

doi:10.19306/j.cnki.2095-8110.2020.05.017

基于SVR的导航传感器自适应野值检测方法

戴海发, 卞鸿巍, 马恒, 王荣颖

(海军工程大学电气工程学院, 武汉 430033)

摘要:针对Lagrange插值法无法处理连续野值的问题,提出了一种基于改进支持向量回归(Support Vector Regression, SVR)的导航传感器自适应野值检测方法。该方法结合了支持向量回归利用小样本数据就能够准确建模和 3σ 准则计算简便的优点,利用支持向量回归在线建立舰船的运动模型对测量值进行实时预测,并利用 3σ 准则自适应地计算阈值,然后通过比较阈值与预测残差来判别测量值是否为野值点。该方法可以自动地学习舰船的运动趋势,建立舰船的真实运动模型,而且不受连续野值点的影响,能够在没有其他传感器辅助的条件下完成野值检测。海试实测数据表明,提出的方法对离散和连续的野值点均具有较好的检测效果,同时可以更好地估计传感器的真实测量值。

关键词:导航传感器;野值剔除;Lagrange插值法;支持向量回归;自适应

中图分类号:U249.3

文献标志码:A

开放科学(资源服务)标识码(OSID):

文章编号:2095-8110(2020)05-0141-06



Adaptive Navigation Sensor Outlier Detection Method Based on SVR

DAI Hai-fa, BIAN Hong-wei, MA Heng, WANG Rong-ying

(Collage of Electrical Engineering, Naval University of Engineering, Wuhan 430033, China)

Abstract: Aiming at the problem that Lagrange interpolation can't deal with continuous outliers, this paper proposes an adaptive outlier detection method for navigation sensors based on improved support vector regression (SVR). This method combines the advantages of support vector regression with small sample data for accurate modeling and 3σ for simple calculation. The support vector regression is used to establish the ship's motion model online to predict the measured values in real time, and adaptively using 3σ criteria. The threshold is calculated, and then the measured value is compared to the predicted residual to determine whether the measured value is an outlier. The method can automatically learn the movement trend of the ship, establish the real motion model of the ship, without being affected by the continuous outlier, and can perform the outlier detection without other sensors assistance. The sea-tested data shows that the proposed method has a good detection effect on both discrete and continuous outliers, and at the same time obtains a better estimate of the true measured value of the sensor.

Key words: Navigation sensor; Outlier elimination; Lagrange interpolation; Support vector regres-

收稿日期:2019-05-31;修订日期:2019-06-13

基金项目:国家自然科学基金(41876222)

作者简介:戴海发(1990-),男,博士生,主要从事组合导航技术方面的研究。E-mail: daihaifa1990@163.com

通信作者:卞鸿巍(1972-),男,教授,博士生导师,主要从事惯性导航技术、极区导航技术方面的研究。

E-mail: travisbian@foxmail.com

sion; Adaptive

0 引言

综导系统是舰艇收集各导航传感器数据、进行集中处理并分发到各个用户的重要部件。舰载导航传感器的种类和数量较多,再加上海洋和电磁环境复杂多变,因此导航传感器的实际测量数据总是存在一些异常的值,这些异常测量值一般称为野值。如果不对这些野值进行处理,直接输入到综导系统的信息集成模块进行融合,就会影响信息融合的结果,甚至使融合结果完全不可信。因此,在信息融合之前对野值进行检测和处理具有相当重要的必要性。

在实际工程上,处理野值的方法主要有两种思路:一是利用外部传感器的辅助进行野值的检测。主要的方法有基于差值的 3σ 准则法、基于卡尔曼滤波器的 χ^2 检验法^[1-4]、基于小波分析的方法^[5-6]等,其中 3σ 准则法需要基准数据,而 χ^2 检验法依赖准确的系统噪声和测量噪声矩阵,基于小波分析的方法对渐变故障的检测效果不理想。上述方法最大的问题是都依赖稳定无野值的参考系统,如果参考系统也有可能产生野值,那么在检测到野值时,将无法判断具体是哪个系统发生了信息故障。

