

模糊支持向量机的非直达波抑制算法

万 群, 王 伟, 黄际彦, 宋玉梅

(电子科技大学电子工程学院 成都 610054)

【摘要】针对支持向量机中由于噪声和野值带来的过拟合问题,提出了基于模糊最小二乘支持向量机(fuzzy LS-SVM)的抑制非直达波的移动定位方法。利用一种基于支持向量数据域描述的模糊隶属度函数模型,根据样本到特征空间最小包含超球球心的距离来确定模糊隶属度。仿真结果表明了该方法的稳健性,提高了LS-SVM的抗噪声能力。

关键词 模糊隶属度; 最小二乘支持向量机; 非直达波定位; 支持向量机数据域描述
中图分类号 TN929.53 **文献标识码** A

Method for NLOS Mitigation Based on Fuzzy Support Vector Machines

WAN Qun, WANG Wei, HUANG Ji-yan, and SONG Yu-mei

(School of Electronic Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract In order to overcome the overfitting problem caused by noises and outliers in support vector machine, a method for non-line-of-sight (NLOS) mitigation based on fuzzy least square support vector machines (LS-SVM) is proposed. Using the fuzzy membership model based on support vector data description (SVDD), the membership values to each input sample is computed according to its distance to the center of the hypersphere with minimal volume containing all objects. Simulation results show that the proposed method is robust in NLOS environments and actually increases the accuracy of LS-SVM.

Key words fuzzy membership; LS-SVM; NLOS location; SVDD

近年来,蜂窝网络的移动定位问题受到国内外学者日益密切的关注。在蜂窝移动定位系统中,非直达波传播成为阻碍定位精度提高的主要因素^[1]。非直达波传播带来的误差由环境决定,随着环境不同而呈不同的误差分布,抑制非直达波实质上就是抑制环境对定位精度的影响。

目前抑制非直达波误差的方法主要有:(1)利用散射体信息^[2];(2)进行加权和约束^[3];(3)直达波鉴别^[4]。由于非直达波误差是由环境决定的,上述几何的方法不能有效地抑制非直达波误差,因此出现了一些学习型算法。文献[5]利用克利金(Kriging)数据内插法建立NLOS误差数据库进行定位。文献[6]提出利用神经网络学习方法抑制非直达波误差,取得了一定效果。但神经网络训练过程遵循经验风险最小化准则,存在过拟合、网络层次节点不易确定、效率取决于样本的复杂性等固有缺陷。文献[7]利用核函数等方法描述测试数据与训练数据之间的相似

性,利用相似程度对训练数据中移动站位置进行加权,最后得到移动站位置。

支持向量机(SVM)是近年来机器学习中最有影响的成果之一,可成功克服神经网络的上述缺陷。文献[8]提出利用最小二乘支持向量机(LS-SVM)抑制非直达波误差,该方法通过对一定数量位置估计样本的学习,得到定位系统无线传播信道的系统函数的估计,对未知TOA数据的移动台进行位置估计。

尽管支持向量机方法具有较好的推广能力,但由于SVM的最优超平面最终只由少部分样本构成,因此SVM对噪声点或野值非常敏感,容噪性较差。针对这种情况,模糊支持向量机方法(FSVM)在SVM的基础上,将模糊数学引入其中,对不同的样本采用不同的模糊隶属度,以致在构造目标函数时,不同的样本有不同的贡献,对含有噪声与野值的样本赋予较小的权值,在一定的程度上可消除噪声与野值的影响。

收稿日期: 2007-07-02; 修回日期: 2008-01-15

基金项目: 国家自然科学基金(60372022; 607721467)

作者简介: 万 群(1971-),男,教授,博士生导师,主要从事移动定位、阵列信号处理方面的研究。

本文提出一种基于支持向量数据域描述的模糊最小二乘支持向量机(FLS-SVM)定位算法,根据样本本偏离数据域的程度赋予不同的隶属度。该方法提高了最小二乘支持向量机的抗噪声能力,仿真结果表明该方法在非直达波环境下有较好的稳健性,提高了定位精度。

