

文章编号: 1672-2892(2012)02-0192-04

具有信任度输出的雷达信号 SVM 回归分析模型

尚 伟, 梁玉英, 朱 赛

(军械工程学院 光学与电子工程系, 河北 石家庄 050003)

摘 要: 针对支持向量机只有点输出, 没有概率输出的问题, 提出了一种加权平均的预测信任度函数, 通过该函数推导计算出每个预测值对应的信任度, 实现对预测值采用一个可信程度的度量。用雷达上常用的线性调频信号仿真验证得出, 该方法不但实现了计算预测信任度, 而且可为信号中噪声含量提供定性分析。

关键词: 支持向量机; 回归算法; 信任度; 预测; 概率输出

中图分类号: TN957.51; TP181

文献标识码: A

Regression analysis model for radar signal based on SVM with credibility evaluation

SHANG Wei, LIANG Yu-ying, ZHU Sai

(Optics and Electronic Engineering Department, Ordnance Engineering College, Shijiazhuang Hebei 050003, China)

Abstract: Support Vector Machine(SVM) has been widely applied to classification and regression problems. Because there is no probability output but lattice output only in SVM, this paper puts forward a kind of credibility prediction function with weighted average. Through this function, the degree of belief corresponding to every predicted value can be calculated, adopting a tolerance of credible level on the predicted value. Simulation results with the general Linear Frequency Modulation(LFM) Chirp signal on radar indicate that, this method not only realizes calculating the degree of belief for prediction, but also can offer qualitative analysis for the noise content in signal.

Key words: Support Vector Machine; regression; credibility; predicting; probability output

支持向量机(SVM)是针对小样本数据问题提出的, 借助于最优化方法解决机器学习问题的新工具。它最初于 20 世纪 90 年代由 Vapnik 提出, 近年来在其理论和算法实现方面都取得了突破性进展^[1]。支持向量机的最大特点是根据 Vapnik 结构风险最小化原则, 尽量提高学习机的泛化能力, 即由有限的训练集样本得到最小的误差, 能够保证对独立的测试集仍然保持小的误差; 另外, 由于支持向量机算法是一个凸优化问题, 因此局部最优解一定是全局最优解^[2]; 支持向量机的这些优点使它成为了克服“维数灾难”和“过学习”等传统困难的有力手段^[3]。支持向量机的一个弱点就是, 它的输出只有点输出, 没有概率输出。这使得人们在使用支持向量机做分类和预测时不能确定所得结果的可信度, 尤其是做预测时, 不能知道预测结果是否可信。在文献[4]中引出了预测信任度的概念, 但是没有考虑不同测试点对预测信任度的不同影响。本文结合雷达某信号, 用 SVM 对其进行回归和预测分析, 通过预测信任度函数计算出每个预测值的信任度, 从而能够指导使用者对预测值进行取舍。

1 支持向量回归算法

设给定的输入样本为 n 维向量, l 个样本及其输出值可表示为^[1]:

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (X \times Y)^l \quad (1)$$

式中: $X = \mathbf{R}^n$; $Y = \mathbf{R}$ 。采用 1 个非线性映射 ϕ 将数据映射到高维特征空间 F , 并在这个空间进行线性回归, 设回归函数为:

$$f(x) = \langle \mathbf{w}, \phi(x) \rangle + b \quad (2)$$

式中: $\mathbf{w}$ 是特征向量; b 是误差项。

收稿日期: 2011-05-25; 修回日期: 2011-06-22

基金项目: 国家自然科学基金资助项目(60472009)

在不敏感损失函数 ε 下, 采用 1 个参数 ξ 来控制误差项, 得到下面的优化问题:

$$\min F(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} C \sum_{i=1}^l \xi_i^2 \quad (3)$$

式中: C 为惩罚因子; ξ_i 为松弛变量。

约束为:

$$|f(x_i) - y_i| \leq \xi_i + \varepsilon, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (4)$$

