

基于双支持向量回归机的增量学习算法

郝运河 张浩峰

(南京理工大学计算机科学与工程学院 南京 210094)

摘要 提出了一种基于双支持向量回归机的增量学习算法。将获取到的新样本加入训练数据集后,该算法无需在整个新的数据集上重新训练双支持向量回归机,而是充分利用增量前的计算信息,从而大大减少了模型更新中逆矩阵的计算量,提高了算法的执行效率。在人工数据集、时间序列预测和 UCI 数据集上的数值实验表明,该算法快速有效。

关键词 双支持向量回归机,增量学习,逆矩阵,时间序列

中图分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.2.048

Incremental Learning Algorithm Based on Twin Support Vector Regression

HAO Yun-he ZHANG Hao-feng

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract This paper proposed an incremental learning algorithm based on twin support vector regression. When a new sample is added to the training set, our algorithm makes full use of old computing information instead of training all the new training set, so it greatly simplifies the calculation of inverse matrix and improves the execution efficiency. Experimental results on artificial datasets, time series and UCI datasets show that our algorithm has remarkable improvement of generalization performance with short training time.

Keywords Twin support vector regression, Incremental learning, Inverse matrix, Time series

1 引言

支持向量机(SVM)是 Vapnik 基于统计学习理论提出的一种新的机器学习方法,已成功地应用到多种领域^[1-3]。通过最大化边界实现结构风险最小化,具有良好的学习性能和泛化能力。对于支持向量回归问题,许多学者进行了研究,已经提出了一些解决方法,如多核线性规划支持向量回归机^[4],最小二乘支持向量回归机^[5](LS-SVR)等。

双支持向量回归机(TSVR)借鉴于双支持向量机^[6](TSVM)思想,TSVM 通过求解两个形如 SVM 的问题来构建两个非平行超平面,将一个大的二次规划问题转化为两个小的二次规划问题,能将训练时间约减到原 SVM 的 1/4。TSVM 提出后,许多学者对 TSVR 做了进一步的研究,如文献^[7]提出了 TSVR 输出的概率模型,使得分类器输出为校准的后验概率,文献^[8]提出了一种最小二乘递归投影双支持向量机,该方法通过添加正则化项使得优化问题正定,可以获得更好的泛化性能,将不等式约束转化为等式约束,避免了求解二次规划问题,提高了训练速度。基于同样的思想,TSVR 也是寻找两个非平行超平面,实验结果证明^[9],TSVR 在泛化性能和训练时间上都具有极大的优势。因此,TSVR 也迅速成为支持向量机领域的研究热点,如文献^[10]提出了一种 ϵ 双支持

向量机(ETSVR),其通过添加正则化项 $\|\omega\|^2 + b^2$ 来实现结构风险最小化,同样,文献^[11]中也采用了这种方法。

由于数据更新使旧的模型失效,从而无法满足现实世界的需求,因此研究增量学习算法具有极大的现实意义,并且取得了一定的成果^[12-15]。增量学习算法具有显著的优越性,首先,无需保存历史训练数据,从而可以节省存储空间;其次,在新的训练中充分利用了历史的训练结果,可以减少后续训练的时间,提高效率;最后,增量学习可以很好地解决渐变问题。

本文结合双支持向量回归机与增量学习的优点,提出了一种基于双支持向量回归机的增量学习算法。在双支持向量回归机的训练过程中,计算量主要集中在矩阵的求逆上。每当有新的训练样本加入时,训练样本集发生变化,旧样本集训练出的回归模型已无法准确地描述新样本集的信息,回归模型需要更新。相应地,回归模型的更新导致需要重新计算矩阵的逆,计算复杂度,内存占用量大,训练效率低,其原因在于完全忽略了以前的计算结果。实际上,在旧样本集上训练旧的回归模型时,已经计算过矩阵的逆,考虑到如果可以充分利用以前的计算结果,将大大地提高回归模型的训练效率,降低内存的占用量。基于以上思想,本文对双支持向量回归机的增量学习进行了深入研究,提出的增量学习方法充分利用了以前的计算结果,无需对新样本集重新训练,使得新增样本

到稿日期:2015-01-12 返修日期:2015-04-13 本文受国家自然科学基金(61101197),水下机器人技术国防科技重点实验室基金(9140C270205120C2701)资助。

郝运河(1988-),男,硕士生,主要研究方向为机器学习、模式识别,E-mail:yunhe_hao@126.com;张浩峰(1983-),男,博士,副教授,主要研究方向为机器学习、模式识别、智能机器人。

后矩阵逆的计算量大大降低,提高了模型的训练速度。

2 双支持向量回归机

考虑下面的回归问题:设训练样本集 (A, Y) ,其中 A 是 $m \times n$ 的矩阵,第 i 行代表第 i 个训练样本, $i=1, 2, \dots, m, Y=(y_1; y_2; \dots; y_m)$ 为输出样本值,其中 $y_i \in \mathbb{R}$,求回归函数 $y=f(x)$ 。