1 模糊最小二乘支持向量机定位算法

设 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_l, y_l)\} \in R^n$ 为训练样本集,其中 $i=1, 2, \dots, l$ 。引入非线性变换 $\varphi(x)$,将样本从输入空间 R^n 映射到高维特征空间 H 。输入空间的回归问题可归结为求解下面的二次规划:

$$\begin{cases} \min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^l e_i^2 \\ \text{s.t. } y_i = w^T \varphi(x_i) + b + e_i \end{cases} \quad (1)$$

式中 $i=1, 2, \dots, l$; w 为权值向量; e_i 为松弛变量; b 为偏差; γ 为惩罚因子。与传统的SVM相比,该方法中二次规划约束条件为等式,且损失函数为二次函数,故称最小二乘支持向量机(LS-SVM)算法。引入Lagrange系数 α_i ,定义如下的Lagrange函数:

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^l \alpha_i \{w^T \varphi(x_i) + b + e_i - y_i\} \quad (2)$$

根据Mercer条件,存在映射 φ 和核函数 $K(\cdot, \cdot)$,使得 $K(x_i, x_j) = (\varphi(x_i), \varphi(x_j))$ 。令 L 对 w 、 b 、 e 和 α 的偏导数等于零,并将得到的等式代入式(2),可以得到矩阵方程:

$$\begin{bmatrix} \mathbf{0} & \mathbf{1}_v^T \\ \mathbf{1}_v & \Omega + \frac{1}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ y \end{bmatrix} \quad (3)$$

式中 $y = [y_1, y_2, \dots, y_l]^T$; $\mathbf{1}_v = [1, 1, \dots, 1]^T$; $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]^T$; $\Omega_j = \varphi(x_i)^T \varphi(x_j)$ 。求解矩阵方程(3),得到最小二乘支持向量机的解为:

$$y(x) = \sum_{i=1}^l \alpha_i K(x, x_i) + b \quad (4)$$

本文采用RBF核对移动站位置进行估计:

$$K(x_k, x_l) = \exp\left(-\frac{|x_k - x_l|^2}{2\sigma^2}\right)$$

为解决SVM对孤立点过分敏感并因此而带来的过拟合问题,文献[9]将模糊隶属度概念引入SVM,模糊化输入样本集,提出了模糊支持向量机(FSVM)。将这一思想引入LS-SVM,为LS-SVM中的每个样本引入模糊隶属度 μ_i ,模糊化输入样本集

$(x_1, y_1, \mu_1), (x_2, y_2, \mu_2), \dots, (x_i, y_i, \mu_i), \dots, (x_l, y_l, \mu_l)$,其中 $0 < \mu_i < 1$ 。将式(1)中的目标函数重写为:

$$\min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \mu_i \sum_{i=1}^l e_i^2 \quad (5)$$

与LS-SVM方法一样,构造Lagrange函数,最后得到矩阵方程:

$$\begin{bmatrix} \mathbf{0} & \mathbf{1}_v^T \\ \mathbf{1}_v & \Omega + \frac{1}{\mu_i} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ y \end{bmatrix} \quad (6)$$

求解式(6),最后得到模糊最小二乘支持向量机的解。与式(3)相比,矩阵方程(6)中多了模糊隶属度 μ_i ,故该方法被称为模糊最小二乘支持向量机算法。

基于支持向量机的定位过程分为机器学习和定位两个阶段。机器学习阶段是为实现函数估计的风险函数最小化,该过程是将各个基站测得的测量参数(本文考虑TOA测量参数,也可以是TDOA、RSS、AOA参数)作为输入向量,将移动台位置作为输出向量。对LS-SVM结构进行训练,得到定位系统无线传播信道的系统函数的估计,用来对待定位移动台进行位置估计。

