

一种改进的序贯最小优化算法

骆世广¹ 杨晓伟¹ 吴广潮¹ 张新华²

(华南理工大学数学科学学院 广州 510640)¹ (华中师范大学计算机科学系 武汉 430079)²

摘要 序贯最小优化(SMO)算法是目前解决支持向量机训练问题的一种十分有效的方法,但是当面对大样本数据时,SMO训练速度比较慢。本文分析了SMO迭代过程中目标函数值的变化情况,进而提出以目标函数值的改变量作为算法终止的判定条件。几个著名的数据集的试验结果表明,该方法可以大大缩短SMO的训练时间,特别适用于大样本数据。

关键词 支持向量机,序贯最小优化算法

An Improved Sequential Minimal Optimization Algorithm

LUO Shi-Guang¹ YANG Xiao-Wei¹ WU Guang-Chao¹ ZHANG Xin-Hua²

(School of Mathematical Science, South China University of Technology, Guangzhou 510640)¹

(Department of Computer Science, Central China Normal University, Wuhan 430079)²

Abstract At present sequential minimal optimization (SMO) algorithm is a very efficient method for training support vector machines (SVM). However, the training speed of SMO is very slow for the large-scale datasets. Analyzing the varieties of the objective function in SMO iterations, we propose a novel improved SMO algorithm in this paper, where the changed value of the objective function is taken as the termination condition. Experiments on several benchmark datasets have been done and the results show that the training time of the proposed algorithm is reduced greatly, especially for the large-scale problems.

Keywords Support vector machine, Sequential minimal optimization algorithm

20世纪60年代末,Vapnik和Cervonenkis建立了现代意义上的统计学习理论,即目前人们所称的VC维理论。1979年,在统计学习理论的VC维理论和结构风险最小化(SRM)准则的基础上,Vapnik等人提出了一种新的机器学习算法——支持向量机(SVM)方法。由于其坚实的理论基础、良好的泛化性能、简洁的数学形式、直观的几何解释等特点,它在许多实际问题的应用中取得了成功。

支持向量机可以归结为解决一个二次规划问题^[1]:

给定输入空间中训练样本:

$$\{(x_i, y_i) | i=1, \dots, l; x_i \in R^n, y_i \in \{-1, +1\}\} \quad (1)$$

SVM是要得到下面的决策函数:

$$f(x) = \text{sign} \left\{ \sum_{i=1}^l \alpha_i y_i k(x_i, x) - b \right\} \quad (2)$$

其中 $k(x_i, x_j)$ 是核函数, α_i 是每个样本对应的Lagrange乘子, b 是阈值。 $\{\alpha_i\}_{i=1}^l$ 是下面QP(Quadratic Programming)问题的解:

$$\begin{aligned} \max_{\alpha} W(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j Q(i, j) \\ \text{s. t. } \sum_{j=1}^l \alpha_j y_j &= 0 \\ 0 \leq \alpha_i &\leq C, i=1, \dots, l \end{aligned} \quad (3)$$

其中 $Q(i, j) = y_i y_j k(x_i, x_j)$ 是一个 $l \times l$ 的半正定矩阵, C 是正则化参数。

在以前的研究中,Vapnik本人提出了Chunking^[2]算法,Osuna等人^[3]提出了固定工作集的分解算法。基于Osuna的分解思想,Joachim^[4]用软件SVM^{light}实现了分解算法。Platt^[5]提出了SMO算法,该算法把大的二次规划问题转化为一系列最小的二次规划子问题,从而可以解析求解。

Keerthi等人^[6]在判别最优条件时用双阈值代替单阈值,在保证算法收敛的情况下,减少了迭代次数。孙剑等人^[7]利用SMO迭代后期到达上界的Lagrange乘子(记为BSV)变化平稳的特性,提出了一种基于决策函数增量的缓存策略,该方法可以有效地减少核函数的计算和训练时间。但如何判断BSV何时变化稳定,严格依赖于训练数据集。基于工作集的选择,李建民等人^[8]提出了一种收益代价平衡的工作集选择算法。该方法可以提高SMO算法的性能,特别适用于样本较多、支持向量较多、非有界支持向量较多的情况;但是平衡参数 $coef$ 的确定没有一定的准则。基于最大梯度法,张浩然等人^[9]提出了新的变量选择策略。该策略使得所选取的优化变量能够使目标函数的下降最多,优化步长更大。La等人^[10]认为SMO的分解策略并不能保证迭代的每一步都是朝着目标函数的最优点靠近的,从而提出了基于外推法的SMO改进算法。至于SMO的收敛性,Keerthi^[11]和Lin^[12]做了重要的工作。