TSVR的思想是寻找下界回归函数: $f_1 = \omega_1^T x + b_1$ 和上界回归函数: $f_2 = \omega_2^T x + b_2$,其中, $\omega_1 \in \mathbb{R}^n, \omega_2 \in \mathbb{R}^n, b_1 \in \mathbb{R}$ 和 $b_2 \in \mathbb{R}$,即TSVR得到两个非平行超平面。引入 ϵ -不敏感损失函数来增加双支持向量回归机的推广能力, ϵ -不敏感损失函数是指当样本点 A_i 的预测值 $f(A_i)$ 与观测值 y_i 之差不超过 ϵ 时,则认为此预测值是无损失的,因此对目标函数来说存在一个无任何损失值的区域,即 ϵ -带。因此,标准线性TSVR解决两个二次规划问题:

$$\begin{aligned} \text{Min}_{\omega_1, b_1} & \frac{1}{2} \|\omega_1\|^2 + c_1 e^T \xi \\ \text{s. t. } & Y - (A\omega_1 + eb_1) \leq \epsilon e_1 - \xi \\ & \xi \geq 0 \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Min}_{\omega_2, b_2} & \frac{1}{2} \|\omega_2\|^2 + c_2 e^T \eta \\ \text{s. t. } & (A\omega_2 + eb_2) - Y \leq \epsilon e_2 - \eta \\ & \eta \geq 0 \end{aligned} \quad (2)$$

其中, $c_1, c_2 > 0, \epsilon_1, \epsilon_2$ 为不敏感损失函数, ξ, η 为松弛变量, e 为相应维数的单位向量。

对于非线性情况,引入核函数 K ,下界回归函数和上界回归函数变为 $f_1 = \omega_1^T K(x^T, A^T) + b_1$ 和 $f_2 = \omega_2^T K(x^T, A^T) + b_2$ 。标准非线性TSVR解决两个二次规划问题:

$$\begin{aligned} \text{Min}_{\omega_1, b_1} & \frac{1}{2} \|\omega_1\|^2 + c_1 e^T \xi \\ \text{s. t. } & Y - (K(A, A^T)\omega_1 + eb_1) \leq \epsilon e_1 - \xi \\ & \xi \geq 0 \end{aligned} \quad (3)$$

$$\begin{aligned} \text{Min}_{\omega_2, b_2} & \frac{1}{2} \|\omega_2\|^2 + c_2 e^T \eta \\ \text{s. t. } & (K(A, A^T)\omega_2 + eb_2) - Y \leq \epsilon e_2 - \eta \\ & \eta \geq 0 \end{aligned} \quad (4)$$

通过拉格朗日乘子法和极值条件将原始的两个二次规划问题转变成其对偶问题,通过对偶问题的求解,得到回归函数 f_1 和 f_2 ,将 f_1 和 f_2 合并得到最终的回归函数 $f = \frac{1}{2}(f_1 + f_2)$ 。

3 基于双支持向量机的增量学习算法

为了进行双支持向量回归机的增量学习,首先给出一个引理。

引理 1^[16] 设 $A \in \mathbb{R}^{n \times n}$ 是非奇异矩阵, $u, v \in \mathbb{R}^n$ 是任意向量,若 $1 + v^T A^{-1} u \neq 0$,则 $A + uv^T$ 非奇异,且其逆矩阵可表示为:

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} uv^T A^{-1}}{1 + v^T A^{-1} u} \quad (5)$$

考虑到拉格朗日支持向量回归机的优越性,将此方法引入TSVR,添加正则化项 $\frac{1}{2}(\|\omega\|^2 + b^2)$,最大化间隔由原来的 n 维空间变为 $n+1$ 维空间,使TSVR更加合理;将松弛

变量由1-范数变为2-范数的平方,从而省去了非负约束。最后,将不等式约束转化为等式约束。将使用以上方法矫正的TSVR记为LTSVR。

3.1 线性双支持向量回归机的增量学习算法

本节采用线性LTSVR来进行增量学习。线性LTSVR模型如下:

$$\begin{aligned} \text{Min}_{\omega_1, b_1} & \frac{1}{2} (\|\omega_1\|^2 + b_1^2) + \frac{c_1}{2} \xi^T \xi \\ \text{s. t. } & Y - (A\omega_1 + eb_1) - \epsilon e_1 = \xi \end{aligned} \quad (6)$$

$$\begin{aligned} \text{Min}_{\omega_2, b_2} & \frac{1}{2} (\|\omega_2\|^2 + b_2^2) + \frac{c_2}{2} \eta^T \eta \\ \text{s. t. } & (A\omega_2 + eb_2) - Y - \epsilon e_2 = \eta \end{aligned} \quad (7)$$

与式(1)、式(2)相比,式(6)、式(7)转化为等式约束的微弱强凸函数的最小化问题,省去了迭代求解的繁琐步骤,通过求解线性方程得到最优化问题的解。

消除松弛变量,得:

$$\begin{aligned} \text{Min}_{\omega_1, b_1} & \frac{1}{2} (\|\omega_1\|^2 + b_1^2) + \frac{c_1}{2} \|Y - (A\omega_1 + eb_1) - \epsilon e_1\|^2 \end{aligned} \quad (8)$$

$$\begin{aligned} \text{Min}_{\omega_2, b_2} & \frac{1}{2} (\|\omega_2\|^2 + b_2^2) + \frac{c_2}{2} \|(A\omega_2 + eb_2) - Y - \epsilon e_2\|^2 \end{aligned} \quad (9)$$