为了得到式(3)的对偶形式, 引入拉格朗日函数 L :

$$L(\mathbf{w}, b, \xi, a, a^*, \gamma) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l a_i [\xi_i + \varepsilon - y_i + f(x_i)] - \sum_{i=1}^l a_i^* [\xi_i + \varepsilon - y_i + f(x_i)] - \sum_{i=1}^l \gamma_i \xi_i \quad (5)$$

式中: α 和 γ 为拉格朗日乘子; $a_i, a_i^*, \gamma_i \geq 0, \quad i = 1, 2, \dots, l$ 。

函数 L 的极值应当满足条件 $\frac{\partial}{\partial \mathbf{w}} L = 0, \quad \frac{\partial}{\partial b} L = 0, \quad \frac{\partial}{\partial \xi_i} L = 0$, 可以得到:

$$\mathbf{w} = \sum_{i=1}^l (a_i - a_i^*) \phi(x_i) \quad (6)$$

$$\sum_{i=1}^l (a_i - a_i^*) = 0 \quad (7)$$

$$C - a_i - a_i^* - \gamma_i = 0 \quad (8)$$

将上面 3 个公式代入拉格朗日函数 L , 求解 $\mathbf{w}$ 的最大值 W , 得到优化问题式(3)的对偶形式:

$$\max W(a, a^*) = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i - a_i^*)(a_j - a_j^*) K(x_i, x_j) + \sum_{i=1}^l (a_i - a_i^*) y_i - \sum_{i=1}^l (a_i + a_i^*) \varepsilon \quad (9)$$

式中 $K(x_i, x_j)$ 是核函数。约束为:

$$\sum_{i=1}^l (a_i - a_i^*) = 0, \quad a_i + a_i^* \leq C \quad (10)$$

解出 $f(x)$ 的表达式^[7]:

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x, x_i) + b \quad (11)$$

$$b = \frac{1}{2} \left\{ \min \left[y_i - \sum_{i=1}^l (a_i - a_i^*) K(x, x_i) \right] + \max \left[y_i - \sum_{i=1}^l (a_i - a_i^*) K(x, x_i) \right] \right\} \quad (12)$$

解出 $f(x)$ 的表达式后, 则可以进行回归预测分析。

2 信任度函数

通过上面的模型解出 $f(x)$ 的表达式进行预测, 得到的预测值是否可信并不能确定。根据文献[1]引出了预测信任度的概念, 用局部预测的思想, 对每个待预测样本都找到它的一个邻近集, 然后以这个邻近集作为训练集, 并且根据这个邻近集来定义该待预测样本的预测信任度^[5-6]。文献[5]中的方法理解为第 x_{M+1} 个预测样本对应的预测值 y_{M+1} 的可信度 $PB(y_{M+1})$, 是根据 x_{M+1} 点前 l 个点对应的回归准确度的平均值。但是考虑到距离 x_{M+1} 点远近不同, 对 y_{M+1} 预测准确度影响不同, 所以对预测信任度的影响不同。为此, 提出了一种加权平均的预测信任度函数, 本文中采用的权重是 $\frac{i}{N}$, $N = \sum_{i=1}^l i$, 点 i 的权重 $\frac{i}{N}$ 与它到 x_{M+1} 点的距离成反比。

设总的训练集为 $\{x_i, y_i\}, i = 1, 2, \dots, M$, M 代表训练数据的个数, $x_i \in \mathbf{R}^M$ 表示输入数据, $y_i \in \mathbf{R}$ 表示输出数据。首先, 给定一个适当的正数 a , 选择到 x_{M+1} 的距离小于等于 a 的样本作为训练集(设数量为 l), 然后送入支持向量回归机模型中训练, 得到 $f(x)$; 然后, 给出能够接受的误差 $e > 0$, 本文的预测信任度函数定义为:

$$PB(y_{M+1}) = \frac{\sum_{i=1}^l i [y_i - f(x_i)]}{N} \quad (13)$$

$$\text{式中 } I(x) = \begin{cases} 1, & x \leq e \\ 0, & x > e \end{cases}.$$

该定义可以理解为, x_{M+1} 点的预测值的可信度等于 x_{M+1} 点前 l 个点对应的回归准确度的加权平均值。

3 实例与结果分析

线性调频信号(LFM)是雷达中常用的信号, T 为周期, 定义为^[7-9]:

$$s(t) = e^{j\pi\omega t^2/T}, \quad -\frac{1}{2}T \leq t \leq \frac{1}{2}T \quad (14)$$