另一种思路是只利用传感器自身的历史测量值自主地判断野值。工程上比较常用的是差分技术、时间多项式外推法、ARMA 模型预测法^[7]、Lagrange 插值法^[8]等。差分技术利用前后时刻的值作差分,通过判断差分值是否合理来判断测量值是否为野值,该方法只对粗大的野值点有效。ARMA 模型预测法通过历史数据训练 ARMA 模型用于预测当前测量值,但是要求测量数据满足平稳性条件。Lagrange 插值法利用历史数据通过插值公式预测当前测量值,对于野值离散的野值点效果较好,但是对于野值连续的情况则效果不理想,有可能出现严重的失真现象。文献[9-10]提出了基于模糊系统的遥感数据野值检测方法,但是模糊规则的确定需要一定的技巧,限制了它的使用范围。

因此,本文提出了一种基于支持向量回归(Support Vector Regression, SVR)的导航传感器野值剔除方法。与 Lagrange 插值法相比,SVR 可以自动地学习舰船的运动趋势,连续受野值点的影响较小,而且能够在没有其他传感器辅助的条件下完成野值检测。该方法的思路是利用在线训练好

的模型预测当前的导航参数,如果测量值偏离预测值(残差)且超过一定限度,则认为测量值是异常值,并用预测值代替测量值;如果连续检测出异常值超过长 T ,则认为系统发生故障,除非故障被清除,否则不再使用该设备的信息。此外,传统的阈值一般是靠经验设定,而本文利用了历史残差数据计算出残差的统计量,然后利用 3σ 准则构建阈值。

本文首先介绍了 SVR 及算法的基本原理,然后推导了基于 SVR 的野值剔除方法,并通过实际海试数据验证了本文所提方法的有效性。

1 SVR 原理及算法

利用支持向量机作回归分析的原理是通过非线性映射 $x \rightarrow \Phi(x)$,将多维非线性关系样本组 x 投影到高维特征空间 G 中变成线性关系,然后在这个高维特征空间中进行线性回归。给定 n 个数据样本 $\{x_i, y_i\}, i=1, 2, \dots, n$,其中, x_i 为实际观测值, y_i 为期望值。采用式(1)来估计 SVR 的回归函数 f ^[11]

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

$$\Phi: \mathbf{R}^n \rightarrow G, \omega \in G$$

式中, b 为偏置量; ω 为高维特征空间 G 中的矢量; \mathbf{R}^n 为 n 维实数空间。

如图 1 所示,SVR 回归的目标是找到一个超平面使得到超平面最远的样本点距离最小,因此得到优化目标函数为

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

满足以下条件

$$\begin{cases} y_i - (\omega, \Phi(x)) - b \leq \varepsilon + \xi_i \\ (\omega, \Phi(x)) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i=1, 2, \dots, k \end{cases} \quad (3)$$

其中, C 为惩罚因子; ξ_i 和 ξ_i^* 为松弛因子; ε 为损失函数。

上述优化模型可以看成是一个二次规划问题,利用 Lagrange 对偶形式,并引入核函数表达式,则式(2)可转化为

$$\max_{\alpha_i, \alpha_i^*} \sum_{i=1}^n (\alpha_i + \alpha_i^*) (y_i - \varepsilon) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) \quad (4)$$

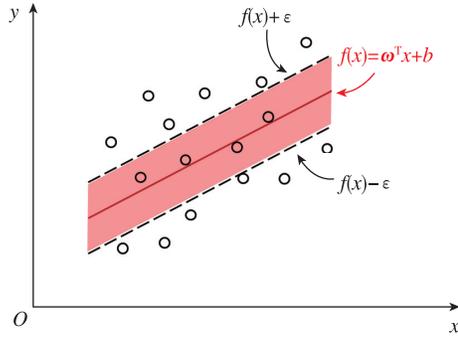


图 1 SVR 示意图

Fig. 1 Support vector regression diagram

其约束条件为

$$\begin{cases} \omega = \sum_i^n (\alpha_i - \alpha_i^*) \Phi(x_i) \\ 0 \leq \alpha_i, \alpha_i^* \leq C \\ \sum_i^n (\alpha_i - \alpha_i^*) = 0 \end{cases} \quad (5)$$