令 $G = [A \ e], f = Y - \epsilon e_1, h = Y + \epsilon e_2, u_1 = [\omega_1^T, b_1], u_2 = [\omega_2^T, b_2]$,则上式表示为:

$$\text{Min}_{u_1} L_1(u_1) = \frac{1}{2} u_1^T u_1 + \frac{c_1}{2} \|f - Gu_1\|^2 \quad (10)$$

$$\text{Min}_{u_2} L_2(u_2) = \frac{1}{2} u_2^T u_2 + \frac{c_2}{2} \|Gu_2 - h\|^2 \quad (11)$$

采用最优性条件,可知:

$$\frac{\partial L_1}{\partial u_1} = u_1 - c_1 G^T (f - Gu_1) = 0 \quad (12)$$

$$\frac{\partial L_2}{\partial u_2} = u_2 + c_2 G^T (Gu_2 - h) = 0 \quad (13)$$

求解上式得:

$$u_1 = (I + c_1 G^T G)^{-1} (c_1 G^T f) \quad (14)$$

$$u_2 = (I + c_2 G^T G)^{-1} (c_2 G^T h) \quad (15)$$

假设新增样本点 (x_{m+1}, y_{m+1}) ,令 $g = [x_{m+1}, e], Q_i = I + c_i G^T G, i=1, 2$,则线性LTSVR变为:

$$\text{Min}_{u_1} L_1(u_1) = \frac{1}{2} u_1^T u_1 + \frac{c_1}{2} \|[f; y_{m+1} - \epsilon_1] - [G; g]u_1\|^2 \quad (16)$$

$$\text{Min}_{u_2} L_2(u_2) = \frac{1}{2} u_2^T u_2 + \frac{c_2}{2} \|[G; g]u_2 - [h; y_{m+1} + \epsilon_2]\|^2 \quad (17)$$

由最优性条件及引理1求解,得:

$$\begin{aligned} u_1 &= (I + c_1 G^T G + c_1 g^T g)^{-1} (c_1 G^T f + c_1 g^T (y_{m+1} - \epsilon_1)) \\ &= (Q_1 + c_1 g^T g)^{-1} (c_1 G^T f + c_1 g^T (y_{m+1} - \epsilon_1)) \\ &= (Q_1^{-1} - \frac{Q_1^{-1} c_1 g^T g Q_1^{-1}}{1 + c_1 g^T Q_1^{-1} g}) (c_1 G^T f + c_1 g^T (y_{m+1} - \epsilon_1)) \end{aligned} \quad (18)$$

$$\begin{aligned} u_2 &= (I + c_2 G^T G + c_2 g^T g)^{-1} (c_2 G^T h + c_2 g^T (y_{m+1} + \epsilon_2)) \\ &= (Q_2 + c_2 g^T g)^{-1} (c_2 G^T h + c_2 g^T (y_{m+1} + \epsilon_2)) \\ &= (Q_2^{-1} - \frac{Q_2^{-1} c_2 g^T g Q_2^{-1}}{1 + c_2 g^T Q_2^{-1} g}) (c_2 G^T h + c_2 g^T (y_{m+1} + \epsilon_2)) \end{aligned} \quad (19)$$

从式(18)、式(19)可以看出,每加入一个新的样本点 (x_{m+1}, y_{m+1}) ,在计算 u_1 和 u_2 时,可以利用以前的计算结果(即 Q^{-1})来完成当前计算,从而大大减少了计算量。

综上所述,线性 LTSVR 的增量学习算法流程如下:

设初始样本集为 (A, Y) ,其中 A 是 $m \times n$ 的矩阵,共 m 个样本, $Y=(y_1; y_2; \dots; y_m)$ 为输出样本值。

Step 1:初始化参数 c_1 和 c_2, ϵ_1 和 ϵ_2 。

Step 2:使用式(14)和式(15)计算得到参数 u_1 和 u_2 ,并保存计算结果 Q^{-1} ,用于增量学习。

Step 3:新增样本 (x_{m+1}, y_{m+1}) ,使用式(18)和式(19)计算参数 u_1 和 u_2 ,并相应地更新 Q^{-1} 。

Step 4:以后新增样本均使用 Step 3 中方法计算参数 u_1 和 u_2 。

3.2 非线性双支持向量回归机的增量学习算法

对于非线性 LTSVR 模型,引入核函数 K ,其模型变为:

$$\text{Min}_{\omega_1, b_1} \frac{1}{2} (\|\omega_1\|^2 + b_1^2) + \frac{c_1}{2} \xi^T \xi \quad (20)$$

$$\text{s. t. } Y - (K(A, A^T)\omega_1 + eb_1) - \epsilon\epsilon_1 = \xi$$

$$\text{Min}_{\omega_2, b_2} \frac{1}{2} (\|\omega_2\|^2 + b_2^2) + \frac{c_2}{2} \eta^T \eta \quad (21)$$

$$\text{s. t. } (K(A, A^T)\omega_2 + eb_2) - Y - \epsilon\epsilon_2 = \eta$$

消除松弛变量,得:

$$\text{Min}_{\omega_1, b_1} \frac{1}{2} (\|\omega_1\|^2 + b_1^2) + \frac{c_1}{2} \|Y - (K(A, A^T)\omega_1 + eb_1) - \epsilon\epsilon_1\|^2 \quad (22)$$

$$\text{Min}_{\omega_2, b_2} \frac{1}{2} (\|\omega_2\|^2 + b_2^2) + \frac{c_2}{2} \|(K(A, A^T)\omega_2 + eb_2) - Y - \epsilon\epsilon_2\|^2 \quad (23)$$

