

基于模糊 sigmoid 核的支持向量机回归建模

刘 涵, 刘 丁

(西安理工大学 自动化与信息工程学院, 陕西 西安 710048)

摘要: 支持向量机中对核函数的要求为对称的半正定矩阵. 来自于神经网络的 sigmoid 核函数在其参数满足一定条件时才成为半正定矩阵, 但是这种核函数在 SVM 中却有很多成功的应用. 本文将 sigmoid 核函数与模糊逻辑相结合并使其模糊化, 从而简化了 SVM 的计算并便于用硬件实现. 通过对混沌时间序列预测以及图像去噪滤波器两个实例的实验研究发现, 使用模糊 sigmoid 核函数可以使 SVM 回归建模在损失较小精度的代价下, 较大地降低平均 CPU 执行时间, 便于硬件实现.

关键词: 支持向量回归; sigmoid 核函数; 模糊逻辑; 混沌时间序列预测; 图像滤波器

中图分类号: TP181

文献标识码: A

Support vector regression based on fuzzy sigmoid kernel

LIU Han, LIU Ding

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an Shaanxi 710048, China)

Abstract: In the support vector machines (SVM) framework, kernel function must meet certain requirements to be symmetric and positive semi-definite (PSD) matrix. Although sigmoid function derived from neural network can become a PSD kernel for proper combinations of its free parameters, it has been used in several practical and successful cases. The sigmoid kernel is combined with fuzzy logic methodology first, which makes the computation of SVM simple and of ease implementation in hardware. Experiments for chaotic time series prediction and image filter are also carried out to show that the average CPU time can be decreased markedly in favor of hardware implementation, in spite of small decrease in prediction precision.

Keywords: support vector regression; sigmoid kernels function; fuzzy logic; chaotic time series prediction; image filter

1 引言(Introduction)

统计学习理论 (statistical learning theory, SLT) 是由 Vapnik 建立的一种专门研究小样本下机器学习规律的理论, 支持向量机 (support vector machines, SVM) 是在这一理论基础之上发展起来的一种新的分类和回归工具^[1,2], 目前已经在模式识别、系统控制领域得到广泛的应用^[3,4]. 支持向量机通过结构风险最小化原理来提高泛化能力, 将求解最优分类面问题转化为求解凸二次规划问题, 较好地解决了小样本、非线性、高维数、局部极小点等实际问题. 在 SVM 问题的求解中, 巧妙地使用对称、半正定的核函数将输入样本映射到高维特征空间, 从而使线性不可分的问题转化为线性可分问题. 核函数在数学上必须满足 Mercer 条件^[5], 常用的核函

数有线性、多项式、高斯和 sigmoid 函数. 前三种核函数已经被广泛应用, 而来自于神经网络的 sigmoid 函数在实际中的应用受到限制, 主要是因为只有满足某些特定条件 (斜坡项 α 和偏置项 β 满足一定的条件), sigmoid 函数才满足对称、半正定的核函数条件. 由于在神经网络的应用中, sigmoid 函数已经被证明具有良好的全局分类性能, 因此其在 SVM 中的应用及特点已经成为非常有意义的研究内容^[6].

由于 sigmoid 函数为非线性函数, 用硬件实现比较困难并且耗时较长. 为解决这个问题, 一种方法是将 sigmoid 函数分段线性化, 但是会出现在分段点处不可微分的问题; 第二种方法是利用泰勒展开提取其线性项 (第一项), 但这种方法将损失核函数的非线性特性. 本文提出一种基于模糊逻辑的

sigmoid 核函数,将模糊逻辑应用于核函数,简化了计算且容易被硬件实现.通过对混沌时间序列预测以及图像去噪滤波器的实验研究,证明使用模糊 sigmoid 核函数和标准的 sigmoid 核函数相比,计算速度更快,并且计算精度没有明显的损失.

2 支持向量机回归的原理 (Principle of support vector regression)

给定训练集 $\{x_k, y_k; k = 1, 2, \dots, N\}$, 其中输入样本为 $x_k \in \mathbb{R}^n$, 输出样本为 $y_k \in \mathbb{R}$, 选择非线性变换 $\Phi(\cdot)$, 将上述训练集从原空间映射到高维特征空间 F , 并在此空间构造最优线性回归函数

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

这样在高维空间的线性回归就对应于低维输入空间的非线性回归.