其中, α_i 和 α_i^* 为支持向量的加权系数; K 为核函数。通过式(4)可求得 α_i 和 α_i^* , 将式(5)代入式(1)可以得到 SVR 的解形如

$$f(x) = \sum_i^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (6)$$

上述过程需满足 KKT(Karush-Kuhn-Tucker) 条件^[12], 即

$$\begin{cases} \alpha_i (f(x_i) - y_i - \epsilon - \xi_i) = 0 \\ \alpha_i^* (y_i - f(x_i) - \epsilon - \xi_i^*) = 0 \\ \alpha_i \alpha_i^* = 0, \xi_i \xi_i^* = 0 \\ (C - \alpha_i) \xi_i = 0, (C - \alpha_i^*) \xi_i^* = 0 \end{cases} \quad (7)$$

由式(7)可以看出, 对任意一个样本 (x_i, y_i) 都有 $(C - \alpha_i) \xi_i = 0$, 且 $\alpha_i (f(x_i) - y_i - \epsilon - \xi_i) = 0$ 。于是, 在得到 α_i 后, 若 $0 < \alpha_i < C$, 则必有 $\xi_i = 0$, 从而有

$$b = y_i + \epsilon - \sum_i^n (\alpha_i - \alpha_i^*) K(x_i, x) \quad (8)$$

2 基于 SVR 的野值剔除方法

导航设备的当前导航参数与过去的导航参数具有相关性, 理论上可以找到某个函数来表达这种关系, 但是这种函数往往是高度非线性的, 关于模型的阶次缺少先验知识, 而且相关的自变量一般是多维的, 传统的回归方法很难解决这类问题。此外, 对于样本 (x, y) , 传统回归模型通常直接基于模型输出 $f(x)$ 与真实输出 y 之间的差别来计算损失, 当且仅当 $f(x)$ 与 y 完全相同时, 损失才为零;

与此不同, SVR 假设能容忍 $f(x)$ 与 y 之间最多有 ϵ 的偏差, 即仅当 $f(x)$ 与 y 之间的差值绝对值大于 ϵ 时才计算损失。正是由于这一特点, 使得 SVR 避免了过拟合的问题^[13]。因此, 本文采用了 SVR 的方法来建立这样的函数关系。

对于导航传感器的输出时间序列 $\{x_1, x_2, \dots, x_t\}$, 如果根据 x_t 的历史数据预测当前时刻的数据 x_{t+1} , 则可以建立映射 $f: R^m \rightarrow R$, 满足^[14]

$$\hat{x}_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-(m-1)}) \quad (9)$$

式中, m 为预测函数中使用的历史数据的个数, 即模型的阶次。

为了利用第 1 节的 SVR 算法来训练预测函数, 需要采用窗口大小为 N 的滑动窗技术, 并对 k 时刻存储的历史窗数据序列 $\{x_{k-N}, x_{k-N+1}, \dots, x_{k-1}\}$ 进行重构, 将数据序列转化为矩阵形式, 作为 SVR 的学习样本。构造的学习样本 (\mathbf{X}, \mathbf{Y}) 为

$$\begin{cases} \mathbf{X} = \begin{bmatrix} x_{k-N} & x_{k-N+1} & \dots & x_{k-N+m-1} \\ x_{k-N+1} & x_{k-N+2} & \dots & x_{k-N+m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{k-m} & x_{k-m+1} & \dots & x_{k-2} \end{bmatrix} \\ \mathbf{Y} = \begin{bmatrix} x_{k-N-m} \\ x_{k-N-m+1} \\ \vdots \\ x_{k-1} \end{bmatrix} \end{cases} \quad (10)$$

对 SVR 进行训练的回归函数为

$$Y_t = \sum_{i=1}^{N-m} (\alpha_i - \alpha_i^*) K(X_i, X_t) + b \quad t = 1, 2, \dots, N - m \quad (11)$$

其中, X_i 表示矩阵 \mathbf{X} 的第 i 行, Y_t 表示矩阵 \mathbf{Y} 的第 t 行。本文选取的核函数为高斯径向基核函数, 其表达式为

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \quad (12)$$

其中, $\|\cdot\|$ 为范数运算符; σ 为函数的宽度参数。

通过对样本的学习可以得到 SVR 的模型, 从而得到当前时刻的预测值为

$$\hat{y}_k = \sum_{i=1}^{N-m} (\alpha_i - \alpha_i^*) K(X_i, X_k) + b \quad (13)$$

式中, $X_k = [x_{k-m}, x_{k-m+1}, \dots, x_{k-1}]$ 。

计算预测值与实际测量值之间的差值

$$\delta x_k = \hat{y}_k - \tilde{x}_k \quad (14)$$

野值判定原则为

$$\begin{cases} \delta x_k > T, \tilde{x}_k \text{ 为野值点} \\ \delta x_k < T, \tilde{x}_k \text{ 为非野值点} \end{cases} \quad (15)$$