令 $G=[K(A, A^T) \ e]$, $f=Y-\epsilon\epsilon_1$, $h=Y+\epsilon\epsilon_2$, $u_1=[\omega_1^T, b_1]$, $u_2=[\omega_2^T, b_2]$,则上式表示为:

$$\text{Min}_{u_1} L_1(u_1) = \frac{1}{2} u_1^T u_1 + \frac{c_1}{2} \|f - Gu_1\|^2 \quad (24)$$

$$\text{Min}_{u_2} L_2(u_2) = \frac{1}{2} u_2^T u_2 + \frac{c_2}{2} \|Gu_2 - h\|^2 \quad (25)$$

采用最优性条件,可知:

$$\frac{\partial L_1}{\partial u_1} = u_1 - c_1 G^T (f - Gu_1) = 0 \quad (26)$$

$$\frac{\partial L_2}{\partial u_2} = u_2 + c_2 G^T (Gu_2 - h) = 0 \quad (27)$$

求解上式得:

$$u_1 = (I + c_1 G^T G)^{-1} (c_1 G^T f) \quad (28)$$

$$u_2 = (I + c_2 G^T G)^{-1} (c_2 G^T h) \quad (29)$$

假设新增样本点 (x_{m+1}, y_{m+1}) ,令 $g=[K(x_{m+1}, A^T), e]$, $Q_i = I + c_i G^T G, i=1, 2$,则非线性 LTSVR 变为:

$$\text{Min}_{u_1} L_1(u_1) = \frac{1}{2} u_1^T u_1 + \frac{c_1}{2} \|[f; y_{m+1} - \epsilon_1] - [G; g]u_1\|^2 \quad (30)$$

$$\text{Min}_{u_2} L_2(u_2) = \frac{1}{2} u_2^T u_2 + \frac{c_2}{2} \|[G; g]u_2 - [h; y_{m+1} + \epsilon_2]\|^2 \quad (31)$$

由最优性条件及引理 1 求解,得:

$$\begin{aligned} u_1 &= (I + c_1 G^T G + c_1 g^T g)^{-1} (c_1 G^T f + c_1 g^T (y_{m+1} - \epsilon_1)) \\ &= (Q_1 + c_1 g^T g)^{-1} (c_1 G^T f + c_1 g^T (y_{m+1} - \epsilon_1)) \\ &= (Q_1^{-1} - \frac{Q_1^{-1} c_1 g^T g Q_1^{-1}}{1 + c_1 g^T Q_1^{-1} g}) (c_1 G^T f + c_1 g^T (y_{m+1} - \epsilon_1)) \end{aligned} \quad (32)$$

$$\begin{aligned} u_2 &= (I + c_2 G^T G + c_2 g^T g)^{-1} (c_2 G^T h + c_2 g^T (y_{m+1} + \epsilon_2)) \\ &= (Q_2 + c_2 g^T g)^{-1} (c_2 G^T h + c_2 g^T (y_{m+1} + \epsilon_2)) \\ &= (Q_2^{-1} - \frac{Q_2^{-1} c_2 g^T g Q_2^{-1}}{1 + c_2 g^T Q_2^{-1} g}) (c_2 G^T h + c_2 g^T (y_{m+1} + \epsilon_2)) \end{aligned} \quad (33)$$

综上所述,非线性 LTSVR 的增量学习算法流程如下:

设初始样本集为 (A, Y) ,其中 A 是 $m \times n$ 的矩阵,共 m 个样本, $Y=(y_1; y_2; \dots; y_m)$ 为输出样本值。

Step 1:初始化参数 c_1 和 c_2, ϵ_1 和 ϵ_2 ,核函数 K 。

Step 2:使用式(28)和式(29)计算得到参数 u_1 和 u_2 ,并保存计算结果 Q_i^{-1} ,用于增量学习。

Step 3:新增样本 (x_{m+1}, y_{m+1}) ,首先使用核函数 K 将样本点映射到高维空间,然后使用式(32)和式(33)计算参数 u_1 和 u_2 ,并相应地更新 Q_i^{-1} 。

Step 4:以后新增样本均使用 Step 3 中方法计算参数 u_1 和 u_2 。

3.3 计算复杂度分析

设初始样本集为 (A, Y) ,其中 A 是 $m \times n$ 的矩阵,共 m 个样本, $Y=(y_1; y_2; \dots; y_m)$ 为输出样本值,令 $G=[A \ e]$ 。

