

支持向量机增量学习中模型参数选择问题研究

张 鹏, 倪世宏, 谢 川

(空军工程大学工程学院, 陕西 西安 710038)

摘要 支持向量机性能主要受模型参数的影响, 而支持向量机增量学习中模型参数选择问题研究较少。针对这一问题, 提出一种支持向量机增量学习中模型参数选择方法。将鲁棒度作为增量学习的性能估计准则, 用拟合误差和比例系数调节解空间取值范围, 采用梯度下降法搜索参数, 用初始模型参数作为梯度下降法的初始值。用该方法对 Logistic 模型和航空发动机振动监控进行实验。结果表明: 与基本遗传算法和梯度法进行比较, 所提方法能充分利用历史学习的结果, 缩小解空间的搜索范围, 加快收敛速度。

关键词 支持向量机; 增量学习; 模型参数选择; 鲁棒度; 拟合误差; 梯度下降法

DOI 10.3969/j.issn.1009-3516.2011.05.002

中图分类号 TP391 **文献标识码** A **文章编号** 1009-3516(2011)05-0005-05

支持向量机(Support Vector Machines, SVM)^[1]是在统计学习理论的基础上发展起来的新一代机器学习算法, 该算法目前在模式识别、故障诊断和时序预测等领域得到了广泛的应用。Vapnik 等人的研究表明, 对于给定的样本, SVM 的性能主要受核参数和系统参数的影响^[2]。选择最优的核参数及系统参数被称为模型参数的选择。

实际中, 样本的采集都是不断积累的, 当有新的样本加入时, 要求对新的样本进行再学习以提高精度; 而一些渐变问题(如机械设备的早期故障期和损耗期)中, 增量样本集所提供的信息量与历史样本集所提供的不同, 需要重新构建 SVM 模型, 这就是增量学习问题^[3]。传统 SVM 参数选择方法主要集中在样本集筛选, 不能有效利用历史学习的结果, 时间代价较高^[4-5]。本文针对 SVM 增量学习中样本集更新后模型参数选择问题, 对性能估计准则、解空间的优化设置和具体实现步骤进行了研究, 提出一种新的参数搜索方法。通过实验验证了所提方法的可行性和有效性。

1 SVM 增量学习及参数选择问题

SVM 的基本思想是通过一个非线性映射 $\varphi(\cdot)$ 将数据 x 映射到高维特征空间, 然后寻找 $\varphi(x)$ 的一个线性函数:

$$f(x) = \omega^T \varphi(x) + b \quad (1)$$

式中: $f(x)$ 为 x 的非线性函数, 同时又是 $\varphi(x)$ 的线性函数; ω 为权值矢量; b 为阈值。

给定训练集 $\{x_i, y_i\}_n$, 根据结构风险最小化准则, $f(x)$ 应使得以下风险函数最小, 式中 L 是惩罚函数, C 为平衡因子:

$$\min J = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n L(f(x_i), y_i) \quad (2)$$

通过采用对偶原理、拉格朗日乘子法及核函数技术, 求解式(2)得分类决策函数:

* 收稿日期: 2011-01-12

基金项目: 国家“863”计划资助项目(2010AA8090514-C)

作者简介: 张 鹏(1982-), 男, 四川成都人, 博士生, 主要从事飞行器故障诊断与健康监控研究。

E-mail: zhangpeng25@gmail.com

$$f(x) = \operatorname{sgn}\left[\sum_{i=1}^n a_i y_i K(x_i, x) + b\right] \quad (3)$$

若是回归问题,可得回归估计函数:

$$f(x) = \sum_{i=1}^n (a_i^* - a_i) K(x_i, x) + b \quad (4)$$

式中: K 为核函数; a_i 与 a_i^* 为拉格朗日常数。 $a_i \neq 0$ 时对应的 x_i 被称为支持向量(Support Vector, SV)。在大多数情况下,支持向量只占训练样本集的一小部分。

当有增量训练样本加入时,原样本中的支持向量和非支持向量会发生变化,需将增量本和初始样本合并成新的训练集,重新求解式(2)的优化问题^[6]。在实际工程中,这个总样本集会越来越大,导致SVM的计算出现困难。