式中, T 为阈值, 关于阈值的选取, 文献[14]是按照经验选取的, 缺少理论依据。本文结合 3σ 准则, 计算滑动窗内残差的均值 $\bar{\delta x}$ 和标准差 $\sigma_{\delta x}$

$$\begin{aligned} \bar{\delta x} &= \frac{1}{N} \sum_{i=1}^N \delta x_i \\ \sigma_{\delta x} &= \frac{1}{N-1} \sqrt{\sum_{i=1}^N (\delta x_i - \bar{\delta x})^2} \end{aligned} \quad (16)$$

进而确定阈值为

$$T = 3\sigma_{\delta x} \quad (17)$$

通常对于野值的处理都是直接剔除, 然而这样会导致信息的缺失, 不利于后续的信息融合。因此, 本文选择重构的思路, 具体的重构方法为当检测出测量值为野值时, 利用 SVR 的预测结果代替测量值。

3 试验与分析

为了验证本文所提出算法的有效性, 选取船舶的实测数据做实验, 并与 Lagrange 插值法进行对比。实测数据是在某次航海试验中采集得到的数据, 船上搭载的导航设备有惯导、卫导、电磁计程仪和天文导航系统等。海试的航行轨迹如图 2 所示。高斯径向基核函数的宽度参数取为 $\sigma = 4$; 损失函数 $\epsilon = 0.01$; 滑动窗宽 $N = 20$; 模型阶数 $m = 5$ 。本文所使用的算法是在开源代码库 LibSVM 的基础上进行修改实现的, 使用的仿真平台为 MATLAB。

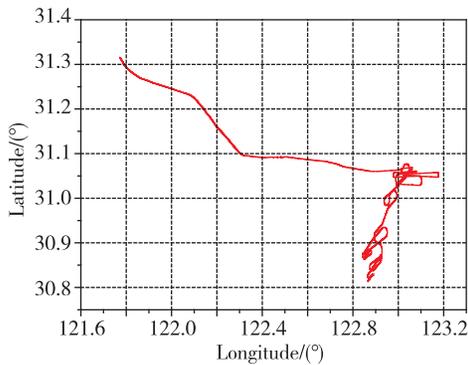


图 2 海试航行轨迹

Fig. 2 Sea trial sailing track

以电磁计程仪的速度信息为例(图 3), 原始测量数据中含有较多的野值点。取其中长度均为 1000

个样本的 2 段数据进行分析, 其中一段只含有少量离散的野值点, 另一段含有几处连续的野值点, 分别利用 Lagrange 插值和 SVR 的方法对数据进行野值处理。

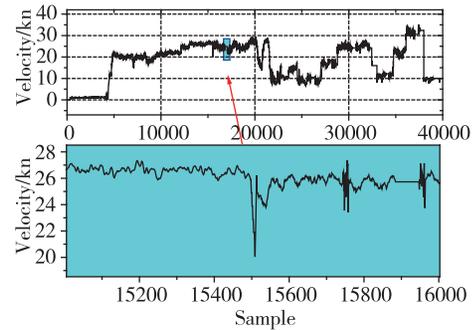


图 3 电磁计程仪原始数据

Fig. 3 Electromagnetic log raw data

1) 场景 1

在场景 1 中, 样本中的速度测量值含有少量离散的点, 且速度保持在基本平稳的状态下, Lagrange 插值法的插值点数设为 $n = 2$, 得到的结果如图 4 所示。

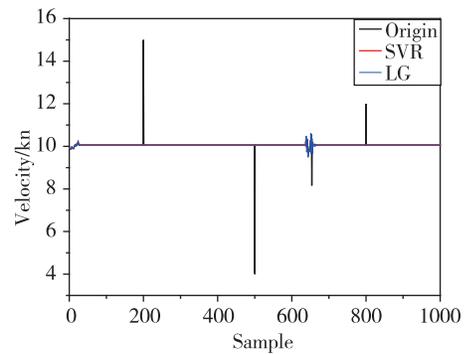


图 4 场景 1 不同野值处理方法结果对比

Fig. 4 Comparison of different outliers processing methods in scenario 1

从图 4 中可以看出, 在该场景中 Lagrange 插值法和 SVR 算法都能够有效地去除离散的野值点, 而且 SVR 算法对波动值的估计值更加平滑。从表 1 可以看出, SVR 算法的计算时间比 Lagrange 插值法只多了 11ms, 但是检测成功率提高了 8%。这表明 SVR 算法建立的预测模型比较准确, 在一定程度上能够抑制噪声的影响。