考虑到计算机做乘法的复杂度远高于加法,因此以下关于时间复杂度的分析只考虑算法所需做乘法的次数。假设新增样本 (x_{m+1}, y_{m+1}) ,令 $g=[x_{m+1}, e]$ 。若不进行增量学习,即在新的样本集上重新训练双支持向量回归机,则需计算矩阵表达式 $(I + c_1 G^T G + c_1 g^T g)^{-1}$ 。令 $M = I + c_1 G^T G + c_1 g^T g$,则其时间复杂度主要由 3 部分组成: $c_1 G^T G, c_1 g^T g$ 和 M^{-1} 。其中 $c_1 G^T G$ 需要做 $(n+1)m + m(n+1)^2$ 次乘法, $c_1 g^T g$ 需要做 $(n+1) + (n+1)^2$ 次乘法,对于 M^{-1} ,由于矩阵求逆有不同的方法,但一般情况下,其乘法次数不小于 $(n+1)^3$ 。若进行增量训练, Q_i^{-1} 已知,则需计算 $Q_2^{-1} - \frac{Q_2^{-1} c_2 g^T g Q_2^{-1}}{1 + c_2 g^T Q_2^{-1} g}$,其大约需要做 $2(n+1) + 4(n+1)^2 + (n+1)^3$ 次乘法。比较两种不同的计算方法,则一般情况下,前者比后者大约需要多做 $(m-1)(n-1) + (m-3)(n+1)^2$ 次乘法。因此,增量学习可以有效地减少训练时间。

4 数值实验与分析

本节进行仿真实验来验证算法的有效性及其优势。为了克服参数寻优的困难,在本文实验中,令 $c_1 = c_2 = 2^i$,指数 i 在 $[-10, 10]$ 的范围内寻找最优值, $\epsilon_1 = \epsilon_2 = 0.01$ 。核函数采用 RBF 核函数 $\exp(-\frac{\|x-z\|^2}{2\sigma^2})$,文献[17]给出了参数 σ 的极性质,当 $\sigma > 0$ 且过小时,可对所有训练样本正确分类,这很容易造成“过学习”的情况;当 σ 过大时,其推广能力或对新样本的正确分类能力为零;因此,参数不宜过大,也不宜过小,在本文实验中,参数 σ 在 $[1, 10]$ 范围内寻找最优值。

本文将从回归性能和训练时间两个方面考察 ETSVR、LTSVR 和 ILTSVR 3 种算法的优劣。实验数据包括两个人工数据集、一个时间序列和 5 个 UCI 数据集。实验的性能评价准则为: $SSE = \sum_{i=1}^m (y_i - \bar{y}_i)^2$;其中, y_i 为准确值, \bar{y}_i 为回归函数值。

4.1 人工数据集

(1) 仿真目标函数为: $y = x^{2/3}$ 。通过以下方法获得训练样本集:

$$y=x^{2/3}+\delta, x\sim U[-10,10], \delta\sim N(0,1^2)$$

其中, U 是在 $[-10,10]$ 范围内的均匀分布, N 是均值为 0、标准差为 1 的正态分布, δ 为加入的噪声数据。本实验获得 3000 个样本点。

(2) 仿真目标函数 $y=\sin(x)/x$, 通过以下方法获得训练样本集:

$$y=\sin(x)/x+\delta, x\sim U[-4\pi,4\pi], \delta\sim N(0,0.2^2)$$

其中, U 是在 $[-4\pi,4\pi]$ 的均匀分布, N 是均值为 0、标准差为 0.2 的正态分布, δ 为加入的噪声数据。本实验获得 5000 个样本点。

(3) 仿真实验结果

两个人工数据集均采用非线性 LTSVR 进行实验, 核函数采用 RBF 核函数。数据集的前 90% 的样本作为训练数据, 剩余的 10% 作为测试数据。人工数据集最优参数值及 SSE 值如表 1 所列。

表 1 人工数据集最优参数值及 SSE 值

	$y=x^{2/3}$			$y=\sin(x)/x$		
	LTSVR	ILTSVR	ETSVR	LTSVR	ILTSVR	ETSVR
i	0	0	1	-7	-7	-3
σ	3	3	4	5	5	10
SSE	1.591	1.591	1.926	0.113	0.113	0.118

拟合函数效果图如图 1、图 2 所示。

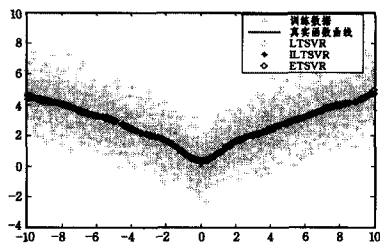


图 1 $y=x^{2/3}$ 的实验结果

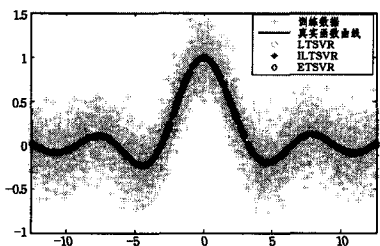


图 2 $y=\sin(x)/x$ 的效果图

从图 1、图 2 和表 1 中可以看出, 在仿真目标函数 $y=x^{2/3}$ 和 $y=\sin(x)/x$ 上, ETSVR、LTSVR 和 ILTSVR 效果相当, 都很好地拟合了真实的函数模型, 但 LTSVR 和 ILTSVR 的 SSE 值略小于 ETSVR。