2 模糊隶属度的确定

在分类问题中,文献[9]提出一种根据样本集的先后顺序确定样本隶属度的模型,该模型认为,最近得到的样本相对要比其他的样本重要,其隶属度也大,但该模型缺乏理论上的依据。文献[10]提出一种基于孤立点检测的模糊隶属度模型,对于主体集中的样本,根据样本到其聚类中心的距离确定模糊隶属度;而对于孤立点集中的样本其模糊隶属度则赋予一个很小的正数。文献[10]并没有区分孤立点集中的样本,仅简单地为每个孤立点赋予相同的模糊隶属度。为了确定模糊隶属度函数的形式,需要衡量一个样本偏离其所在类总体的程度。本文采用支持向量数据域描述方法(support vector domain description, SVDD)。将数据样本映射到一个高维的空间,然后在这个高维空间中寻找其最小包含超球,并根据样本超球球心的距离确定其隶属度值。

2.1 支持向量数据域描述

支持向量数据域描述方法^[11]可描述为,给定训练样本集 $X = \{x_1, x_2, \dots, x_i, \dots, x_l\}$,其中 $x_i \in R^n$ 和 l 分别为输入空间和样本个数;为了建立样本的数据域描述模型,需要寻找样本的最小包含超球;当输入空间中的样本为非球形分布时,引入映射 $\Phi: R^n \rightarrow F$,将输入空间中的样本映射到一个高维

的特征空间 F , 然后求解下面的二次规划:

$$\begin{cases} \min W = R^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t. } \|\Phi(x_i) - a\|^2 \leq R^2 + \xi_i \end{cases} \quad (7)$$

式中 R 为最小包含超球的半径; a 为球心; $\xi_i \geq 0$ 为松弛变量; C 为惩罚因子。引入Lagrange系数, 得到Wolf对偶为:

$$\begin{cases} \max Q = \sum_{i=1}^l K(x_i, x_i) \beta_i - \sum_{i,j=1}^l K(x_i, x_j) \beta_i \beta_j \\ \text{s.t. } 0 \leq \beta_i \leq C \quad i = 1, 2, \dots, l \end{cases} \quad (8)$$

求解上面的二次规划, 即可得到最优的Lagrange系数及样本在特征空间中的数据描述。

2.2 数据域描述的模糊隶属度函数模型

输入空间中的点 x_i 在特征空间中的映射 $\Phi(x_i)$ 到最小包含超球球心 a 的距离定义为 $D^2(x_i) = \|\Phi(x_i) - a\|^2$, 考虑 $a = \sum_{i=1}^l \beta_i \Phi(x_i)$, 有:

$$D^2(x_i) = \sum_{i,j=1}^l K(x_i, x_j) \beta_i \beta_j + K(x_i, x_i) - 2 \sum_{j=1}^l K(x_j, x_i) \beta_j$$

式中 $i, j = 1, 2, \dots, l$; $0 < \beta_i < C$ 。

定义 $X_{\text{NBSV}} = \{x_1, x_2, \dots, x_k, \dots, x_m\}$ 为输入空间的子集, 其中 m 为非边界支持向量的个数。特征空间中最小包含超球半径满足 $R = D(x_i) | x_i \in X_{\text{NBSV}}$ 。当 R 和 a 确定, 就得到了给定数据集的数据域描述。

定义 $D_{\text{max}} = \max(D(x_i) | x_i \in X)$, $D_{\text{min}} = \min(D(x_i) | x_i \in X)$ 分别为样本到最小超球球心最大、最小距离。

定义模糊隶属度函数如下:

$$\mu_i = \begin{cases} \left(1 - \frac{D(x_i) - D_{\text{min}}}{D_{\text{max}} - D_{\text{min}}}\right)^f + \delta & R < D(x_i) \leq D_{\text{max}} \\ 1 - \frac{D(x_i) - D_{\text{min}}}{D_{\text{max}} - D_{\text{min}}} & D_{\text{min}} \leq D(x_i) \leq R \end{cases} \quad (10)$$

式中 $\delta < 1$ 为足够小的正实数; $f \geq 2$ 。本文选取 $f = 2$; $\delta = 0.01$ 。

对于输入空间中的点 x_i , 其在特征空间中的映射 $\Phi(x_i)$ 到最小包含超球球心 a 的距离满足 $D_{\text{min}} \leq D(x_i) \leq D_{\text{max}}$ 。当 $D_{\text{min}} \leq D(x_i) \leq R$ 时, 表示 x_i 满足数据域描述, $\Phi(x_i)$ 位于超球内或球面上, 其模糊隶属度随 $\Phi(x_i)$ 的增大而线性减少; 当 $R < D(x_i) \leq D_{\text{max}}$ 时, 样本 x_i 偏离数据域描述, 也就是偏离其所在的类总体, $\Phi(x_i)$ 位于超球之外, 其模

