
- 14 J. M. Montes, C. Paul, A. E. Melchinger. Quality assessment of rapeseed accessions by means of near-infrared spectroscopy on combine harvesters[J]. *Plant Breeding*, 2007, **126**: 329~330
- 15 Feng Lei, Fang Hui, Zhou Weijun *et al.*. Nitrogen stress measurement of canola based on multi-spectral charged coupled device imaging sensor[J]. *Spectroscopy and Spectral Analysis*, 2006, **26**(9): 1749~1752
冯雷, 方慧, 周伟军等. 基于多光谱视觉传感技术的油菜氮含量诊断方法研究[J]. 光谱学与光谱分析, 2006, **26**(9): 1749~1752
- 16 Fang Hui, Song Haiyan, Cao Fang *et al.*. Study on the relationship between spectral properties of oilseed rape leaves and their chlorophyll content [J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(9): 1731~1734
方慧, 宋海燕, 曹芳等. 油菜叶片的光谱特征与叶绿素含量之间的关系研究[J]. 光谱学与光谱分析, 2007, **27**(9): 1731~1734
- 17 Wu Jianguo, Shi Chunhai, Zhang Haizhen *et al.*. Study on developing calibration models of fat acid composition in intact rapeseed by near infrared reflectance spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2006, **26**(2): 259~262
吴建国, 石春海, 张海珍等. 构建整粒油菜籽脂肪酸成分近红外反射光谱分析模型的研究[J]. 光谱学与光谱分析, 2006, **26**(2): 259~262
- 18 Qiu Zhengjun, Song Haiyan, He Yong *et al.*. Variation rules of the nitrogen content of the oilseed rape at growth stage using SPAD and visible-NIR [J]. *Transaction of the CSAE*, 2007, **23**(7): 150~154
裘正军, 宋海燕, 何勇等. 应用 SPAD 和光谱技术研究油菜生长期间的氮素变化规律[J]. 农业工程学报, 2007, **23**(7): 150~154
- 19 V. N. Vapnik. The Nature of Statistical Learning Theory[M]. New York: Springer-Verlag, 1995
- 20 Guo Hui, Liu Heping, Wang Ling. Method for selecting parameters of least squares support vector machines and application[J]. *J. System Simulation*, 2006, **18**(7): 2033~2037
郭辉, 刘贺平, 王玲. 最小二乘支持向量机参数选择方法及其应用研究[J]. 系统仿真学报, 2006, **18**(7): 2033~2037
- 21 J. A. K. Suykens, J. Vanderwalle. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, **9**(3): 293~300