上述SMO算法的重点主要集中在SMO的核函数的缓存上和寻找最大步长变量上,并没有从根本上解决SMO后期进展缓慢的瓶颈。本文在详细分析SMO迭代后期情况的基础上,提出了一种基于最大目标差的算法终止条件。试验结果表明,在几乎不损失精度的情况下,训练速度有很大的提高。

1 标准SMO算法

为了解决大规模分类问题,基于Osuna的分解思想,Platt^[5]提出了SMO算法。它把子问题规模限制到最小,每次优化两个拉格朗日乘子(不妨记为 α_1, α_2),这样子问题就可以由解析的方法解出。Platt为了加快SMO的训练速度,用

杨晓伟 博士,副教授,主要研究方向为支持向量机、优化计算与数据挖掘、区间分析和结构重分析。

了启发式的工作集选择策略,由两重循环完成:首先外层循环在所有样本中(examineAll=1)选择一个违反 KKT 条件的样本 α_1 ,然后进入内层循环,选择另外一个乘子和它配对优化。内层循环首先在所有样本中找能和 α_1 引起最大优化步长的 α_2 ,实际上这一步比较粗略,有可能找到的 α_2 不能使目标函数做进一步的调整;这时,算法遍历所有具有非边界拉格朗日乘子的样本(非边界样本),继续寻找能与 α_1 配对优化的 α_2 。如果这样的样本仍然找不到,就遍历所有样本。这两次遍历都是从随机位置开始的,以免算法每一次开始都是向固定的方向寻找。如果在极端退化的情况下,找不到能引起优化进展的 α_2 ,这时就放弃第一个样本。外层循环总是在 examineAll=1 和 examineAll=0 之间交替进行。直到在所有的样本中找不到可以优化的乘子时,算法结束。除了第一次之外,优先遍历非边界样本,因为非边界样本是最有可能需要调整的,而边界样本在优化稳定之后绝大部分都会停留在边界上,如图 1(其中,一轮即是一次 examineAll=1 或 examineAll=0)。

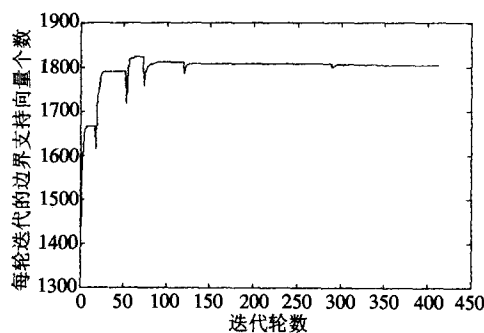


图 1 边界支持向量趋势图

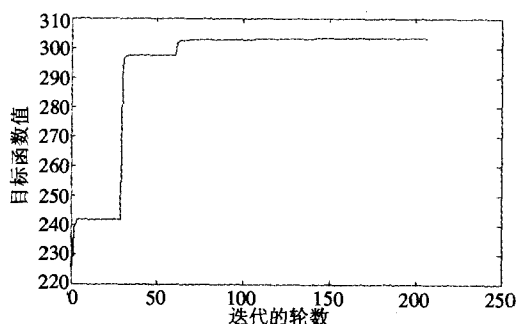


图 2 目标函数值趋势图

Paltt 把子问题规模变为最小,避免了子问题求解的数值算法和由此引起的误差积累,不受计算机内存的限制,使支持向量机应用于大规模学习问题可行。但是子问题的规模和迭代的次数是一对矛盾,SMO 将工作样本集的规模减少到 2,一个直接的后果就是迭代次数的增加,所以 SMO 实际上是将求解子问题的耗费转嫁到迭代上,当问题规模很大时,就需要大量的迭代运算。因此,SMO 除了在处理具有稀疏二进制特点的数据时训练速度较快之外,对一般数据是非常慢的。SMO 的训练时间实际上是由核函数的计算支配着的,核函数的计算实际上就是样本的内积计算。由于 SMO 选择样本集的两层循环机制,实际上访问样本的代价比穷举还大得多,由此使得内层循环优先选择引起最大步长的乘子效果减弱。这是因为 SMO 实际上是一个爬山的过程,刚开始时很快,当接近最优解时,解空间已经变得很小。但是 SMO 还是用原来