糊隶属度随着 $D(x_i)$ 的增大而减小。当 $D(x_i)$ 接近于 D_{max} 时, 其模糊隶属度已接近于一个非常小的实数 δ , 这样可以减少这些点的影响。

2 仿真研究

仿真环境参照文献[5]为一个城市的街道环境, 如图1所示。图中方块部分为建筑物, 空白处为街道。参与定位的有3个基站, 其位置分别为(0,0)、(460,0)、(460,460), 距离单位为米。利用3个基站测得的TOA测量值进行定位计算。移动台在街道上均匀分布。非直达波误差由Dijkstra algorithm^[5]模型产生。测量误差服从零均值, 标准差为 σ 的高斯分布。

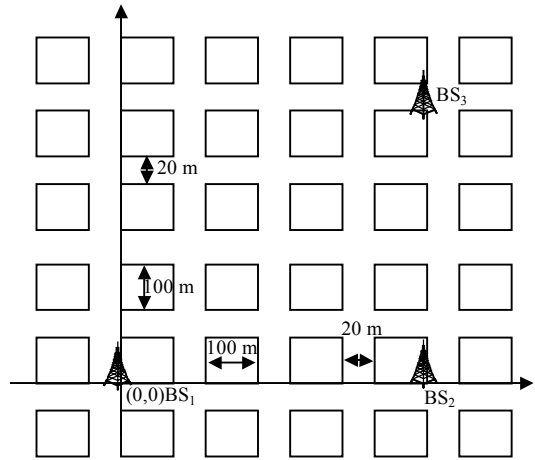


图1 城市街道环境

训练样本中的野值通常是由于多径效应和非直达波的影响, GPS无法得到训练样本的真实位置造成的。仿真中, 在移动台位置上加入均值为100、标准差为100的高斯噪声, 野值在样本集的比例为 P 。定义 $MLE = E(\sqrt{(MS\hat{x} - MSx)^2 + (MS\hat{y} - MSy)^2})$ 为平均定位误差, 其中 (MSx, MSy) 为移动站真实位置; $(MS\hat{x}, MS\hat{y})$ 为移动站估计位置。仿真中本文方法与LS-SVM方法^[8]和核函数方法^[7]做比较。

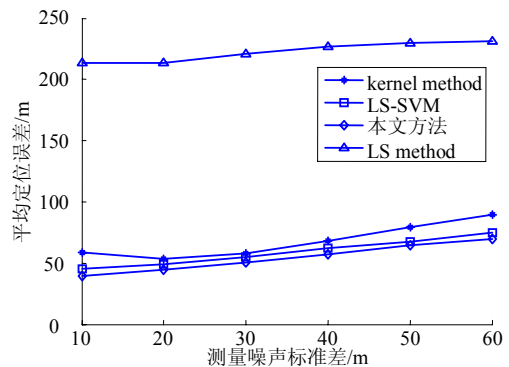


图2 平均定位误差和测量噪声标准差的关系曲线
图2为 $P = 20\%$ 、训练样本数为30时, 平均定位

误差与测量噪声标准差的关系曲线。由图可见,与LS-SVM方法^[8]和核函数方法^[7]相比,本文算法对定位精度有一定的改善。

图3总结了 $\sigma = 30\text{ m}$ 及 $P = 20\%$ 时所得到的仿真结果,主要反映了平均定位误差与训练样本数量的关系曲线。可以看出,平均定位误差随训练样本的数量增加迅速地减少,当样本训练数量达到一定数量时,样本的数量对定位误差影响不大。但具体需多少训练样本,具体需达到何种误差精度,可能还需由具体的实际应用环境决定。

