

基于改进的共轭梯度算法实现的最小二乘 隐空间支持向量机

王 玮, 娄 莉

(西安石油大学 计算机学院, 陕西 西安 710065)

摘 要: 本文研究了最小二乘隐空间支持向量机的优化问题. 文中采用基于对称超松弛预处理技术改进共轭梯度算法, 改进的共轭梯度算法只需求解一个阶数为 $L-1$ 的线性代数方程组即可, 大大节省了计算时间. 最后将其应用于最小二乘隐空间支持向量机中建立数学模型, 并通过实例验证了该算法的优越性.

关键词: 最小二乘隐空间支持向量机; 改进的共轭梯度算法; 预测

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2012)12-0099-04

The Implementation of Least Squares Hidden Space Support Vector Machines Based on Improved Conjugate Gradient Algorithm

WANG Wei, LOU Li

(School of Computer Science, Xi'an Shiyu University, Xi'an 710065, China)

Abstract: This research main focused on the improvement on Least Squares Hidden Space Support Vector Machines. Using Symmetric Successive Over-Relaxation technology to improve Conjugate Gradient Algorithm, it saves a lot of execution time costs for solving the system of linear algebraic equations with $L-1$ orders. At the end of this paper, the advantage of this improved algorithm is proved by applying it on mathematics modeling.

Key words: least squares hidden space support vector machines; improved conjugate gradient algorithm; prediction

1 引言

支持向量机 (Support Vector Machine, SVM) 是一种基于统计学习理论的机器学习技术系统^[1], 在小样本、非线性及高维模式识别等领域具有更好的处理性能. SVM 的数学模型可以归纳为一种对二次凸规划问题的探究, 这类问题是一种求解具有线性约束条件的二次凸规划问题, 并且通过引入核方法, 将样本空间中的内积运算替换成核函数, 从而克服了机器学习理论中的维数灾难以及局部极小等问题.

无论对于二分类问题还是回归问题, 原始支持向量机都需要解一个带不等式约束的二次规划问题. 为了简化 SVM 的运算过程, 节省运算时间, 提

高机器学习效率以及稳定性, 文献[2]中基于等式约束和最小二乘损失函数提出了解二分类问题的最小二乘支持向量机, 解回归问题的最小二乘支持向量机, 它实际上是一种岭回归. 文献[3]在此基础上针对带噪声的回归问题提出了加权最小二乘支持向量机. 文献[4-5]分别提出了一种基于隐空间的最小二乘支持向量机以及针对分类问题的基于矩阵模式的最小二乘支持向量机. 目前, 最小二乘支持向量机在分类和回归中得到了广泛的应用^[6-7]. 由于共轭梯度算法迭代次数较多, 收敛将很慢. 为减少迭代次数, 加快收敛速度, 可以设法先降低系数矩阵的条件数. 本文采用基于对称超松弛预处理技术改进共轭梯度算法, 并将其应用于最小二乘隐空间支持向量机中建立数学模型, 最后, 将建立的数学模型应用于

收稿日期: 2012-04-03; 修回日期: 2012-05-06

基金项目: 陕西省自然科学基金基础研究项目(2012JM8037, 2010JM8032); 陕西省教育厅科学研究计划项目(2010JK772)

股票价格的短期预测,得到了理想的预测效果.

2 最小二乘支持向量机

对于二分类问题,假设训练集 T 由 l 个样本点组成,

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \quad (1)$$

其中, $x_i \in R^n$ 是输入向量, $y_i \in \{-1, 1\}$ 是 x_i 所属的类别. 基于等式约束和最小二乘损失函数的优化问题为^[8]:

$$\min_{w, b, e} J(w, b, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^l e_k^2 \quad (2)$$

s. t.

$$y_k [w^T \varphi(x_k) + b] = 1 - e_k, \quad k = 1, \dots, l \quad (3)$$

式中 w 是权向量, γ 是正则化参数, e_k 是误差变量, $\varphi(\cdot)$ 从输入空间到高维特征空间的非线性映射, b 是一个偏量.