SVM增量学习基本流程可简述如下^[7]:设训练得到初始分类函数为 f_1 ,根据 a_i 与 C 的关系将支持向量集分为边界支持向量集 M_B 和非边界支持向量集 M_U ;用 f_1 对增量训练样本 $\{x'_k, y'_k\}_m$ 进行测试,得到测试错误集 M_{err} 和测试正确集 M_{ok} 。若 $M_{\text{err}} \neq \emptyset$,将 $M_{\text{err}} \cup M_B$ 作为新的训练集 T_{test} ,将 $M_{\text{ok}} \cup M_U$ 作为新的增量样本集 T_{add} 。重复以上步骤直到训练集收敛。

2 SVM增量学习中模型参数选择方法

支持向量机模型参数选择问题就是通过设置合适的SVM模型参数(如平衡因子 C),提高SVM的性能。根据SVM增量学习的特点,模型参数选择需要解决以下3方面的问题:①SVM增量学习中性能估计准则的选择;②利用历史学习的结果,缩小模型参数搜索范围;③权衡效率和精度,选择合适的搜索方法。

2.1 基于鲁棒度的SVM性能估计准则

SVM增量学习是一个不断加入新样本的学习过程,在有限样本情况下更需要考虑其推广性,即泛化性能。传统的模型参数方法如留一法、交叉验证法等,都基于最大间隔原理,要求拥有完整的训练集,并不能保证泛化性能。因此本文采用鲁棒度作为SVM增量学习时的性能估计标准,下面给出定义。

式(1)中函数 $f(x)$ 在点 x_k 处的抗扰动能力^[8]可表达为:

$$R(f, x_k) = |f(x_k)| \cdot \|(\nabla f(x_k))\|^{-1} \quad (5)$$

式中 $\nabla f(x_k)$ 是 $f(x)$ 在点 x_k 处的梯度, $\|(\nabla f(x_k))\|$ 刻画了 $f(x)$ 在点 x_k 处的变化率。

将函数 $f(x)$ 的鲁棒度定义为在边界支持向量集上 $R(f, x_k)$ 的平均值, $\text{sum}(M_B)$ 为边界支持向量个数:

$$R(f) = \frac{1}{\text{sum}(M_B)} \sum_{x_k \in M_B} R(f, x_k) \quad (6)$$

鲁棒度 $R(f)$ 刻画了所得函数对样本集的推广能力。鲁棒度越大,模型对未来输出进行正确预测的能力越强。此时,SVM的参数选择问题可转化为求鲁棒度最大的优化问题。

2.2 设置模型参数取值范围

SVM模型参数的选取直接影响SVM的性能。在增量学习中,由于样本集发生变化,原来经过优化的最优参数不再适合新的训练集。大量实验发现,若初始模型对新训练集测试的误差较小,则新训练集对应的SV与原SV的区别也小,最优参数值的变化也小。随着新训练集与初始训练集的差距增大,新的最优参数与初始最优参数之间的差距也随之增大。根据上面的分析,本文采用测试误差均值来衡量新训练集与初始训练集的差距,用于缩小模型参数的搜索范围,减少计算量。

以SVM回归为例,设初始训练样本集为 $\{x_i, y_i\}_n$,经过训练得到的回归函数见式(4)。设增量样本为 $\{x'_k, y'_k\}_m$,输入 x'_k 可得回归值:

$$\tilde{y}_k = \sum_{i=1}^n (a_i^* - a_i) K(x_i, x'_k) + b \quad (7)$$

增量样本的回归误差均值可表示为:

$$E_{\text{reg}} = \frac{1}{m} \sum_{k=1}^m \frac{|\tilde{y}_k - y'_k|}{|y'_k|_{\max}} \quad (8)$$

设要优化的参数为 η ,设置 η 的可行取值范围为 $[\eta_{\min}, \eta_{\max}]$,设初始模型经过优化得到的最优参数为

η_{best1} , 则增量学习时 η 的值域可取为 $[\eta'_{min}, \eta'_{max}]$, 其中: $\eta'_{min} = \max\left(\eta_{best1} - \frac{(\eta_{max} - \eta_{min})E_{reg}}{2}\beta, \eta_{min}\right)$, $\eta'_{max} = \min\left(\eta_{best1} + \frac{(\eta_{max} - \eta_{min})E_{reg}}{2}\beta, \eta_{max}\right)$, 式中 β 为比例系数, 用于调节效率和精度之间的关系。可认为增量学习的最优参数 η_{best2} 以相当大概率包含于 $[\eta'_{min}, \eta'_{max}]$ 中。