表 1 不同检测方法的性能对比

Tab. 1 Performance comparison of different detection methods

	计算时间/ms	检测成功率/%
Lagrange 插值法	32	90
SVR 算法	43	98

2) 场景 2

如图 5 所示,场景 2 中,舰船运动状态不稳定,存在加速和减速运动,速度的测量值存在 3 处连续的野值群(如图 6 中蓝色虚线椭圆所标示)。此时,不能够确定 Lagrange 插值法的插值点数,因此分别设定了几种插值点方案,得到的结果与真实测量值的对比如图 5 所示。图 6 展示了使用本文所提出的方法估计的结果与真实测量值的对比情况。

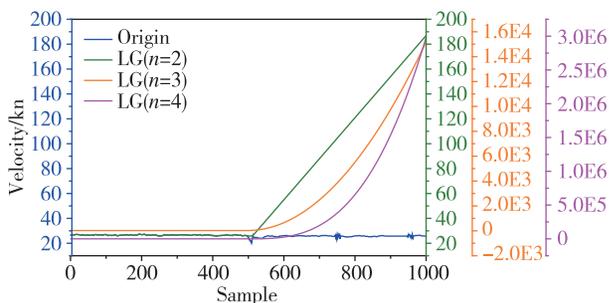


图 5 场景 2 Lagrange 插值结果

Fig. 5 Lagrange interpolation results in scenario 2

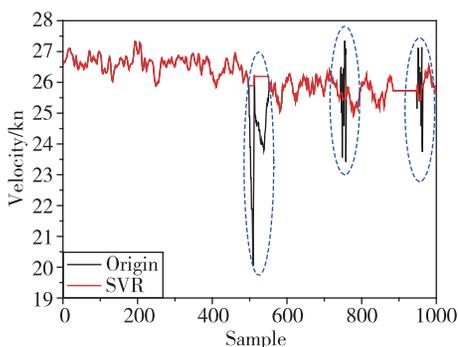


图 6 场景 2 SVR 结果

Fig. 6 SVR results in scenario 2

从图 5 可以看出,利用 Lagrange 插值法得到的结果刚开始与实际测量值比较接近,说明该方法能够剔除少数离散的野值点;但是从第 500 个数据点开始,估计值逐渐偏离真实值,而且随着插值点数量 n 的增加,偏离得越激烈。这是因为此时出现了较多连续的野值点,而 Lagrange 插值法会将误差逐

渐累积,插值点个数越高,模型的阶次越高,误差膨胀得越快。在 1000s 时,速度误差分别达到 160m/s、 1.6×10^4 m/s、 3×10^6 m/s,已经完全失去了修正功能,这说明任意插值点数的 Lagrange 插值法无法检测出连续的野值点。

从图 6 可以看出,蓝色椭圆虚线标示出的 3 处连续的野值已经被基于 SVR 的方法有效剔除,得到的结果与真实测量值比较吻合,而且平滑性较好,较好地反映了舰船的真实运动速度。这表明基于 SVR 的方法对于导航传感器连续野值点的情况具有较好的抑制作用。表 2 展示了不同野值检测方法在场景 2 下的计算时间和检测成功率,从表 2 中可以看出,SVR 算法在该场景下依然具有较少的计算时间和较高的检测成功率。

表 2 不同检测方法的性能对比

Tab. 2 Performance comparison of different detection methods

		计算时间/ms	检测成功率/%
Lagrange 插值法	$n=2$	32	10
	$n=3$	34	5
	$n=4$	40	5
SVR 算法		45	95

4 结论

针对 Lagrange 插值法无法处理连续野值的问题,本文提出了一种基于改进 SVR 的导航传感器自适应野值检测方法。该方法结合了 SVR 只需要小样本数据就能够准确建模和 3σ 准则计算简便的优点,利用 SVR 在线建立舰船的运动模型对测量值进行实时预测,并利用 3σ 准则自适应地计算阈值;然后通过比较阈值与预测残差来判别测量值是否为野值点,如果测量值被检测为野值点,则用 SVR 的预测值替代该野值点。海试实测数据表明,本文提出的方法能够在存在连续野值点的情况下检测出野值点,同时得到较好的传感器真实测量值的估计,而且算法简单容易实现,具有较好的工程应用价值。