对应于优化问题(2)~(3)的拉格朗日函数为:

$$L(w, b, e, \alpha) = J(w, b, e) - \sum_{k=1}^l \alpha_k \{y_k [w^T \varphi(x_k) + b] - 1 + e_k\} \quad (4)$$

其中 α_k 是拉格朗日乘子, 我们称对应于 $\alpha_k \neq 0$ 的样本点为支持向量.

相应的 KKT 条件为

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^l \alpha_k y_k \varphi(x_k) \quad (5)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^l \alpha_k y_k = 0 \quad (6)$$

$$\frac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k, \quad k = 1, \dots, l \quad (7)$$

$$\frac{\partial L}{\partial \alpha_k} = 0 \rightarrow y_k [w^T \varphi(x_k) + b] - 1 + e_k = 0 \quad (8)$$

可以表示为下列方程组的形式

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \quad (9)$$

式中, $Z = [\varphi(x_1) y_1, \dots, \varphi(x_l) y_l]^T$, $Y = [y_1, \dots, y_l]^T$, $\vec{1} = [1, \dots, 1]^T$, $\alpha = [\alpha_1, \dots, \alpha_l]^T$.

令 $\Omega_{ij} = y_i y_j \varphi(x_i)^T \varphi(x_j) = y_i y_j k(x_i, x_j)$, 则方程(1~9)可以表示为

$$\begin{bmatrix} 0 & Y^T \\ Y & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \quad (10)$$

解方程(1~10)得到 α 和 b 后, 对于新的输入向量 x , 其类别可以根据下式进行判别

$$y(x) = \text{sign} \left[\sum_{k=1}^l \alpha_k y_k k(x, x_k) + b \right] \quad (11)$$

对于回归问题, 新的输入向量 x , 其输出值 $y(x)$ 可以根据下式计算

$$y(x) = \sum_{k=1}^l \alpha_k k(x, x_k) + b \quad (12)$$

3 最小二乘隐空间支持向量机

设训练集 T 由式(1)给出, 定义一个由实值函数集合 $\{\varphi_i(x) | i=1, 2, \dots, d_1\}$ 构成的向量

$$\varphi(x) = (\varphi_1(x), \varphi_2(x), \dots, \varphi_{d_1}(x))^T \quad (13)$$

则向量 $\varphi(x)$ 将 n 维输入空间上的点映射到一个 d_1 维空间上. 由于函数集合 $\{\varphi_i(x)\}$ 所起的作用类似于前向神经网络中的隐单元, 因而将 $\varphi_i(x)$ ($i=1, 2, \dots, d_1$) 称为隐函数, 相应的空间 $\{z | z = (\varphi_1(x), \varphi_2(x), \dots, \varphi_{d_1}(x))^T, x \in R^n\}$ 称为隐空间或特征空间.

考虑一种特殊的隐函数: 实对称核函数, 即 $k(x, x') = k(x', x)$, 令 $\varphi_i(x) = k(x_i, x)$, 则相应于输入空间的隐空间为

$$\{z | z = (k(x_1, x), k(x_2, x), \dots, k(x_l, x))^T, x \in R^n\} \quad (14)$$

在隐空间中构造最优分类超平面可以得到隐空间支持向量机^[9]

$$\min_{w, b, \xi} J(w, b, \xi) = \frac{1}{2} w^T w + C \sum_{k=1}^l \xi_k \quad (15)$$

s. t.

$$y_k [w^T z_k + b] \geq 1 - \xi_k \quad (16)$$

$$\xi_k \geq 0, \quad k = 1, \dots, l \quad (17)$$

由于 $w^T z = \sum_{i=1}^l \alpha_i y_i \sum_{j=1}^l k(x_j, x_i) k(x_j, x)$, 所以隐空间支持向量机在决策时将会用到所有的训练样本. 对于大规模的问题来说, 这是不现实的.