对于模型训练时间, 我们将训练数据分成 9 份, 做 9 次训练, 第 n 次训练样本数为前 n 份样本。训练时间结果如图 3、图 4 所示。

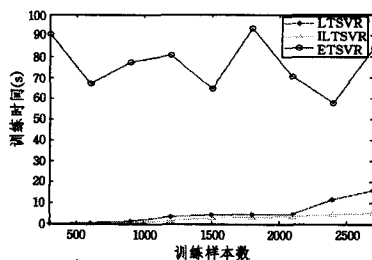


图 3 $y=x^{2/3}$ 的训练时间

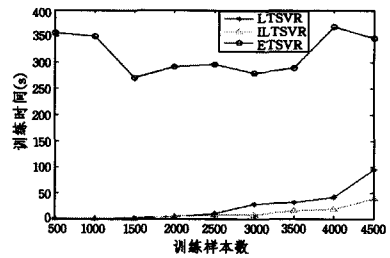


图 4 $y=\sin(x)/x$ 的训练时间

从图 3 和图 4 可以看出, 首先, 相比 ETSVR, LTSVR 和 ILTSVR 大大减少了模型的训练时间, 其原因在于 ETSVR 算法采用 SOR 技术迭代求解模型, 而 LTSVR 与 ILTSVR 只是进行简单的矩阵计算; 其次, 随着训练样本的增加, LTSVR 的训练时间逐渐高于 ILTSVR, 其原因在于随着训练样本增加, LTSVR 总是重新训练回归模型, 而 ILTSVR 是在以前的计算基础上训练回归模型, 因此 ILTSVR 具有更高的执行效率。

4.2 时间序列分析

进行时间序列预测, 首先需要根据嵌入定理重构输入输出空间。时间序列的时间延迟嵌入状态空间为:

$$X(T)=(x(t), x(t-td), \dots, x(t-(Em-1)td))$$

其中, $X(T)$ 是嵌入矢量, $x(t)$ 是第 t 时间点的序列值, td 是时延量, Em 是嵌入维数。此时可以通过嵌入矢量 $X(T)$ 预测第 $t+v$ 时间点的序列值 $x(t+v)$, v 表示预测步长。

Mackey 和 Glass 于 1977 年在时滞系统中发现混沌现象, 时滞混沌系统便引起了人们的广泛关注。Mackey-Glass 混沌时间序列常用来检验非线性系统模型的性能, 其如下式所示:

$$x(t+1)=(1-b)x(t)+\frac{ax(t-\tau)}{(1+x(t-\tau))^{10}}$$

当 $\tau > 17$ 时该时间序列呈现混沌性, τ 值越大, 混沌程度越高。本小节分别选取 $\tau=20$ 做单步预测 ($Em=4, td=4, v=1$) 和多步预测 ($Em=6, td=4, v=20$), 参数 $a=0.1, b=0.1$ 。单步预测产生 5000 个样本, 多步预测产生 3000 个样本, 分别选取前 90% 做训练, 后 10% 做测试。采用非线性 LTSVR 进行实验, 核函数为 RBF 核函数, 实验结果如图 5、图 6 所示, Mackey-Glass 最优参数值及 SSE 值如表 2 所列。

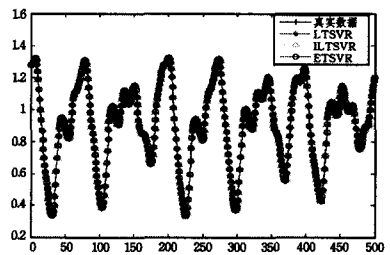


图 5 单步预测实验结果

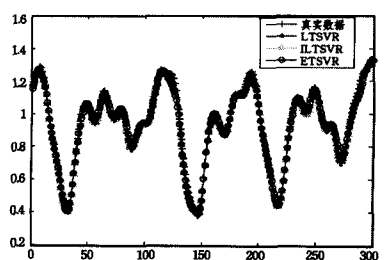


图 6 多步预测实验结果

表2 Mackey-Glass 最优参数值及 SSE 值

	单步			多步		
	LTSVR	ILTSVR	ETSVR	LTSVR	ILTSVR	ETSVR
i	10	10	-5	10	10	-8
σ	1	1	1	1	1	1
SSE	0.002	0.002	0.003	0.042	0.042	0.094

从图5、图6和表2可以看出,ETSVR、LTSVR和ILTSVR对Mackey-Glass时间序列的预测结果旗鼓相当,都达到了较好的预测效果,对于SSE值,LTSVR和ILTSVR略小于ETSVR。

下面对比三者的训练时间,将训练数据分成9份,做9次训练,第n次训练样本为前n份样本,实验结果如图7、图8所示。

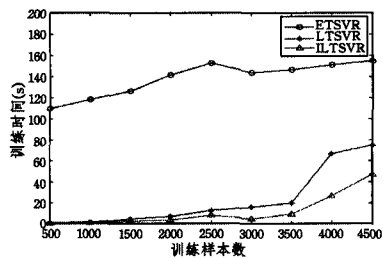


图7 单步预测训练时间

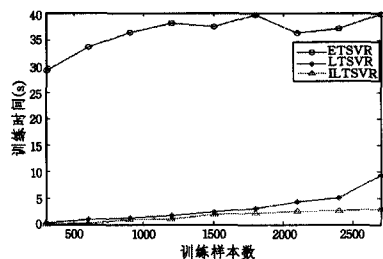


图8 多步预测训练时间

从图7和图8可以看出,在Mackey-Glass时间序列预测的实验中,LTSVR和ILTSVR的训练时间同样大大少于ETSVR的训练时间,且随着训练样本增加,ILTSVR训练时间逐渐少于LTSVR,且增长缓慢。