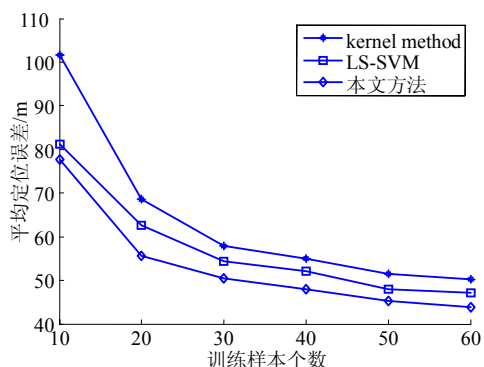


图3 平均定位误差和训练样本数量的关系曲线

图4为 $\sigma = 30\text{ m}$ 、训练样本数为30时,平均定位误差和野值在训练样本集的比例 P 的关系曲线。定位精度随野值点的增多而降低。本文方法能较好地抑制野值的影响。

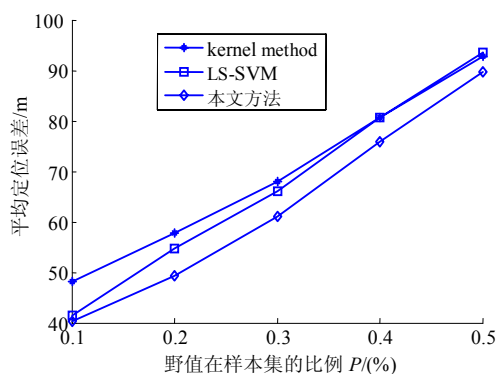


图4 平均定位误差和野值在训练样本集的比例 P 的关系曲线

4 结 论

本文针对SVM对噪声敏感,提出了一种基于支持向量数据域描述的模糊最小二乘支持向量机定位算法,根据样本偏离数据域的程度赋予不同的隶属度。仿真结果表明该方法在非直达波环境下有较好的稳健性,提高了定位精度,增强了LS-SVM的抗噪声能力。

参 考 文 献

- [1] CAFFERY J J. Wireless location in CDMA cellular radio systems[M]. Boston: Kluwer Academic Publishers, 1999.
- [2] WAN Q, YANG W L, PENG Y N. Simplified calculation of mobile location using scatter information[C]//2004 international conference on communications, circuits and systems. Chengdu, China: [s.n.], 2004, 1: 475-478.
- [3] 万 群, 王 军, 罗咏娟, 等. 抑制非直达波误差的高阶矩方法[J]. 电子学报, 2005, 33(10): 1822-1824.
WAN Qun, WANG Jun, LUO Yong-jie, et, al. High-order moment method-for mitigation of NLOS error[J]. Acta Electronica Sinica, 2005, 33(10): 1822-1824
- [4] CHAN Y T, TSUI W Y, SO H C, et al. Time-of-arrival based localization under NLOS conditions[J]. IEEE Transactions on Vehicular Technology, 2006, 55(1): 17-24.
- [5] LI B H, CHRIS R, LEE H K. Utilizing Kriging to generate a NLOS error correction map for network based mobile positioning[J]. Journal of Global Positioning Systems, 2005, 4(1): 27-35.
- [6] SALCIC Z. GSM mobile station location using reference stations and artificial neural networks[J]. Wireless Personal Communications, 2001, 19(3): 205-226.
- [7] MCGUIRE M, PLATANIOTIS K N, VENETSANOPOULOS A N. Location of mobile terminals using time measurements and survey points[J]. IEEE Trans on Vehicular Technology, 2003, 52(4): 999-1011.
- [8] SUN G L, GUO W. Robust mobile Geo-Location algorithm based on LS-SVM[J]. IEEE Transactions on Vehicular Technology, 2005, 54(3): 1037-1041.
- [9] LIN C F, WANG S D. Fuzzy support vector machines[J]. IEEE Trans on Neural Networks, 2002, 13(2): 464-471.
- [10] HUANG H P, LIU Y H. Fuzzy support vector machines for pattern recognition and data mining[J]. International Journal of Fuzzy System, 2002, 4(3): 826-835.
- [11] DAVID M J, ROBERT P W D. Data domain description using support vectors[C]//European Symposium on Artificial Neural Networks. Bruges: [s.n.], 1999: 251-256.

编 辑 熊思亮