基于最小二乘的思想, 在隐空间中构造最优超平面, 文献[4]给出了最小二乘隐空间支持向量机.

$$\min_{w, b} \frac{1}{2\gamma} w^T w + \sum_{i=1}^l ((w^T \cdot z_i + b) - y_i)^2 \quad (18)$$

把式(14)代入式(18)得

$$\min_{w, b} \frac{1}{2\gamma} w^T w + \sum_{i=1}^l \left(\sum_{j=1}^l w_j \cdot k(x_i, x_j) + b - y_i \right)^2 \quad (19)$$

上述表达式可写为

$$\min_{w, b} \frac{1}{2\gamma} w^T w + (\Omega w + \vec{b}\vec{1} - Y)^T (\Omega w + \vec{b}\vec{1} - Y) \quad (20)$$

去掉常数项, 则式(20)等价于下列标准二次规划问题

$$\min_{\alpha} \frac{1}{2} \alpha^T H \alpha + d^T \alpha \quad (21)$$

其中

$$\alpha = (w_1, w_2, \dots, w_l, b)^T \quad (22)$$

$$H = \begin{bmatrix} \Omega^T \Omega + \frac{I}{2\gamma} & \Omega \vec{1} \\ \vec{1}^T \Omega & l \end{bmatrix} \quad (23)$$

$$d = \begin{bmatrix} -\Omega & Y \\ -Y^T & \vec{1} \end{bmatrix} \quad (24)$$

由于对任意的 $\alpha = (w_1, w_2, \dots, w_l, b)^T \neq 0$, 下式成立

$$\begin{aligned} \alpha^T H \alpha &= H \alpha^T \begin{bmatrix} \Omega^T \Omega + \frac{I}{2\gamma} & \Omega \vec{1} \\ \vec{1}^T \Omega & l \end{bmatrix} \alpha \\ &= \frac{1}{2\gamma} w^T w + (\Omega w + \vec{b} \vec{1})^T (\Omega w + \vec{b} \vec{1}) > 0 \end{aligned} \quad (25)$$

所以矩阵 H 是对称正定的, 从而式(21)是一个无约束凸二次规划, 可以用共轭梯度法进行求解。

4 改进的共轭梯度算法

共轭方向法的基本定理告诉我们, 共轭性和精确线性搜索产生二次终止性, 共轭梯度法就是使得最速下降方向具有共轭性, 从而提高算法的有效性和可靠性^[10]。共轭梯度算法的计算复杂性为 $O(kN^2)$, 这里 k 为迭代步数。对于大规模学习问题, 当对称正定方程组系数矩阵的条件数远大于 1 时, 共轭梯度算法迭代次数较多, 收敛将很慢。为减少迭代次数, 加快收敛速度, 可以设法先降低系数矩阵的条件数, 这就是预处理技术。预处理技术包括多项式预处理和对称超松弛预处理等技术, 本文采用基于对称超松弛预处理技术, 为说明该方法, 先给出两个定理。

定理 1 假设对称正定矩阵

$$H = \begin{bmatrix} \bar{H} & q \\ q^T & h_u \end{bmatrix} \quad (26)$$

其中 $\bar{H} \in R^{(l-1) \times (l-1)}$, $q \in R^{(l-1)}$, $h_u \in R$ 。那么矩阵

$$\tilde{H} = \bar{H} - \vec{1}_{l-1} \cdot q^T - q \cdot \vec{1}_{l-1}^T + h_u \cdot \vec{1}_{l-1} \cdot \vec{1}_{l-1}^T \quad (27)$$

是正定矩阵。

证明: 设

$$M = \begin{bmatrix} I_{l-1} & 0_{l-1} \\ -\vec{1}_{l-1}^T & 1 \end{bmatrix} \quad (28)$$

则有

$$M^T H M = \begin{bmatrix} I_{l-1} & -\vec{1}_{l-1} \\ 0_{l-1}^T & 1 \end{bmatrix} \begin{bmatrix} \bar{H} & q \\ q^T & h_u \end{bmatrix} \begin{bmatrix} I_{l-1} & 0_{l-1} \\ -\vec{1}_{l-1}^T & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{H} & \tilde{q} \\ \tilde{q}^T & h_u \end{bmatrix} \quad (29)$$