的工作机制选择工作集,使得刚开始的步步最优失去了一定的意义。试验中我们观察到,刚开始目标函数上升很快,后来变慢,而且越到后来越慢,甚至出现不变情况,如图 2。

2 改进的 SMO 算法

针对 SMO 迭代后期进展缓慢的特点,我们改进了 SMO 的停止条件。由图 2 可知,如果以目标函数改变量来确定算法终止,可以节省很多时间。但是每轮循环完计算一次(3)式,计算代价很大,试验中发现得不偿失。考虑到 SMO 每次只涉及到改变(3)式的两个乘子,我们直接计算在一次优化完后(3)式的改变量(记为 ΔW)为(不妨记本次优化中 α_{i1}, α_{i2} 变为 $\alpha_{i1}^{new}, \alpha_{i2}^{new}$)

$$\begin{aligned} \Delta W = & \Delta\alpha_{i1} + \Delta\alpha_{i2} \\ & - \Delta\alpha_{i1} y_{i1} \sum_{j \neq i1, i2} \alpha_j y_j k(x_{i1}, x_j) \\ & - \Delta\alpha_{i2} y_{i2} \sum_{j \neq i1, i2} \alpha_j y_j k(x_{i2}, x_j) \\ & - (\alpha_{i1}^{new} \alpha_{i2}^{new} - \alpha_{i1} \alpha_{i2}) \cdot y_{i1} y_{i2} k(x_{i1}, x_{i2}) \\ & - \frac{1}{2} [(\alpha_{i1}^{new} \alpha_{i1}^{new} - \alpha_{i1} \alpha_{i1}) \cdot y_{i1} y_{i1} k(x_{i1}, x_{i1}) \\ & + (\alpha_{i2}^{new} \alpha_{i2}^{new} - \alpha_{i2} \alpha_{i2}) \cdot y_{i2} y_{i2} k(x_{i2}, x_{i2})] \end{aligned}$$

其中

$$\Delta\alpha_{i1} = \alpha_{i1}^{new} - \alpha_{i1}, \Delta\alpha_{i2} = \alpha_{i2}^{new} - \alpha_{i2}。$$

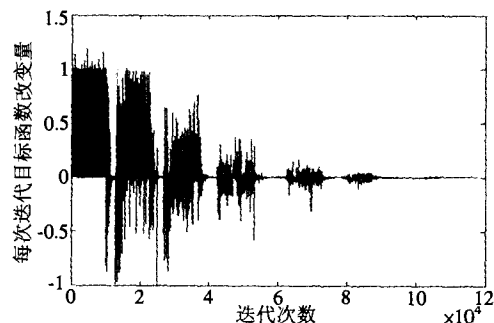


图 3 ΔW 变化趋势图

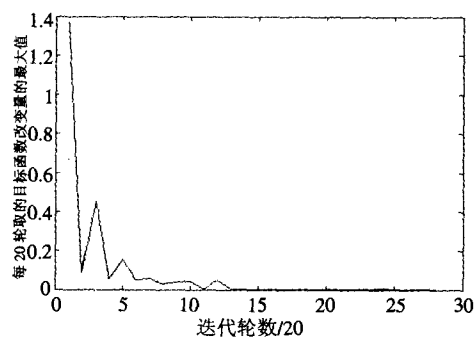


图 4 $\Delta\Delta W_{max}$ 变化趋势图(20 轮)

在整个优化过程中, ΔW 变化如图 3 所示。由图 3 可看出,后期趋于平稳,但不是很明显,不易确定算法何时终止。所以我们在 SMO 的每一轮优化中取出最大的那个 ΔW , 记为 ΔW_{max} 。在这些 ΔW_{max} 中,每几轮取出一个这几轮的最大值(试验中我们发现取 20 轮时试验效果比较好),记为 $\Delta\Delta W_{max}$, 它的变化趋势如图 4 所示。当 $\Delta\Delta W_{max}$ 小于一个设定的阈值(在我们的试验中,这个阈值取 0.001)后,我们将判定违反 KKT 条件的阈值扩大 10 倍,这样就越过一些违反 KKT 条件但是几乎不能引起优化进展的样本。这样 SMO 就在迭代后期,随着解空间的缩小,相应地收缩了工作

集搜索空间,减少了 SMO 的训练时间。

3 试验结果及分析

为了说明新算法的高效性,我们用 VC++ 6.0 编写了相关程序,并在内存为 256MB、CPU 为 1.8GHz 的 PC 机上测试了 UCI 的 6 个数据集和著名的 Checkerboard 问题。在我们的实验中,核函数是 RBF 函数, $C=0.5$ 。对于不同的数据集,参数 σ 的设置见表 1。标准的 SMO(P-SMO)和改进的 SMO(I-SMO)所得到的结果列在表 2 中。

从表 2 中可以看出,I-SMO 比 P-SMO 缩短了至少一半的训练时间,尤其是对大样本数据集而言,效果更显著,而且算法的训练精度和测试精度都没有大的损失。这说明标准的 SMO 在优化后期调整分类面时,摆动是比较厉害的。由于