在 Vapnik ε 不敏感损失函数的定义下^[1],

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & |y - f(x)| \leq \varepsilon, \\ |y - f(x)| - \varepsilon, & \text{其他}. \end{cases} \quad (2)$$

式(1)的求解转化为下式的最优问题求解:

$$\min_{\omega, b, \xi, \xi^*} J(\omega, \xi, \xi^*) = \frac{1}{2} \omega^T \omega + C \sum_{k=1}^N (\xi_k + \xi_k^*). \quad (3)$$

约束:

$$\begin{cases} y_k - \omega^T \Phi(x_k) - b \leq \varepsilon + \xi_k, \\ \omega^T \Phi(x_k) + b - y_k \leq \varepsilon + \xi_k^*, \\ \xi_k, \xi_k^* \geq 0. \end{cases} \quad (4)$$

其中: ε 为需要近似的精度, ξ_k, ξ_k^* 为处理函数 f 在 ε 精度下不能估计的数据所引入的松弛变量, $C > 0$ 为惩罚参数. 式(3)的求解可以转化为 Lagrange 对偶问题求解:

$$\begin{aligned} \max_{\alpha, \alpha^*} L(\alpha, \alpha^*) = & -\frac{1}{2} \sum_{k,l=1}^N (\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*) K(x_k, x_l) - \\ & \varepsilon \sum_{k=1}^N (\alpha_k + \alpha_k^*) + \sum_{k=1}^N y_k (\alpha_k - \alpha_k^*). \end{aligned} \quad (5)$$

约束:

$$\begin{cases} \sum_{k=1}^N (\alpha_k - \alpha_k^*) = 0, \\ \alpha_k, \alpha_k^* \in [0, C]. \end{cases} \quad (6)$$

其中 α, α^* 定义为非负的 Lagrange 乘子. 核函数 $K(x_k, x_l)$ 定义为

$$K(x_k, x_l) = \Phi(x_k)^T \cdot \Phi(x_l), \quad (7)$$

因此非线性的 SVM 求解不用去考虑映射 Φ 的具体

表达形式,只需要选择合适的核函数. 最后 SVM 对非线性函数回归的结果为

$$f(x) = \sum_{k=1}^N (\alpha_k - \alpha_k^*) K(x, x_k) + b. \quad (8)$$

3 基于模糊逻辑的 sigmoid 核函数 (Sigmoid kernel based on fuzzy logic)

目前在 SVM 中常用的核函数有线性核函数、多项式核函数、高斯核函数以及 sigmoid 核函数, 它们的表达式分别如下:

$$K(x_i, x_j) = x_i \cdot x_j, \quad (9)$$

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d, \quad (10)$$

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2), \quad (11)$$

$$K(x_i, x_j) = \tanh(\alpha \cdot x_i \cdot x_j + \beta). \quad (12)$$

只有当核函数为对称、半正定的矩阵时才满足 Mercer 条件,但在实际应用时,有一些非半正定的核也被应用在式(5)中. 最常见的一个例子就是上式所列的 sigmoid 核. Vapnik 最早在 1995 年就指出 sigmoid 函数当参数 α 和 β 取特定值时其核矩阵为非半正定矩阵 (Non-PSD)^[1]. 当核函数为非半正定阵时,式(3)和(5)的对偶关系就不存在,并且式(7)不成立. 然而, sigmoid 核却被成功用在许多实际应用中, Scholkopf 在文献[7]中给出了部分解释. 文献[6]深入研究了非半正定核函数在 SVM 中的应用以及相应的学习算法,证明了 sigmoid 核为有条件的半正定矩阵,并给出了一些有意义的结论.

为了易于完成硬件实现,文献[8]将 sigmoid 函数与模糊逻辑结合起来应用在神经网络学习中,得到了满意的结果. 受到上述思想的启发,本文将 sigmoid 核函数模糊化,应用于支持向量机的回归建模中,期望得到满意的结果.

基于模糊逻辑的 sigmoid 核函数就是将式(12)用模糊语言的形式描述出来^[8],即

$$K(x_i, x_j) = \begin{cases} -1, & x_i \cdot x_j \text{ 为 low,} \\ +1, & x_i \cdot x_j \text{ 为 high,} \\ a \cdot x_i \cdot x_j, & x_i \cdot x_j \text{ 为 medium.} \end{cases} \quad (13)$$

其中 a 为常数,表示了 sigmoid 函数的光滑程度,如图 1 所示.

