1000-9825/2002/13(06)1169-04 ©2002 Journal of Software 软 件 学 报 Vol.13, No.6

回归型支持向量机的简化算法*

田盛丰, 黄厚宽

(北方交通大学 计算机与信息技术学院,北京 100044)

E-mail: sftian@center.njtu.edu.cn

http://www.njtu.edu.cn

摘要: 针对支持向量机应用于函数估计时支持向量过多所引起的计算复杂性,提出一种简化算法,可以大幅度地减少支持向量的数量,从而简化其应用.采用简化算法还可以将最小平方支持向量机算法和串行最小化算法结合起来,达到学习效率高且生成的支持向量少的效果.

关 键 词: 支持向量机;回归;机器学习;计算复杂性;算法

中图法分类号: TP181 文献标识码: A

自从 Vapnik[1]等人引入支持向量机(SVM)理论以来,支持向量机在模式识别和函数估计方面取得了越来越多的应用.支持向量机是在统计学习理论的基础上形成的,力图实现结构风险的最小化,从而提高了学习机的泛化能力.与人工神经网络的方法相比,由于支持向量机没有局部最优问题,并且提高了泛化能力,因此有较大的优越性.但是,人工神经网络的复杂度是由神经元的个数决定的,而支持向量机的复杂度是由具有非零权值的样本即支持向量的个数决定的,对于大规模样本数据的情况而言,支持向量机的计算量是比较大的,因此限制了它的应用.

为了简化支持向量机,降低其计算复杂度,已有了一些研究成果.比如,Burges 提出根据给定的支持向量机生成缩减的样本集,从而在给定的精度下简化支持向量机[2],但生成缩减样本集的过程也是一个优化过程,计算比较复杂;Schölkopf 等人在目标函数中增加了参数 ν 以控制支持向量的数目,称为ν-SVR[3],证明了参数 ν 与支持向量数目及误差之间的关系,但支持向量数目的减少是以增大误差为代价的.

本文提出的简化回归型支持向量机的算法克服了上述方法的不足,在不需要其他优化算法的情况下,可以在容许的误差范围内大幅度地减少支持向量的数目,从而减小计算的复杂度.下面各节分别给出了简化算法、大规模数据情况下的高效学习算法及实验结果.

1 回归型支持向量机及其简化算法

设给定的输入样本 $\boldsymbol{x}$ 为 n 维向量,k 个样本及其输出值可表示为

$$(\boldsymbol{x}_1,y_1),\ldots,(\boldsymbol{x}_k,y_k)\in R^n\times R, \tag{1}$$

则回归型支持向量机的学习问题就是一个二次规划问题,通常采用 Vapnik 的ε-不敏感损失函数,即指定容许误差 ε,若样本 $\boldsymbol{x}$ 误差为 ξ,则当 $|\xi|\le\varepsilon$ 时不计损失,否则损失计为 $|\xi|-\varepsilon$.回归函数可表示为

$$f(\boldsymbol{x})=\sum_{i=1}^{k}(\alpha_i-\alpha_i^*)K(\boldsymbol{x},\boldsymbol{x}_i)+b\ , \tag{2}$$

其中 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\alpha}^*$ 为求解的 k 维向量,$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\varphi(\boldsymbol{x}_i)\cdot\varphi(\boldsymbol{x}_j)$ 称为核函数,$\varphi(\boldsymbol{x})$ 为从样本空间到高维特征空间的映射函数,核

* 收稿日期: 2000-08-23; 修改日期: 2001-04-03

基金项目: 国家铁道部科技研究开发项目(2000X030-A)

作者简介: 田盛丰(1944-),男,北京人,教授,主要研究领域为模式识别,人工智能;黄厚宽(1940-),男,四川遂宁人,教授,博士生导师,主要研究领域为人工智能,知识工程,KDD.

函数表示为两个$\varphi(\boldsymbol{x})$的点积,核函数的采用使得映射函数$\varphi(\boldsymbol{x})$不必明确求出,因此,求解非线性回归成为可能.式(2)中对应于权值$(\alpha_i-\alpha_i^*)$不为0的样本$\boldsymbol{x}_i$称为支持向量.显然,支持向量的数目决定了计算的复杂度.对上述回归型支持向量机而言,当$|f(\boldsymbol{x}_i)-y_i|\geq\varepsilon$时,权值$(\alpha_i-\alpha_i^*)$非零,因此当噪声较大时,支持向量的比例是相当大的.例如,对于噪声在±1范围内均匀分布的情况,若取$\varepsilon=0.1$,则支持向量约占样本总数的90%.选取大的ε值将减少支持向量数目,但将增加误差.下面描述的简化算法可以在误差容许的范围内大幅度地减少支持向量数目.