2.3 梯度下降法用于模型参数选择

梯度下降法(Gradient descent algorithm, GDA)是一种快速的局部寻优算法, 根据连续可微的显式解析表达式计算下一次迭代点, 搜索具有方向性, 局部搜索效率高。但参数初始值对 GDA 的影响非常大^[9], 该方法用于 SVM 寻参时容易陷入局部极值^[10]。可用初始模型最优参数作为 GDA 的参数初始值, 提高该方法的搜寻效率。

梯度下降法用于 SVM 增量学习模型参数选择的具体实现步骤如下:

Step1 初始化, 确定调整步长 Δ 、最大循环次数 t_{max} 和收敛精度 ε , 令当前迭代数 $t=0$, 令梯度下降法的参数初始值 η_0 等于初始模型最优参数 η_{best1} , 用 SVM 训练增量样本, 得到 η_0 对应的鲁棒度 R_0 , 并令 $\eta_{best} = \eta_0$, $R_{best} = R_0$;

Step2 计算 $\eta_{t+1} = \eta_t + \nabla R_t |_{\eta_t} \Delta$, 其中有:

$$\nabla R_t |_{\eta_t} = \begin{cases} 1 & , t > 0 \text{ 时} \\ \frac{R_t - R_{t-1}}{\eta_t - \eta_{t-1}} & , t = 0 \text{ 时} \end{cases} \quad (9)$$

并求出 η_{t+1} 对应的鲁棒度 R_{t+1} ;

Step3 若 $R_{t+1} > R_{best}$, 说明搜索到更优点, 则令 $\eta_{best} = \eta_{t+1}$, $R_{best} = R_{t+1}$;

Step4 若满足收敛条件 $|R_{best} - R_t| < \varepsilon$, ($i = t, t-1, \dots, t-L$), 中止迭代, 令增量样本最优参数 $\eta_{best2} = \eta_{best}$, 否则下一步;

Step5 若 $t = t_{max}$, 中止迭代, 否则 $t = t + 1$ 并转 Step2。

3 实验分析

3.1 Logistic 序列仿真实验

为了检验本文所提方法的有效性, 下面以 Logistic 系统为例, 用最小二乘支持向量回归机 LS-SVR 进行仿真验证。

Logistic 系统定义为 $x_{n+1} = ax_n(1 - x_n)$, 取初始值 $x_0 = 0.05$, 初始状态参数 $a = 3.7$, 训练样本集为 $n = [501, 600]$, 选用高斯核函数训练最小二乘支持向量机, 用遗传算法对参数进行优化, 得到初始模型最优参数为 $\gamma = 2.80, \sigma = 0.024$ 。

用参数 a 的变化模拟 Logistic 系统的渐变行为, 设 a 从 3.7 逐步变为 3.9, 幅度为 0.01, 去掉其中 3 个循环序列, 共 17 个样本集, 令步长 $\Delta_\gamma = 0.01, \Delta_\sigma = 0.001$, 最大迭代次数 $t_{max} = 200$ 。将本文所提方法在不同比例系数 β 的情况下与遗传算法、梯度法进行比较, 其中遗传算法采用 20 位二进制编码, 种群大小为 20, 交叉概率 0.3, 变异概率 0.08, 梯度法的参数初始值从全局值域中随机获得, 步长等参数与本文所提方法相同, 17 次计算求平均的实验结果见表 1。

表 1 实验结果比较

Tab.1 Compare of simulating result

算法	平均有效迭代次数	平均时间/s	鲁棒度偏差 (%)	最优解次数
本文算法 $\beta=0.1$	11.1	21.8	0.025 3	16
本文算法 $\beta=1$	26.95	38.6	0	17
遗传算法	408	731	0	17
梯度法	124	185	0.161 0	13

表 1 中平均有效迭代次数为正确搜索到最优参数的迭代次数, 未搜索最优解到或搜索到局部最优解的迭代次数不计入其中。平均时间为平均有效迭代次数所耗时间。鲁棒度偏差为 17 次搜索的总鲁棒度偏差。

最优解次数为 17 次搜索中正确搜索到最优参数的次数。

从表 1 中可以看出,由于本文所提方法合理利用历史学习的结果选择参数初始值和取值范围,用较少的迭代次数找到最优解,复杂度低,精度高,适用于 SVM 增量学习模型参数选择。该算法全局搜索能力偏弱,有可能陷入局部极值,可以通过调节比例系数 β 改善搜索能力。