根据上式, sigmoid 核可以被定义为一组隶属度函数. 为简单起见,采用三角形的隶属度函数,如图 2 所示. 模糊分区分别由 low, medium, high (低, 中, 高) 组成,当然也可以由 very low, low, medium, high, very high (很低, 低, 中, 高, 很高) 构成,但是这样会增加计算的负担. 由于 sigmoid 核函数必须是连续的,因

此计算三角形函数的上、下界的值分别为 $(-\beta \pm 1)/\alpha$. 最终式(13)可以写为如下 α 和 β 的函数:

$$K(x_i, x_j) = \begin{cases} -1, & x_i \cdot x_j \leq (-\beta + 1)/\alpha, \\ +1, & x_i \cdot x_j \geq (-\beta - 1)/\alpha, \\ M(x_i, x_j, \alpha, \beta), & \text{其他.} \end{cases} \quad (14)$$

其中

$$M(x_i, x_j, \alpha, \beta) = 2(x_i \cdot x_j + \beta/\alpha) - \alpha^2(x_i \cdot x_j + \beta/\alpha) | x_i \cdot x_j + \beta/\alpha |, \quad (15)$$

即 sigmoid 核函数可以由式(16)来描述:

$$K(x_i, x_j) = (-1)\mu_1(x_i \cdot x_j) + \alpha x_i \cdot x_j \mu_2(x_i \cdot x_j) + (+1)\mu_3(x_i \cdot x_j). \quad (16)$$

模糊化后的 sigmoid 核函数有以下优点:

- 1) 函数在整个定义域内每一点都是可微的;
- 2) 图2所定义的隶属度函数可以由微处理器、数字信号处理器等硬件简单地实现;
- 3) 通过这种模糊化的方法, 选择不同数量、不同复杂程度的隶属度函数就可以近似不同非线性程度的核函数.

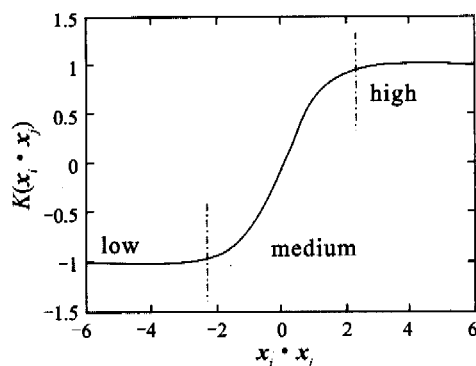


图1 Sigmoid核函数的模糊语言表达示意图

Fig. 1 Sigmoid kernel used fuzzy linguistic

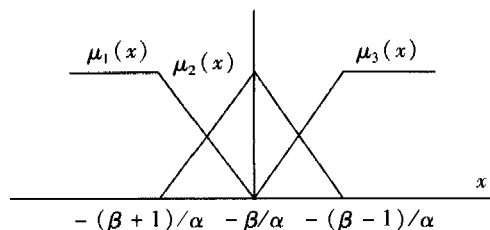


图2 模糊 sigmoid 核函数的隶属度函数

Fig. 2 Schematic of 3 membership functions

4 实验验证 (Experiments)

为了验证提出的模糊 sigmoid 核函数的方法, 分别用 SVM 对混沌时间序列预测以及图像去噪滤波器两个问题进行了研究. 在实验中, SVM 的参数 C 以及核函数的参数 α 和 β 通过 10-fold 交叉检验的方法进行整定, 对 SVM 的训练采用 IRWLS (Iterated

Reweighted Least Squares) 算法^[9], 该算法比标准的二次规划算法在执行时间和内存空间的使用上更有效, 实验在 Pentium IV 2.4GHz 512MB 内存的计算机上进行.

4.1 Mackey-Glass 时间序列预测 (Mackey-Glass chaotic time series prediction)

Mackey-Glass 时间序列的动力学方程定义为

$$\dot{x} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t). \quad (17)$$

该方程在初始条件 $x(0) = 1.2$ 和 $\tau = 17$ 时进入混沌状态. 混沌时间序列预测就是用时间序列 t 时刻前的值预测未来 $t+p$ 内点的值. 预测的方法是建立一个从间隔为 Δ 的时间序列的 D 个点, 即

$$[x(t-(D-1)\Delta), \dots, x(t-\Delta), x(t)]$$

到被预测的未来值 $x(t+p)$ 间的映射, 取 $D=4, \Delta=p=6$. 为了获得每一积分时间点的时间序列值, 用四阶龙格-库塔法求方程(17)的数值解, 得到 $x(t)$. 从 $x(t)$ 中抽取 1000 对输入输出数据:

$$[x(t-18), x(t-12), x(t-6), x(t), x(t+6)].$$

(18)

其中 $t=118 \sim 1117$, 前 500 个数据用于 SVM 训练, 后 500 个数据用于测试.