为了在 MATLAB 中表示线性调频信号, 必须对 $s(t)$ 进行采样, 采样频率一般取 $f_s = p\omega$, $p \geq 1$ 为过采样因子, ω 为扫频带宽, 离散时间的线性调频信号可表示:

$$s(n) = \exp\left[j2\pi\alpha\left(n - \frac{1}{2}N\right)^2\right], \quad 0 \leq n \leq N \quad (15)$$

式中: $N = p\omega T$; $\alpha = \frac{1}{2p^2\omega T}$, 看到它仅与 p 和 ωT 有关。据此编写出线性调频信号的函数程序(chirp.m)。产生的时域信号见图 1, 2 条曲线分别为没有加噪声和加过噪声的线性调频曲线。本文将对加过噪声的曲线进行回归与预测分析, 并计算预测信任度。参数 p 取值为 10, ωT 取值为 50, 幅度值为单位 1, 产生 500 个数据点, 取正半轴的数据为训练所用数据, 其中后 50 个数据用于预测验证。图 2 为正半轴 250 个数据曲线。图 1 中, 点状线表示的是没有噪声的线性调频信号, 连续曲线表示的是加过噪声后的曲线。

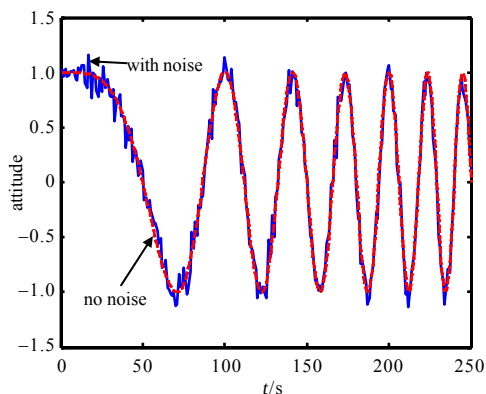


Fig.1 Chirp signal curve of time domain
图 1 线性调频信号时域曲线

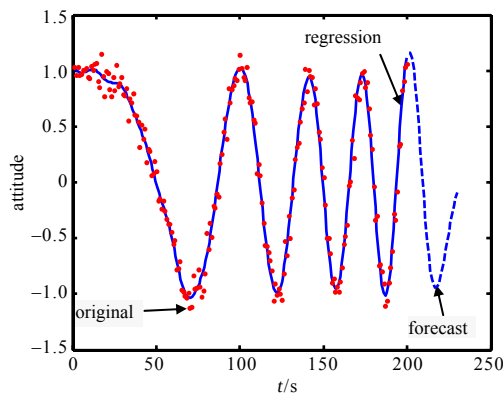


Fig.2 Regression curve based on SVM
图 2 用 SVM 模型拟合曲线

用支持向量机模型对混有噪声的线性调频信号进行训练和预测, e 取 0.1, 对其前 200 个数据用作回归训练, 后 50 个数据用于预测验证, 得到回归曲线和预测值见图 2, 图中“:”曲线表示为原始数值, “-”为回归曲线, “--”为预测曲线。SVM 模型较好完成了回归, 并做出了预测和预测信任度计算, 表 1 中给出了前 10 个预测值及其信任度的计算结果。

从表 1 可看出, 每 1 个预测值都对应有一个信任度值, 在一定程度上可避免盲目性。信任度的逐渐降低是因为采用某一点前面的回归精度来推算该点的预测信任度, 误差传递导致预测精度下降, 预测信任度逐渐降低。

同样用支持向量机模型对没有混有噪声的线性调频信号进行训练和预测, e 取 0.1, 对其前 200 个数据用作回归训练, 后 50 个数据用于预测验证。得到的预测值和预测可信度见表 2。