首先,对上述支持向量机,利用式(2)对每个样本$\boldsymbol{x}_i$计算

$$z_i=\begin{cases}1 & \text{当 } f(x_i)\geq y_i\\ -1 & \text{其他}\end{cases},\quad i=1,\dots,k. \tag{3}$$

新的回归函数也用$f(\boldsymbol{x})$表示:

$$f(\boldsymbol{x})=\boldsymbol{w}\cdot\varphi(\boldsymbol{x})+b, \tag{4}$$

其中$\boldsymbol{w}$为权向量,b为标量,“·”表示向量的点积,$\varphi(\boldsymbol{x})$为一个非线性映射.优化问题可表示为最小化目标函数:

$$R(\boldsymbol{w},\boldsymbol{\xi})=\frac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w}+\mathrm{C}\sum_{i=1}^{k}\xi_i, \tag{5}$$

其约束条件为

$$z_i(\boldsymbol{w}\cdot\varphi(\boldsymbol{x}_i)+b-y_i)\geq\varepsilon-\xi_i,\quad \xi_i\geq 0,\quad i=1,\dots,k. \tag{6}$$

其中式(5)的第1项使函数更为平坦,提高了泛化能力,第2项则为减少误差,常数C对两者作出折衷,ε为一个正常数.这是一个二次优化问题,引入拉格郎日函数

$$L(\boldsymbol{w},\boldsymbol{b},\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\gamma})=\frac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w}+\mathrm{C}\sum_{i=1}^{k}\xi_i-\sum_{i=1}^{k}\alpha_i[z_i(\boldsymbol{w}\cdot\varphi(\boldsymbol{x}_\mathrm{i})+b-y_i)-\varepsilon+\xi_i]-\sum_{i=1}^{k}\gamma_i\xi_i, \tag{7}$$

其中$\alpha_i\geq 0,\gamma_i\geq 0,i=1,\dots,k$.

函数L应对α_i和γ_i最大化,且对$\boldsymbol{w},b$和ξ_i最小化.函数L的极值应满足条件:

$$\frac{\partial}{\partial\boldsymbol{w}}L=0,\ \ \frac{\partial}{\partial b}L=0,\ \ \frac{\partial}{\partial\xi_i}L=0. \tag{8}$$

从而得到

$$\boldsymbol{w}=\sum_{i=1}^{k}\alpha_i z_i\varphi(\boldsymbol{x}_i);\quad \sum_{i=1}^{k}\alpha_i z_i=0;\quad \mathrm{C}-\alpha_i-\gamma_i,\quad i=1,\dots,k. \tag{9}$$

将以上3式代入式(7),可以得到优化问题的对偶形式,最小化函数

$$W(\boldsymbol{\alpha})=\frac{1}{2}\sum_{i,j=1}^{k}\alpha_i\alpha_j z_i z_j K(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^{k}(z_i y_i+\varepsilon)\alpha_i, \tag{10}$$

其约束为

$$\sum_{i=1}^{k}\alpha_i z_i=0,\quad 0\leq\alpha_i\leq\mathrm{C},\quad i=1,\dots,k. \tag{11}$$

这是一个典型的二次优化问题,已有高效的算法求解,回归函数可表示为

$$f(\boldsymbol{x})=\sum_{i=1}^{k}\alpha_i z_i K(\boldsymbol{x},\boldsymbol{x}_\mathrm{i})+b. \tag{12}$$

按照优化理论的Kuhn-Tucker定理,在鞍点,对偶变量与约束的乘积为0,即

$$\alpha_i[z_i(\boldsymbol{w}\cdot\varphi(\boldsymbol{x}_i)+b-y_i)-\varepsilon+\xi_i]=0,\quad \gamma_i\xi_i=0,\quad i=1,\dots,k. \tag{13}$$

当$0<\alpha_i<\mathrm{C}$时,由式(9)知$0<\gamma_i<\mathrm{C}$,由式(13)知$\xi_i=0$,且方括号内的项应为0,因此有

$$z_i(f(\boldsymbol{x}_i)-y_i)=\varepsilon; \tag{14}$$

当$\alpha_i=0$时,由式(9)知$\gamma_i=\mathrm{C}$,由式(13)知$\xi_i=0$,且方括号内的项可不为0,因此有

$$z_i(f(\boldsymbol{x}_i)-y_i)\geq\varepsilon; \tag{15}$$

当$\alpha_i=\mathrm{C}$时,由式(9)知$\gamma_i=0$,由式(13)知$\xi_i\geq 0$,且方括号内的项应为0,因此有

$$z_i(f(\boldsymbol{x}_i)-y_i)\leq\varepsilon. \tag{16}$$

由以上3式可知,当噪声较大时,支持向量的比例将是相当小的.对上例噪声在±1的范围内均匀分布的情况,若取$\varepsilon=0.1$,则支持向量约占样本总数的 10%,大大减小了复杂度.但是,当ε进一步减小时,由于 $f(\boldsymbol{x})$将会接近位于均值附近的样本,因此支持向量的比例不会按比例减小,而是趋近于一个常数.