4.3 UCI数据集

为了进一步验证本文算法的有效性,从UCI机器学习标准数据库中选取了5个数据集进行实验,随机选取每个样本集中10%的样本作为测试数据。Wine、COIL、Housing采用线性LTSCR,MPG、Parkinson采用非线性LTSVR,核函数采用RBF函数。训练前数据采用下式做归一化处理:

$$x_i = \frac{x_i - \text{mean}(x_i)}{\sqrt{\text{var}(x_i) + C}}, i = 1, 2, \dots, m$$

其中,mean(x_i)为均值,var(x_i)为方差,C为常数,用于防止除以零的现象发生。

5个数据集基本属性及实验结果如表3、表4所列。

从表4可以看出,在5个UCI数据集上,LTSVR和ILTSVR的SSE值较小,回归效果良好。同样,ETSVR也可以获得良好的结果。

表3 数据集基本属性

数据集	Wine	MPG	Parkinson	COIL	Housing
样本数	4898	392	5875	5822	506
维数	12	8	26	86	14

表4 UCI数据集最优参数值及SSE值

数据集	LTSVR			ILTSVR			ETSVR		
	i	σ	SSE	i	σ	SSE	i	σ	SSE
Wine	10	/	0.011	10	/	0.011	-10	/	0.023
MPG	10	1	0.003	10	1	0.003	-10	1	0.004
Parkinson	9	10	0.045	9	10	0.045	-8	10	0.059
COIL	9	/	0.005	9	/	0.005	-10	/	0.013
Housing	8	/	0.002	8	/	0.002	-10	/	0.007

对比训练时间,将训练数据分成9份,做9次训练,第n次训练样本为前n份样本。以Parkinson和COIL为例,其他3个数据集训练时间图走势大体相同。从图9和图10可以看出,ETSVR训练时间较长,增长速度较快;LTSVR训练时间少于ETSVR,增长速度相对平缓;ILTSVR训练时间最短,增长速度最缓慢。因此,相比ETSVR,LTSVR与ILTSVR的执行效率更高,可以有效地节省训练时间,且随着训练样本的增加,ILTSVR在训练时间上的优势会进一步凸显出来。

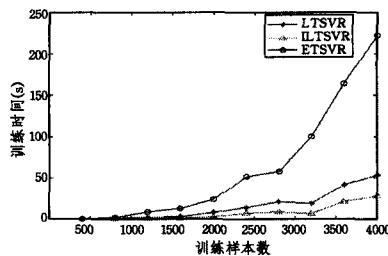


图9 Parkinson数据集训练时间

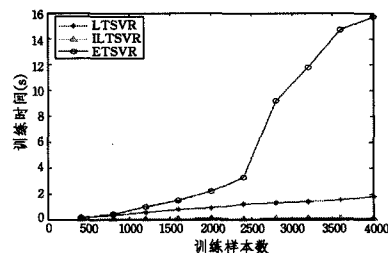


图10 COIL数据集训练时间

结束语 本文提出了基于双支持向量回归机的增量学习算法,算法结合了双支持向量回归机与增量学习的优势,在保证预测精度的前提下,可以大大缩短算法的计算时间。实验结果表明,算法快速有效。

随着训练样本的不断增多,样本集的规模会越来越大,导致计算量增大,训练效率降低。为了提高算法的执行效率,必须有效地缩减训练样本集的规模,因此下一步的研究工作是设计出高效删除冗余样本的方法,从而进一步满足实际应用的需求。

参考文献

- [1] Vapnik V. The nature of statistical learning theory[M]. Springer, 2000
- [2] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167
- [3] Cristianini N, Shawe-Taylor J. Kernel methods for pattern analysis[M]. Cambridge University Press, 2004
- [4] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3): 293-300

(下转第249页)

Hydrology, 2014, 34(4): 15-20 (in Chinese)

王继民, 朱跃龙, 李薇, 等. 多度量水文时间序列相似性分析[J]. 水文, 2014, 34(4): 15-20

- [2] Serrà J, Li J, Arcos. An empirical evaluation of similarity measures for time series classification[J]. Knowledge-Based Systems, 2014, 61(9): 305-314
- [3] Li Wei, Sun Hong-lin. Analysis and study on hydrological time series similarity search[J]. Journal of China Hydrology, 2009, 29(6): 76-80 (in Chinese)
李薇, 孙洪林. 水文时间序列相似性查询的分析与研究[J]. 水文, 2009, 29(6): 76-80
- [4] Gu Xin-chen, Wan Ding-sheng, Fan Long. Research and Application of Hydrological Time Series Similarity Based on Hadoop[J]. Computer & Digital Engineering, 2014, 42(11): 1-5 (in Chinese)
顾昕辰, 万定生, 樊龙. 基于 Hadoop 的水文时间序列相似性研究与应用[J]. 计算机与数字工程, 2014, 42(11): 1-5
- [5] Cheng Xi-feng, Wan Ding-sheng, Wang Ya-ming. Similarity search optimization algorithm in hydrological time series[J]. Computer Engineering and Design, 2013, 34(11): 4046-4050 (in Chinese)
程习锋, 万定生, 王亚明. 水文时间序列相似性查询优化算法[J]. 计算机工程与设计, 2013, 34(11): 4046-4050
- [6] Lin J, Keogh E, Lonardi S, et al. A symbolic representation of time series, with implications for streaming algorithms[C] // Proc of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. New York: ACM Press, 2003: 2-11
- [7] Yan Qiu-yan, Meng Fan-rong. A key point based SAX improving algorithm[J]. Journal of Computer Research and Development,

