

一种新的基于 KKT 条件的错误驱动 SVM 增量学习算法^①

张灿淋, 姚明海, 童小龙, 张何栋

(浙江工业大学 信息工程学院, 杭州 310023)

摘要: 分析了 SVM 增量学习过程中, 样本 SV 集跟非 SV 集的转化, 考虑到初始非 SV 集和新增样本对分类信息的影响, 改进了原有 KKT 条件, 并结合改进了的错误驱动策略, 提出了新的基于 KKT 条件下的错误驱动增量学习算法, 在不影响处理速度的前提下, 尽可能多的保留原始样本中的有用信息, 剔除新增样本中的无用信息, 提高分类器精度, 最后通过实验表明该算法在优化分类器效果, 提高分类器性能方面上有良好的作用。

关键词: 增量学习; SVM; KKT 条件; 错误驱动

A New Error-Driven Incremental SVM Learning Algorithm Based on KKT Conditions

ZHANG Can-Lin, YAO Ming-Hai, TONG Xiao-Long, ZHANG He-Dong

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract: The transformation between the SV set and non-SV set is analyzed during the process of incremental SVM learning. Considering the initial non-SV set and new samples which will influence the accuracy of classification, it improves the KKT rule and error-driven rule. With these rules the new error-driven incremental SVM learning algorithm based on KKT conditions is proposed. With this algorithm, the useful information of original sample can be preserved as much as possible, the useless information of new samples can be removed accurately without affecting the processing speed. Experimental results show that this new algorithm has a good effect on both optimizing classifier and improving classification performance.

Key words: incremental learning; SVM; KKT condition; error-driven

1 引言

正如人脑学习一样, 人类也希望机器人的学习是渐进的、具有增值能力的, 能够不断的学习来自环境中新样本的新知识, 并能保留以前学过的有用的旧知识, 合理遗忘无用的旧知识, 而不是全部遗忘以前学过的知识, 从头开始训练^[1]. 因此, 增量学习最早在心理学家对人类以及其他生物的学习过程的研究中就被提出过. 早在 1962 年, Coppock 和 Freund 就在 Science 上发表过有关于增量学习的文章^[2].

支持向量机(SVM)作为一种根据统计学习理论提出的新的机器学习方法, 以其独特的优点, 近年来得到了广泛的发展^[3,4]. 以 SVM 为基础的增量学习,

也越来越成为了研究的热点问题. Syed 最早提出了基于支持向量机的增量学习算法^[5]——固定划分增量学习算法, 即 Batch SVM. 其基本思想是每次增量过程均只保留支持向量集(SV 集), 舍弃非支持向量集, 并将本次 SV 集与所有新增样本作为下一次增量学习的训练样本. 但事实上, 被舍弃的非支持向量, 也有可能隐性的包含了分类信息, 不做任何处理, 过早的、完全的舍弃非 SV 集将会在成分类精度的下降, 以至于不能反应真实的分类情况. 而同时有些新增样本也有可能包含完全无用的信息, 将新增样本不做任何取舍, 全部拿来作为下一次训练的样本, 必将导致处理时间和存储空间上的浪费. 文献[6,7]引入了

^① 基金项目:国家自然科学基金(61070113)

收稿时间:2013-06-18;收到修改稿时间:2013-07-09

KKT 条件, 来提取新增样本中的有用信息, 文献[8]引入了错误驱动的概念, 把错分类样本也考虑在内. 本文在分析了上述几种 SVM 增量学习算法的基础上, 改进了原有 KKT 条件, 并结合改进了的错误驱动策略, 提出了一种新的基于 KKT 条件的错误驱动增量 SVM 学习算法, 尽可能多的保留原始样本中的有用信息, 并舍弃新增样本中的无用信息, 来提高算法的分类精度和处理速度.

2 增量式支持向量机理论

支持向量机是由 Vapnik 等在统计学理论上提出的一种新的机器学习方法. 由于其以 VC 维为理论, 并在学习模型中考虑了结构风险最小化原则. 因此具有很强的学习能力和泛化能力, 并在克服维数灾难和收敛速度等问题上也有很好的表现^[9].