参考文献

[1] Wang R, Xiong Z, Liu J, et al. Chi-square and SPRT combined fault detection for multisensor navigation [J]. IEEE Transactions on Aerospace and Electronic Systems, 2016, 52(3): 1352-1365.
 [2] Ma H S, Wang D, Sun X M, et al. Research on

- tightly coupled IRS/GPS integrated arithmetic based on residual chi-square fault detection[J]. Avionics Technology, 2013, 44(4): 24-29.
- [3] Wen R. Improved residual χ^2 inspection in integrated navigation fault detection of application[J]. Electronic Measurement Technology, 2012, 35(11): 124-128.
- [4] Liu L, Fu J. Improved state- χ^2 fault detection of navigation systems based on neural network[C]// Proceedings of 2010 Chinese Control and Decision Conference(CCDC), 2010: 3932-3937.
- [5] Ding D F, Zhao X M. Navigation system failure detection based on FM gauss wavelet transforms[J]. Applied Mechanics & Materials, 2014, 602-605: 1874-1878.
- [6] Lin X, Wang J, Luo H. Fault detection and information fusion method of GPS/SINS integrated navigation system based on wavelet aided [J]. Ordnance Industry Automation, 2012, 31(2):14-16.
- [7] 颜东, 张洪铖. 基于辨识 ARMA 模型的野值剔除方法与卡尔曼滤波修正算法[J]. 信息与控制, 1995, 24(3): 183-188.
Yan Dong, Zhang Hongcheng. The outlier elimination method based on identification ARMA model and Kalman filter correction algorithm[J]. Information and Control, 1995, 24(3): 183-188(in Chinese).
- [8] 雷雨, 赵丹宁, 高玉平, 等. 基于滑动式 Lagrange 插值方法的 GPS 精密星历内插分析[J]. 测绘工程, 2013, 22(2): 34-36.
Lei Yu, Zhao Danning, Gao Yuping, et al. GPS precise ephemeris interpolation analysis based on sliding Lagrange interpolation method[J]. Survey and Mapping Engineering, 2013, 22(2): 34-36(in Chinese).
- [9] 朱学锋. 基于聚类模糊系统的动态数据野值剔除方法[J]. 飞行器测控学报, 2011, 30(5): 81-84.
Zhu Xuefeng. Dynamic data outlier elimination method based on clustering fuzzy system [J]. Journal of Spacecraft Measurement and Control, 2011, 30(5): 81-84(in Chinese).
- [10] 朱学锋, 韩荣阁, 杨若红. 基于模糊预测系统的观测数据野值剔除方法[J]. 系统工程与电子技术, 2006, 28(3): 478-482.
Zhu Xuefeng, Han Rongge, Yang Ruohong. Elimination method of observation data outlier based on fuzzy prediction system[J]. Systems Engineering and Electronics, 2006, 28(3): 478-482(in Chinese).
- [11] Ma J, Theiler J, Perkins S. Accurate on-line support vector regression[J]. Neural Computation, 2003, 15(11): 2683-2703.
- [12] Duan L, Xie M, Bai T, et al. A new support vector data description method for machinery fault diagnosis with unbalanced datasets[J]. Expert Systems with Applications, 2016, 64: 239-246.
- [13] 王定成, 方廷健, 高理富, 等. 支持向量机回归在线建模及应用[J]. 控制与决策, 2003, 18(1): 89-91.
Wang Dingcheng, Fang Tingjian, Gao Lifu, et al. Online modeling and application of support vector machine regression[J]. Control and Decision, 2003, 18(1): 89-91(in Chinese).
- [14] 柳敏, 赖际舟, 刘建业, 等. 基于 SVR 的惯性/卫星组合导航系统故障诊断方法[J]. 控制与决策, 2016, 31(10): 1-5.
Liu Min, Lai Jizhou, Liu Jianye, et al. Fault diagnosis method for inertial/satellite integrated navigation system based on SVR[J]. Control and Decision, 2016, 31(10): 1-5(in Chinese).