2009, 46(Suppl.): 483-490 (in Chinese)

闫秋艳, 孟凡荣. 一种基于关键点的 SAX 改进算法[J]. 计算机研究与发展, 2009, 46(Suppl.): 483-490

- [8] Zhu Yue-long, Wang Yong-mei, Wan Ding-sheng, et al. Similarity mining of hydrological time series based on semantic similarity measures[J]. Journal of China Hydrology, 2011, 31(1): 35-40 (in Chinese)
朱跃龙, 王咏梅, 万定生, 等. 基于语义相似的水文时间序列相似性挖掘[J]. 水文, 2011, 31(1): 35-40
- [9] Li Ying. Symbolization time series clustering based on DTW[J]. Microcomputer & Its Applications, 2011, 30(18): 3-5 (in Chinese)
李迎. 一种基于 DTW 的符号化时间序列聚类算法[J]. 微型机与应用, 2011, 30(18): 3-5
- [10] Ouyang Ru-lin, Ren Li-liang, Zhou Cheng-hu. Similarity search in hydrological time series[J]. Journal of Hohai University (Natural Sciences), 2010, 38(3): 241-245 (in Chinese)
欧阳如琳, 任立良, 周成虎. 水文时间序列的相似性搜索研究[J]. 河海大学学报(自然科学版), 2010, 38(3): 241-245
- [11] Ouyang Ru-lin, Ren Li-liang, Cheng Wei-ming, et al. Similarity search and pattern discovery in hydrological time series data mining[J]. Hydrol. Process, 2010, 24(9): 1198-1210
- [12] Zhang Xiao-hang, Liu Jia-qi, Du Yu, et al. A novel clustering method on time series data[J]. Expert Systems With Applications, 2011, 38(9): 11891-11900
- [13] Zhu Yue-long, Peng Li, Li Shi-jin, et al. Research on hydrological time series motifs mining[J]. Journal of Hydraulic Engineering, 2012, 43(11): 1422-1430 (in Chinese)
朱跃龙, 彭力, 李士进, 等. 水文时间序列模式挖掘[J]. 水利学报, 2012, 43(11): 1422-1430

(上接第 234 页)

- [5] Zhou Jin-zhu, Huang Jin. Multiple kernel liner programming support vector regression incorporating prior knowledge[J]. Acta Automatica Sinica, 2011, 37(3): 360-370 (in Chinese)
周金柱, 黄进. 集成先验知识的多核线性规划支持向量回归[J]. 自动化学报, 2011, 37(3): 360-370
- [6] Khemchandani R, Chandra S. Twin support vector machines for pattern classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905-910
- [7] Shao Y H, Deng N Y, Yang Z M, et al. Probabilistic outputs for twin support vector machines[J]. Knowledge-Based Systems, 2012, 33: 145-151
- [8] Shao Y H, Deng N Y, Yang Z M. Least squares recursive projection twin support vector machine for classification[J]. Pattern Recognition, 2012, 45(6): 2299-2307
- [9] Peng X. TSVR: an efficient twin support vector machine for regression[J]. Neural Networks, 2010, 23(3): 365-372
- [10] Shao Y H, Zhang C H, Yang Z M, et al. An ϵ -twin support vector machine for regression[J]. Neural Computing and Applications, 2013, 23(1): 175-185
- [11] Peng X. Primal twin support vector regression and its sparse approximation[J]. Neurocomputing, 2010, 73(16): 2846-2858
- [12] Lu G F, Zou J, Wang Y. Incremental learning of complete linear discriminant analysis for face recognition[J]. Knowledge-Based Systems, 2012, 31: 19-27
- [13] Deguchi T, Ishii N. On Memory Capacity in Incremental Learning with Appropriate Refractoriness and Weight Increment[C] // 2011 First ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering (CNSI). IEEE, 2011: 427-430
- [14] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1-3): 125-141
- [15] Khreich W, Granger E, Miri A, et al. A survey of techniques for incremental learning of HMM parameters[J]. Information Sciences, 2012, 197: 105-130
- [16] Yuan Ya-xiang, Sun Wen-yu. Optimization theory and methods [M]. Beijing: Science Press, 1997 (in Chinese)
袁亚湘, 孙文瑜. 最优化理论与方法 [M]. 北京: 科学出版社, 1997
- [17] Zhang Xiao-yun, Liu Yun-cai. Performance analysis of support vector machines with Gauss kernel[J]. Computer Engineering, 2003, 29(8): 22-25 (in Chinese)
张小云, 刘允才. 高斯核支撑向量机的性能分析[J]. 计算机工程, 2003, 29(8): 22-25