2 大规模数据的高效算法

Suykens 等人提出的最小平方支持向量机(least square support vector machine,简称 LS-SVM)算法具有很高的学习效率,对大规模数据可采用共轭梯度法求解[4],但由于该算法得到的回归函数中非零权值的样本数接近于全部样本数,因此,计算复杂度较大.Platt 提出的串行最小化方法 SMO(sequential minimal optimization)对分类问题有较高的效率[5],但对支持向量数较多的回归问题效率欠佳.

本文提出的方法是,首先用 LS-SVM 算法求解回归问题,再用 SMO 算法求解本文描述的简化式,可获得运行时间短且支持向量少的效果.

3 实验结果

本节的例子包括 100 个样本,来自 $y=10\sin x/x$,并迭有噪声.实验采用二次规划的内点算法实现回归型支持向量机(记为 QPA),算法中选取 RBF 核函数,参数$\sigma=2$,常量ε选为 0.6,0.1 和 0.01,结果见表 1,其中平均误差 *error* 定义为

$$error=\frac{1}{100}\sum_{i=1}^{100}|f(x_i)-y_i|. \tag{17}$$

可见,当$\varepsilon=0.01$时,支持向量数为 99;当$\varepsilon=0.6$时,误差增大,但支持向量数减少为 42.本文的方法则分两步进行,首先用$\varepsilon=0.01$ 实现支持向量机,再用简化算法(记为 SPA)进行简化见表 1.可见,在误差类似时,支持向量数减少到 6.

事实上,如果第 1 步采用最小平方支持向量机(记为 LSA),第 2 步采用串行最小化方法实现简化算法,还可以提高学习效率,并适用于大规模数据的情况.表 2 分别给出了以上两步的结果.

Table 1 Applications of the simplification algorithm in the example

表 **1** 简化算法在例中的应用

Algorithm	ε	Average error	No. of SVs	Running time (s)
QPA	0.6	0.716	42	1.68
QPA	0.1	0.704	90	1.72
QPA	0.01	0.701	99	1.64
SPA	0.01	0.715	6	0.18

Table 2 Applications of the efficient learning algorithm in the example

表 **2** 高效学习算法在例中的应用

Algorithm	ε	Average error	No. of SVs	Running time(s)
LSA	–	0.708	100	0.10
SPA	0.01	0.700	9	0.20

4 结束语

在有噪声的情况下,采用二次规划算法的回归型支持向量机中,仅仅靠近样本均值的少量样本对应的权值为零,因此复杂度较高.本文提出的简化算法反其道而行之,仅仅靠近样本均值的少量样本对应的权值非零,因此,在容许的误差范围内大大降低了计算复杂度.对于大规模数据的情况,本文提出的方法充分利用了最小平方支持向量机学习效率高的优点,同时也克服了其复杂度过高的缺点,是较为实用的学习算法.

References:

[1] Cortes, C., Vapnik, V. Support vector networks. Machine Learning, 1995,20:273~297.

[2] Burges, C.J.C. Simplified support vector decision rules. In: Saitta, L., ed. Proceedings of the 13th International Conference on Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1996. 71~77.

[3] Schölkopf, B., Bartlett, P., Smola, A., *et al*. Support vector regression with automatic accuracy control. In: Niklasson, L., Bodén, M., Ziemke, T., eds. Proceedings of the ICANN'98, Perspectives in Neural Computing. Berlin: Springer-Verlag, 1998. 111~116.

[4] Suykens, J.A.K., Lukas, L., Van Dooren, P., *et al*. Least squares support vector machine classifiers: a large scale algorithm. In: European Conference on Circuit Theory and Design, ECCTD'99. 1999. 839~842. http://www.kernel-machines.org/papers/SuyDooMooVan99.ps.gz

[5] Platt, J. Fast training of support vector machines using sequential minimal optimization. In: Schölkopf, B., Berges, C.J.C., Smola, A.J., eds. Advances in Kernel Methods——Support Vector Learning. Cambridge, MA: MIT Press, 1999. 185~208.

A Simplification Algorithm to Support Vector Machines for Regression*

TIAN Sheng-feng, HUANG Hou-kuan

(*School of Computer and Information Technology*, *Northern Jiaotong University*, *Beijing* 100044, *China*)

E-mail: sftian@center.njtu.edu.cn

http://www.njtu.edu.cn

Abstract: Aiming at the computational complexity resulted from the large amounts of support vectors when the support vector machines (SVMs) are used in function estimation, a simplification algorithm is presented to reduce the number of support vectors and simplify applications. By the adaptation of the simplification algorithm, the LS-SVM (least square support vector machine) algorithm can be combined with SMO (sequential minimal optimization) algorithm to achieve good results with high learning efficiency and a few number of support vectors.

Key words: support vector machine; regression; machine learning; computational complexity; algorithm

---

* Received August 23, 2000; accepted April 3, 2001
Supported by the Science and Technology Development Program of the Railway Ministry of China under Grant No.2000X030-A