支持向量机最初是用于解决二类线性可分问题的, 假设对于训练数据集:

$$\Omega = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

其中, $x \in R^n$, $y \in \{-1, 1\}$.

支持向量机就是在数据空间 Ω 上, 寻找一个超平面 $(w \bullet x) + b = 0$, 如图 1 所示. 使得两类数据集能够

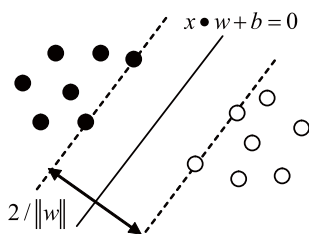


图 1 最优超分类面示意图

尽可能无错误的分开, 归一化后满足:

$$|y_i(w \bullet x_i + b)| \geq 1 \quad (1)$$

其中 $i = 1, 2, \dots, n$; 并使得两类之间的分类间隔最大, 等价于使 $\|w\|^2 / 2$ 最小, 即:

$$\min(J) = \min_{w,b} \|w\|^2 / 2 \quad (2)$$

因此, 寻找最优超平面即为在满足式(1)的前提下, 求解式(2)的最优解. 引入拉格朗日函数:

$$L(w, b, \alpha) = \frac{1}{2}(w \bullet w) - \sum_{i=1}^n \alpha_i \{[(x \bullet w) + b]y_i - 1\} \quad (3)$$

式中, α_i 为拉格朗日乘子, 该问题为一个凸二次规划

问题, 存在唯一的最优解. 最后求得分类函数为:

$$f(x) = \text{sign} \left[\sum_{i \in SV} y_i \alpha_i (x_i \bullet x) + b \right] \quad (4)$$

其中 SV 集为数据空间 Ω 的支持向量集合, k 为 SV 集的个数. 由式(4)可知, 支持向量机的决策函数仅与其当前空间的 SV 集有关, 与其余数据无关.

对于增量式 SVM 来说, 新增样本的加入打破了原有 SV 集与整个训练数据集的等价关系. 那么如何寻找新的 SV 集, 使这种等价关系继续保持. 这就是增量式 SVM 所需要做的.

假如加入的新增样本不包含新知识, 那么原 SV 集不变; 若加入的新增样本数据包含新知识, 那么原来的 SV 集必将发生变化, 以体现样本集的改变^[10].

考虑新增样本对原 SV 集的影响, 当且仅当对于每个样本 x 都满足如下 KKT 条件时, $\alpha = [\alpha_1, \dots, \alpha_k]$ 才是凸二次规划问题(式 3)的最优解:

$$\begin{cases} \alpha_i = 0 \Rightarrow |f(x_i)| \geq 1 \\ 0 < \alpha_i < C \Rightarrow |f(x_i)| = 1 \\ \alpha_i = C \Rightarrow |f(x_i)| \leq 1 \end{cases} \quad (5)$$

对于 SVM 训练得到的分类器而言, $\alpha = 0$ 对应于样本分布在分类间隔之外, 如图 2 中的 A 所示; $0 < \alpha < C$ 对应于样本分布在分类间隔上, 如图 2 中的 C 所示; $\alpha = C$ 对应于分布在分类间隔内, 如图 2 中的 B 所示.

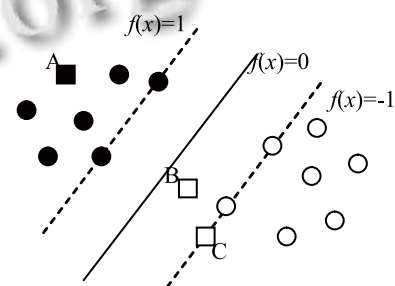


图 2 新增样本(方形)与 KKT 条件关系

综合上述三类情况, 假设 $f(x)$ 为原样本集的决策函数, $\{x_i, y_i\}$ 为新增样本, 当满足 KKT 条件, 即 $|f(x_i)| \geq 1$ 时, 该新增样本不会使原 SV 集发生变化; 当不满足 KKT 条件, 即 $|f(x_i)| < 1$ 时, 该新增样本将影响原 SV 集.