Mackey-Glass 时间序列预测的结果及预测误差分别如图3和图4所示. 为了比较使用模糊 sigmoid 核函数对 SVM 性能的影响, 对使用标准 sigmoid 核函数和模糊 sigmoid 核函数的混沌时间序列预测的两个指标进行了测试: 均方根误差和平均的 CPU 执行时间 (预测时间). 均方根误差 (RMSE) 反映了 SVM 回归建模的精度, 平均的 CPU 执行时间反映了回归建模的执行效率. 表1为相应的数据对比, 可看出, 虽然由核函数的模糊化带来了建模精度的下降 (约 6.3%), 但是平均的 CPU 的执行时间却有明显的提高 (约 35.3%).

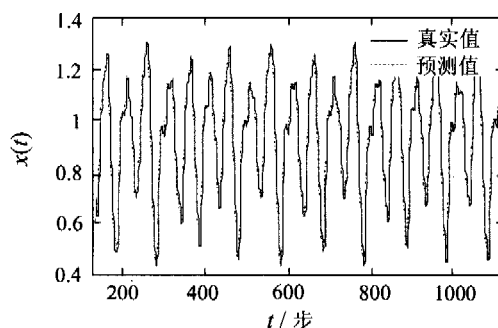


图3 $t=124 \sim 1123$ 时的 Mackey-Glass 时间序列预测结果

Fig. 3 Mackey-Glass time series prediction result

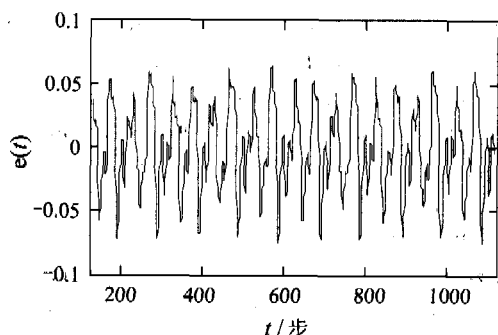


图4 混沌时间序列的预测误差

Fig. 4 Prediction error of chaotic time series

表1 混沌时间序列预测的性能指标比较

Table 1 Characteristic index of chaotic time series prediction

	标准 sigmoid 函数	模糊 sigmoid 函数
RMSE	0.0160	0.0169
CPU 时间/s	1.7	1.1

4.2 图像去噪滤波器 (Denoised image filter)

作者在文献[10]中提出了一种遗传-模糊神经网络图像去噪滤波器,对图像建立模糊神经网络去噪模型,并通过遗传算法来训练网络,获得了很好的去噪效果.文中所用的方法也可以通过 SVM 回归建模的方法来实现.

假设要处理的是具有 L 级灰度的噪声图像,设 $x(i,j)$ 为像素点 (i,j) 处的亮度值,而 $x_1, x_2, \dots, x_8$ 分别为其相邻 8 个像素点的亮度值,如图 5 所示.该方法的基本思想是利用 $x(i,j)$ 周围的 8 个像素点的亮度值去估计 $x(i,j)$ 的真实值,即完成对像素点 $x(i,j)$ 的滤波.完成这样的非线性映射可以通过 SVM 来实现,输入向量由 $x_1, x_2, \dots, x_8$ 构成,输出为真实无污染的样本 $x(i,j)$. 实验中采用标准的 256×256 Cameraman 图像作为训练图像,用标准的 256×256 Lenna 图像作为测试图像.使用标准 sigmoid 核函数和模糊 sigmoid 核函数分别对该问题进行了研究,图 6 和图 7 分别为冲击噪声概率为 0.3 时 256×256 的 Lenna 图像以及经过 SVM 回归去噪后的 Lenna 图像.为了评价不同方法去噪的效果,仍然使用均方根误差(RMSE)作为测试指标,同时用平均 CPU 执行时间作为算法效率的指标,表 2 为用 3 种方法所作实验相应数据指标的对比.

表 2 的数据表明采用模糊 sigmoid 函数回归建模虽然精度有少量的降低(约 4.7%),但是平均 CPU 执行时间有明显的提高(约 31.0%).