因为 H 是正定矩阵, 所以 $\begin{bmatrix} \tilde{H} & \tilde{q} \\ \tilde{q}^T & h_u \end{bmatrix}$ 也是正定矩阵,

从而 $\tilde{H}$ 也是正定矩阵。

定理 2 设 $\tilde{\alpha}^*$ 是线性代数方程组

$$\tilde{H} \tilde{\alpha} = \tilde{Y} - y_l \cdot \vec{1}_{l-1} \quad (30)$$

的解, 其中 $\tilde{Y} = (y_1, y_2, \dots, y_{l-1})^T$ 。则

$$\alpha^* = (\tilde{\alpha}^*, -\vec{1}_{l-1}^T \cdot \tilde{\alpha}^*)^T \quad (31)$$

$$b^* = y_l + h_u \cdot (\vec{1}_{l-1}^T \cdot \tilde{\alpha}^*) - q^T \cdot \tilde{\alpha}^* \quad (32)$$

是式(10)的解。

证明: 因为 $\tilde{H}$ 是正定的, 所以 $\tilde{\alpha}^*$ 是唯一的。由(27)和(30)得

$$\begin{aligned} & \tilde{H} \tilde{\alpha}^* - q \cdot \vec{1}_{l-1}^T \cdot \tilde{\alpha}^* - \tilde{Y} \\ &= (\tilde{H} + \vec{1}_{l-1} \cdot q^T + q \cdot \vec{1}_{l-1}^T - h_u \cdot \vec{1}_{l-1} \cdot \vec{1}_{l-1}^T) \\ & \quad \cdot \tilde{\alpha}^* - q \cdot \vec{1}_{l-1}^T \cdot \tilde{\alpha}^* - \tilde{Y} \\ &= \tilde{H} \cdot \tilde{\alpha}^* + \vec{1}_{l-1} \cdot q^T \cdot \tilde{\alpha}^* + q \cdot \vec{1}_{l-1}^T \cdot \tilde{\alpha}^* - h_u \\ & \quad \cdot \vec{1}_{l-1} \cdot \vec{1}_{l-1}^T \cdot \tilde{\alpha}^* - q \cdot \vec{1}_{l-1}^T \cdot \tilde{\alpha}^* - \tilde{Y} \\ &= -y_l \vec{1}_{l-1} + \vec{1}_{l-1} \cdot q^T \cdot \tilde{\alpha}^* - h_u \cdot \vec{1}_{l-1} \cdot \vec{1}_{l-1}^T \\ & \quad \cdot \tilde{\alpha}^* \\ &= (q^T \cdot \tilde{\alpha}^* - h_u \cdot (\vec{1}_{l-1}^T \cdot \tilde{\alpha}^*) - y_l) \cdot \vec{1}_{l-1} \\ &= -b^* \cdot \vec{1}_{l-1} \end{aligned} \quad (33)$$

把(32)和(33)写成矩阵形式, 可以得到

$$H \cdot \alpha^* + b^* \cdot \vec{1}_l = Y \quad (34)$$

其中 $Y = (y_1, y_2, \dots, y_{l-1}, y_l)^T$ 。由于 $\vec{1}_l^T \cdot \alpha^* = 0$, 所以 α^* 和 b^* 是式(10)的解。

从式(30)可知, 通过采用基于对称超松弛预处理技术对称正定方程组系数矩阵降低其条件数, 可得到改进的共轭梯度算法只需求解一个阶数为 $l-1$ 的线性代数方程组即可, 而共轭梯度算法需要求解两个阶数为 l 的线性代数方程组, 因此, 对于两种算法的求解的复杂性来说, 改进后的共轭梯度算法的求解复杂性明显降低, 从而节省了求解过程的计算时间。