从表 2 可看出, 每 1 个预测值也都对应 1 个信任度, 但都是 1。这是因为所用的训练数据中没有加入任何噪声, 就是标准的线性调频信号, 所以训练得出的回归精度特别高, 基本没有什么误差, 这使得模型预测比较准确。

表 1 预测与信任度计算结果(有噪声)
Table1 Regress and credibility(with noise)

real	regression	credibility
0.904	1.161	0.686
1.034	1.111	0.644
1.060	1.014	0.603
0.839	0.875	0.603
0.873	0.702	0.603
0.652	0.503	0.490
0.611	0.292	0.455
0.152	0.074	0.450
-0.015	-0.138	0.417
-0.001	-0.095	0.385

表 2 预测与信任度计算结果(无噪声)
Table2 Regress and credibility(no noise)

real	regression	credibility
1.000	1.001	1
0.968	0.967	1
0.875	0.875	1
0.725	0.725	1
0.527	0.526	1
0.294	0.294	1
0.040	0.041	1
-0.218	-0.219	1
-0.462	-0.463	1
-0.676	-0.678	1

由上面2个例子看出,该预测信任度可以在一定程度上反映信号中含有噪声的情况,可以用作一般性要求不高的定性分析。

4 结论

分析了支持向量回归算法的理论基础,针对SVM没有概率输出的缺点,定义了1个预测可信度函数,使得预测值有了1个可信的量度,并且能够在一定程度上分析数据中噪声的含量。通过线性调频信号的仿真分析,验证了可信度函数正确可行。但是信任度函数的精确程度需要进一步验证并不断提高,能够接受的误差 e 与支持向量机参数 ε 的相互作用关系以及对信任度的影响需要进一步研究。

参考文献:

- [1] 吴今培,孙德山. 现代数据分析[M]. 北京:机械工业出版社, 2006.
- [2] 崔和,龙玉峰. 支持向量机器学习算法的研究现状与展望[J]. 信息与电子工程, 2008,6(5):328-332. (CUI He, LONG Yufeng. Status Quo and Expectation of Support Vector Machine Learning Algorithm[J]. Information and Electronic Engineering, 2008,6(5):328-332.)
- [3] Vapnik V N. Statistical Learning Theory[M]. New York:Wiley, 1998.
- [4] 张浩然,韩正之,李昌刚. 支持向量机[J]. 计算机科学, 2002,29(12):135-137. (ZHANG Haoran, HAN Zhengzhi, LI Changgang. Support Vector Machine[J]. Computer Science, 2002,29(12):135-137.)
- [5] 孙德山,吴今培. 支持向量回归中的预测信任度[J]. 计算机科学, 2003,30(8):126-127. (SUN Deshan, WU Jinpei. Predicting Credibility Based on Support Vector Machine[J]. Computer Science, 2003,30(8):126-127.)
- [6] 李占闯,肖国强,代毅,等. 基于SVM输出概率和后置滤波的运动目标分类[J]. 计算机应用研究, 2010,27(2):778-780. (LI Zhanchuang, XIAO Guoqiang, DAI Yi, et al. Moving target classification using SVM Probability and post-filtering[J]. Application Research of Computers, 2010,27(2):778-780.)
- [7] 楼顺天,刘小东,李博菡. 基于MATLAB 7.x的系统分析与设计-信号处理[M]. 西安:西安电子科技大学出版社, 2005.
- [8] Turk M, Pentland A. Eigenfaces for Recognition[J]. Journal of Cognitive Neuroscience, 1991,3(1):71-86.
- [9] 李超峰. 基于支持向量机的非线性系统广义预测控制研究与应用[D]. 杭州:浙江大学, 2010. (LI Chaofeng. The research and application of Nonlinear GPC based on SVM[D]. Hangzhou:ZHEJIANG UNIVERSITY, 2010.)

作者简介:



尚 伟(1986-),男,陕西省佳县人,在读硕士研究生,研究方向为武器系统性能检测与故障诊断. email:noise51@126.com.

梁玉英(1968-),女,石家庄市人,副教授,研究方向为武器系统性能检测与故障诊断.

朱 赛(1987-),男,河南省周口市人,在读博士研究生,研究方向为武器系统性能检测与故障诊断.