3.2 航空发动机振动监控仿真实验

下面以航空发动机振动监控^[11]为例对本文所提方法进行验证。

航空发动机振动监控的原理为:首先建立同一工作状态下发动机的振动值与其它性能参数之间的变化关系模型,然后连续监控发动机振动值随飞行架次的变化趋势。具体实现过程如下:以发动机装机后前几个架次的起飞振动值作为训练样本,用 C-SVR 建立发动机振动模型,对后期的振动数据进行拟合分析,若拟合误差较大则认为发动机振动异常。实际中,经过一段时间的使用,发动机内部机械结构磨损、各气动部件效率降低等因素,使得各参数间的协调关系逐渐偏离初始值,从而实际模型渐渐偏离初始模型,为了提高监控的精度,需要对模型进行更新。

初始训练样本选择前 5 个架次起飞振动数据,采用 C-SVR 算法建立支持向量机模型,用交叉验证法优化得到最优参数为 $C = 617.43, \sigma = 0.284$ 。实验数据选择第 21-25 架次的总数据作为增量学习样本,根据式(7)对增量学习样本回归,根据公式(8)的拟合误差为 0.252 3。采用本文所提方法进行增量学习参数选择,取步长 $\Delta_c = 0.01, \Delta_\sigma = 0.001$,比例系数 $\beta = 1$ 。为观察参数优化过程,在搜索空间上运用穷举法求出鲁棒度曲面,经 23 次迭代后得到最优参数为 $C' = 323.11, \sigma' = 0.239$,鲁棒度增加了 3.6%,优化过程见图 1。

为更好的分析本文所提方法的性能,用基本遗传算法进行对比实验,遗传算法的参数设置和本文所提方法的参数设置与 Logistic 序列仿真实验一致。选择第 6-25 架次中每一个架次数据作为增量学习样本,对每个架次的样本进行参数选择,20 次实验参数优化过程中的平均最优解与迭代次数的关系见图 2。

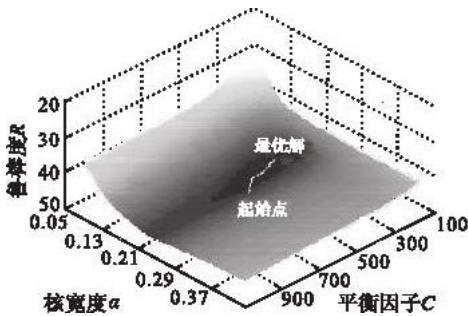


图 1 参数优化过程

Fig. 1 Process of optimizing parameters

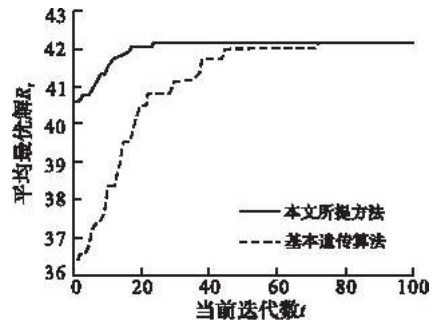


图 2 平均收敛速度

Fig. 2 Average convergent rate

从图 1、图 2 可以看出本文所提方法比基本遗传算法的收敛速度快很多。原因为:遗传算法的初始群体在整个解空间中随机生成,搜索效率较低。而梯度下降法在限制解空间大小和设置初始值后,在不损失精度的情况下收敛速度加快,具有很好的工程应用价值。

4 结束语

针对当前支持向量机模型参数选择方法用于增量学习时存在的问题,本文提出一种利用历史学习结果的参数选择方法,并给出了具体步骤,最后通过仿真实验验证了本文所提方法的可行性和有效性。该方法物理意义明确,易于计算机实现,能充分利用历史学习的结果,有效减少了选择参数的时间,在保障精度的前提下减少迭代次数,适合于解决 SVM 增量学习模型参数选择问题。

参考文献:

- [1] Vapnik V N. An overview of statistical learning theory[J]. IEEE transactions on neural networks, 1999, 10(5):988-999.
 - [2] Vapnik V N. 统计学习理论[M]. 许建华,张学工,译.北京:电子工业出版社,2004:181-223.
- Vapnik V N. The nature of statistical learning theory[M]. XU Jianhua, ZHANG Xuegong, translate. Beijing: Publishing