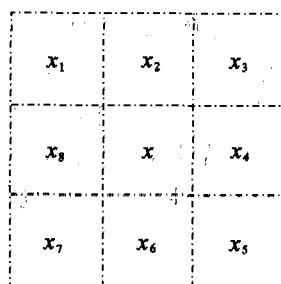


图5 3*3 相邻像素点的示意图

Fig. 5 Pixels belonging to 3×3 neighborhood

图6 256×256 冲击噪声概率为 0.3 的 Lenna 图像

Fig. 6 256×256 Lenna image with impulse noise



图7 SVM 回归去噪后的 Lenna 图像

Fig. 7 Lenna image denoised by SVM

表2 图像去噪滤波器的性能指标比较

Table 2 Characteristic index of image filter

		RMSE	CPU 时间/s
SVM	标准 sigmoid 函数	57.1	18.7
	模糊 sigmoid 函数	59.8	12.9
	FNN 滤波器[10]	69.2	67.0

从以上两个实验可以看出,模糊化的 sigmoid 核函数提供了一种简单的核函数的实现方法,通过对 SVM 回归建模的研究,这种简化虽然比标准的 sigmoid 核函数精度有稍许的降低,但是在平均的 CPU 执行时间上却有明显的提高,为硬件实现 SVM 计算提供了基础.

5 结论 (Conclusion)

本文提出了使用模糊化的 sigmoid 函数作为核函数进行 SVM 回归建模的方法.通过对混沌时间

序列预测以及图像去噪滤波器两个问题的实验结果看,这种简化的核函数在损失较小回归精度的代价下,提供了较快的执行时间,并且便于用硬件实现.在此基础上,对其它的核函数,比如 RBF 核函数的模糊化,对 SVM 回归性能的影响也将成为有意义的研究内容.

参考文献(References):

- [1] VAPNIK V. *The Nature of Statistical Learning Theory* [M]. New York: Springer-Verlag, 1995.
- [2] VAPNIK V. An overview of statistical learning theory[J]. *IEEE Trans on Neural Network*, 1999, 10(5): 988 - 999.
- [3] 王定成,方廷健.一种基于支持向量机的内模控制方法[J]. *控制理论与应用*, 2004, 21(1): 85 - 88.
(WANG Dingcheng, FANG Tingjian. Internal model control approach based on support vector machines[J]. *Control Theory & Applications*, 2004, 21(1): 85 - 88.)
- [4] 刘涵,刘丁,郑岗,等.基于最小二乘支持向量机的天然气负荷预测[J]. *化工学报*, 2004, 55(5): 828 - 832.
(LIU Han, LIU Ding, ZHENG Gang, et al. Natural gas load forecasting based on least squares support vector machine[J]. *J of Chemical Industry and Engineering*, 2004, 55(5): 828 - 832.)
- [5] COURANT R, HILBERT D. *Methods of Mathematical Physics* [M]// Interscience Publications, New York: Wiley, 1953.
- [6] LIN H T, LIN C J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods [EB/OL]. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, 2004, available at <http://www.csie.ntu.edu.tw/~cjlin/papers/tanh.pdf>.
- [7] SCHOLKOPF B, BURGE C, SMOLA A. *Advances in Kernel Methods: Support Vector Machines* [M]. Cambridge, MA: MIT Press, 1998.
- [8] SOIRA E, MARTIN J, CAMPS G, et al. A low complexity fuzzy activation function for artificial neural network [J]. *IEEE Trans on Neural Networks*, 2003, 14(6): 1576 - 1579.
- [9] PÉREZ-CRUZ F, ARTÉS-RODRÍGUEZ A. A new optimizing procedure for v-support vector regressor [C] // *Proc of Int Conf on Acoustics, Speech and Signal Processing, ICASSP 01*. Salt Lake City, UT: IEEE Press, 2001, 2: 1265 - 1268.
- [10] 刘涵,刘丁,李琦.一种遗传-模糊神经网络滤波器[J]. *仪器仪表学报*, 2004, 25(3): 310 - 312.
(LIU Han, LIU Ding, LI Qi. A GA fuzzy neural network image filter[J]. *Chinese J of Scientific Instrument*, 2004, 25(3): 310 - 312.)

作者简介:

刘涵 (1972—),男,工学博士,副教授,研究方向为复杂系统建模与控制、机器学习、智能信息处理等, E-mail: liuhan@xaut.edu.cn;

刘丁 (1957—),男,教授,博士生导师,现任西安理工大学校长,研究方向为智能控制、复杂系统建模等, E-mail: liud@xaut.edu.cn.