最小二乘隐空间支持向量机实质上是求解线性方程组 $H\alpha = -d$, 当 H 为 n 阶对称正定阵时, 为得到处理后的方程组仍为正定阵, 设 $M = SS^T$, M 是对称正定阵, 则方程组等价于:

$$Fu = g, F = S^{-1}HS^{-T}, u = S^T\alpha, g = S^{-1}(-d) \quad (35)$$

通过这样的变换, 并选取相应的预处理阵, 可有效降低原系数矩阵 H 的条件数。改进的共轭梯度算法的步骤如下:

(1) 给定初值 $x^{(0)}$, $k=0$, 令 $r^{(0)} = Hx^{(0)} - d$.

(2) 解方程组 $Mz^{(0)} = r^{(0)}$, 求出 $z^{(0)}$, 令 $p^{(0)} = z^{(0)}$.

(3) 当 $k=0, 1, 2, \dots$ 时

$$\alpha^{(k)} = -\frac{r^{(k)} p^{(k)}}{p^{(k)} H p^{(k)}} \quad (36)$$

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)} \quad (37)$$

$$r^{(k+1)} = r^{(k)} + \alpha^{(k)} H p^{(k)} \quad (38)$$

$$z^{(k+1)} = M^{-1} r^{(k+1)} \quad (39)$$

$$\beta^{(k)} = \frac{r^{(k+1)} r^{(k+1)}}{r^{(k)} r^{(k)}} \quad (40)$$

$$p^{(k+1)} = -r^{(k+1)} + \beta^{(k)} p^{(k)} \quad (41)$$

(4) 直到 $\|r^{(k+1)}\| < \epsilon$, 输出 $\alpha^{(k+1)}$; 否则, 转到 (3).

5 实例分析

股票价格是一种影响因素庞杂的非线性时间序列, 普通方法的预测结果往往不尽人意. 而支持向量机在股价预测方面拥有相当优异的预测效果. 本文将改进后的算法模型应用于股票价格的预测, 检验了模型的优越性.

6 结束语

文中采用基于对称超松弛预处理技术改进共轭梯度算法, 改进的共轭梯度算法只需求解一个阶数为 $l-1$ 的线性代数方程组即可, 节省了计算时间. 最后将其应用于最小二乘隐空间支持向量机中建立数学模型, 将其应用于股票价格的短期预测中, 得到了良好的学习性能和预测效果.

参考文献:

[1] Vapnik V N. Statistical learning theory[M]. New

York: Wiley, 1998.

- [2] Suykens J, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999(3):293-300.
- [3] Zhang Li, Zhou Weida, Jiao Licheng. Hidden space support vector machines[J]. IEEE Transactions on Neural Networks, 2004, 15(6):1424-1434.
- [4] 王玲, 薄列峰. 最小二乘隐空间支持向量机[J]. 计算机学报, 2005, 28(8):1302-1307.
- [5] Zhe Wang, Songcan Chen. New least squares support vector machines based on matrix patterns[J]. Neural Processing Letters, 2007(26):41-56.
- [6] 何昊. 基于最小二乘支持向量机的 Jakes 衰落信道预测[J]. 微电子学与计算机, 2010, 27(7):222-224.
- [7] 于重重, 王静燕, 谭丽, 等. LS-SVM 在桥梁结构健康预测评估中的应用[J]. 微电子学与计算机, 2011, 28(6):148-152.
- [8] Suykens J, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3):293-300.
- [9] Zhang Li, Zhou Weida, Jiao Licheng. Hidden space support vector machines[J]. IEEE Transactions on Neural Networks, 2004, 15(6):1424-1434.
- [10] Hestenes M R, Stiefel E. Method of conjugate gradient for solving linear system[J]. J Res Nat Bur Stand, 1952(49):409-436.

作者简介:

王 玮 男, (1985—), 研究生. 研究方向为计算机理论与体系结构.

娄 莉 女, (1970—), 硕士, 副教授. 研究方向为计算机理论与通信工程.