- house of electronics industry, 2004;181 – 223. (in Chinese)
- [3] 李凯,黄厚宽. 支持向量机增量学习算法研究[J]. 北方交通大学学报, 2003,27(5):34 – 37.
LI Kai, HUANG Houkuan. Research on incremental learning algorithm of support vector machine[J]. Journal of northern jiao tong university, 2003,27(5):34 – 37. (in Chinese)
- [4] 曹冲锋,杨世锡,周晓峰,等. 改进支持向量机模型的旋转机械故障诊断方法[J]. 振动、测试与诊断, 2009,28(3):270 – 273,367.
CAO Chongfeng, YANG Shixi, ZHOU Xiaofeng, et al. Fault diagnosis of rotating machinery based on improved support vector machines model[J]. Journal of vibration, measurement & diagnosis, 2009,28(3):270 – 273,367. (in Chinese)
- [5] 李本威,张赞,孙涛. 基于免疫粒子群算法的滑油屑末支持向量机预测模型设计[J]. 航空动力学报, 2009, 24(7):1639 – 1643.
LI Benwei, ZHANG Yun, SUN Tao. Design of forecasting model for aero – engine lubrication debris support vector machines based on immune – particle swarm optimization algorithm[J]. Journal of aerospace power, 2009, 24(7):1639 – 1643. (in Chinese)
- [6] 周伟达,张莉,焦李成. 支持向量机推广能力分析[J]. 电子学报, 2001,29(5):590 – 594.
ZHOU Weida, ZHANG Li, JIAO Licheng. An analysis of SVMs generalization performance[J]. Acta electronica sinica, 2001, 29(5):590 – 594. (in Chinese)
- [7] 萧嵘,王继成,孙正兴. 一种 SVM 增量学习算法 α – ISVM[J]. 软件学报, 2001,12(12):1818 – 1824.
XIAO Rong, WANG Jicheng, SUN Zhengxing. An incremental SVM learning algorithm α – ISVM[J]. Journal of software, 2001,12(12):1818 – 1824. (in Chinese)
- [8] 罗林开. 支持向量机的核选择[D]. 厦门:厦门大学,2007.
LUO Linkai. Research on kernel selection of support vector machine[D]. Xiamen:Xiamen university, 2007. (in Chinese)
- [9] Keerthi S. Efficient tuning of SVM hyper parameters using radius/margin bound and iterative algorithms [J]. IEEE trans neural networks,2002,13(5):1225 – 1229.
- [10] 董春曦,饶鲜,杨邵全. 支持向量机参数选择方法研究[J]. 系统工程与电子技术,2004,26(8):1117 – 1120.
DONG Chunxi, RAO Xian, YANG Shaoquan. Method for selecting the parameters of support vector machines[J]. Systems engineering and electronics,2004,26(8):1117 – 1120. (in Chinese)
- [11] 倪世宏,史忠科,沙孟春. 一种基于支持向量回归的发动机振动监控方法[J]. 空军工程大学学报:自然科学版, 2006, 7(5):1 – 3.
NI Shihong, SHI Zhongke, SHA Mengchun. A monitoring method of aero – engine vibration based on the SVR[J]. Journal of air force engineering university:natural science edition, 2006, 7(5):1 – 3. (in Chinese)

(编辑:徐敏)

Parameter Selection of Support Vector Machine Based Incremental Learning Method

ZHANG Peng, NI Shi – hong, XIE Chuan

(Engineering Institute, Air Force Engineering University, Xi'an 710038, China)

Abstract: The performance of Support vector machine (SVM) is affected mainly by model parameters, but there is no special method for the model parameter selection of SVM based incremental learning. A new method is proposed in this paper, i. e. taking robustness as criteria for performance evaluation of incremented learning, the range of solution space is designed by fitting error and scale factor, then the gradient descent algorithm is used to search the parameters. The experiments with this new method are made on the Logistic model regressing and aero engine vibration monitoring, and the comparison of this new method with the genetic algorithm and the gradient descent algorithm is made. The result indicates that the use of the proposed method can take full advantage of the results of historical learning, thus the solution space is narrowed, and iteration steps are reduced.

Key words: support vector machine; incremental learning; model parameter selection; robustness; fitting error; gradient descent algorithm