SMO 的工作集选择策略,使得刚开始算法进展得非常迅速。而当找到绝大部分的支持向量后,解空间已经缩小很多了,影响分类面的样本已经非常少了,此时 SMO 应当缩小搜索空间,否则增加花费。I-SMO 通过迭代后期增加违反 KKT 条件的判定阈值,来缩小搜索空间,达到了提高 SMO 性能的目的。

表 1 各个数据集中使用的参数 σ

σ	Data Set
0.1	Checkerboard
1.0	Tic-Tac-Toe, KRKPA7, KRK
5.0	Letter Image Recognition
10.0	Image Segmentation
15.0	Sat Image

表 2 P-SMO 和 I-SMO 的结果比较

Dataset ($l+p$) $\times n$	Algorithm	Training Correctness (%)	Testing Correctness (%)	Time (secs)	# of SVs
Tic-Tac-Toe (900+58) $\times 9$	P-SMO	100	100	27.525	525
	I-SMO	100	100	11.874	400
Segmentation (2200+109) $\times 19$	P-SMO	98.25	95.18	587.75	1671
	I-SMO	98.543	96.3	234.3	1425
KRKPA7 (3000+196) $\times 36$	P-SMO	99.355	98.469	1476.67	1663
	I-SMO	99.4	98.725	506	1206
Sat Image (4435+2000) $\times 36$	P-SMO	99.39	93.15	3820	2369
	I-SMO	99.293	93.3	1563.3	2067
Checkerboard (10000+999) $\times 2$	P-SMO	98.5375	98.2983	1347.75	1945
	I-SMO	98.08	97.625	447.5	1923
Letter Recognition (12000+3000) $\times 16$	P-SMO	94.64	93.24	10060	4203
	I-SMO	94.04	92.614	4874	3773
KRK (15000+1000) $\times 6$	P-SMO	99.81	99.85	12700	4076
	I-SMO	99.822	99.73	1426.7	2263

结论 针对标准 SMO 算法在优化后期进展缓慢的特点,我们改进了算法的停止条件。该策略在发现 SMO 进入到无显著收益的迭代后,加强了违反 KKT 条件的判定力度,收缩了样本空间,有效地缩短了 SMO 的训练时间,提高了 SMO 的性能。当处理大规模训练样本集时,该策略更为有效。这是因为数据集越大,SVM 初始分类面越易提早确定。未来的工作是:一方面,在 SMO 迭代后期,扔掉那些对 SVM 分类面没有影响的样本;另一方面,把新的策略和别的算法结合,进一步加快 SMO 的计算速度。

参考文献

- Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer Verlag, 1995
- Boser B E, Guyon I M, Vapnik V N. A Training Algorithm for Optimal Margin Classifiers. In: Haussler D, ed. Proceedings of the 5th Annual ACM Workshop on COLT, ACM press, Pittsburgh, 1992, 144~152
- Osuna E, Freund R, Girosi F. An Improved Training Algorithm for Support Vector Machines. In: IEEE Workshop on Neural Networks and Signal Processing, IEEE Press, Amelia Island, 1997, 276~285
- Joachims T. Making Large - Scale Support Vector Machine Learning Practical. In: Schoölkopf B, Burges C J C, Smola A J, eds. Advances in Kernel Methods - Support Vector Learning, MIT Press, Cambridge, MA, 1998, 169~184
- Platt J C. Sequential Minimal Optimization-A Fast Algorithm for Training Support Vector Machines. In: Schoölkopf B, Burges C J C, Smola A J, eds. Advances in Kernel Methods - Support Vector Learning, MIT Press, Cambridge, MA, 1998, 185~208
- Keerthi S S, Shevade S K, Bhattacharyya C, et al. Improvements to Platt's SMO Algorithm for SVM Classifier Design. Neural Computation, 2001, 13 (3): 637~649
- 孙剑,郑南宁,张志华.一种训练支持向量机的改进序贯最小优化算法.软件学报,2002,13(10):2007~2012
- 李建民,张钊,林福宗.序贯最小优化的改进算法.软件学报,2003,14(5):918
- 张浩然,韩正之.回归支持向量机的改进序列最小优化算法.软件学报,2003,14(12):2006~2013
- Lai D, Mani N, Palaniswami M. An Extrapolated Sequential Minimal Optimization Algorithm for Support Vector Machines. In: Proceedings of International Conference on Intelligent Sensing and Information Processing, Chennai, India, 2004, 415~420
- Keerthi S S, Gilbert E G. Convergence of A Generalized SMO Algorithm for SVM Classifier Design. Machine Learning, 2002, 46 (1-3): 351~360
- Lin C J. Convergence of the Decomposition Method for Support Vector Machines. IEEE Transaction on Neural Networks, 2001, 12(6): 1288~129