因此, 基于 KKT 条件的 SVM 增量学习, 在学习过程中, 只考虑违背 KKT 条件的新增样本, 并将这些样本与原 SV 集组成新的训练样本进行增量学习, 得到增量后的新的 SV 集. 该算法将原来的非 SV 集全部舍弃, 否认了非 SV 对分类信息的影响, 必将造成分类进度上的下降.

3 一种新的基于KKT条件的错误驱动增量学习算法

在分析了基于 KKT 条件的增量 SVM 后, 考虑到原始样本中非 SV 集对分类信息的影响, 本文改进了 KKT 条件下的增量 SVM, 将原始样本中的非 SV 集进行了一定取舍, 保留了其中有可能成为 SV 集的样本, 并在此基础上考虑到新增样本中, 可能会存在虽然满足 KKT 条件, 但却包含了新的分类知识的样本, 即被当前分类器错分类的样本. 将两者结合提出一种新的 KKT 条件下的错误驱动增量学习算法.

3.1 改进的 KKT 条件

由前面分析可知, 当新增样本中存在不满足 KKT 条件的点时, 那么分类器的原 SV 集即将发生变化: 原来属于 SV 集的样本可能退化为非 SV; 原来属于非 SV 集的样本也可能转化为 SV.

如图 3 所示, 黑细线表示前一次的分类器, 红粗线表示增量后的分类器. 原 SV 集由样本 A1、A2、B1、B2 构成, 当新增样本 B3 不满足 KKT 条件时, 属于原 SV 集的 A1、B2 退化为非支持向量; 而原非 SV 集的 A3 转化为支持向量.

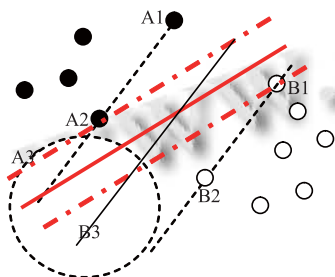


图 3 SV 集与非 SV 集的转化图

本文提出在下一步增量学习时, 必须考虑到原非 SV 集中的有关向量, 改进了原有 KKT 条件, 假设新增样本 $\{x_i, y_i\}$ 不满足 KKT 条件, 那么在该样本附近的非支持向量最有可能转化为支持向量. 只要找到不满

足 KKT 条件的样本附近的非支持向量, 并将其加入到下一步增量学习的训练样本中, 即可以更好的保留原样本集中的分类信息, 提高分类精度. 在此, 引入样本类间距离 θ_m , 若在不满足 KKT 条件的样本附近 θ_m 的范围之内, 存在 n 个非支持向量, 那么在下一步增量学习时, 同时考虑这 n 个向量.

3.2 错误驱动策略

由于采集样本的有限性, 当前所学样本不能反映真实分布, 导致学习知识的不全面, 系统的泛化能力不足以识别新知识时, 可能会出现分类错误. 当错误样本出现在分类间隔之外时, 系统当前的 KKT 条件已无法识别, 即假设新增样本为 $\{x_i, y_i\}$, 而其 $|f(x_i)| \geq 1$. 如图 4 中样本 A 所示.

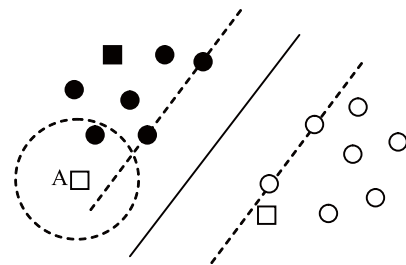


图 4 错分类样本图

若以基于 KKT 条件的增量算法来考虑, 必将舍弃这些在分类间隔之外的误分类样本, 而事实上这些样本, 包含了很大一部分新知识. 舍弃这些样本必将导致学习的不全面, 最终降低分类器的精度.

因此, 提出错误驱动策略, 不仅将这些分类间隔之外的误分类样本加入到下一次增量学习的训练集中, 同时, 以此为基础, 在这些样本附近 θ_m 的范围之内, 搜索原样本中的非支持向量, 并同样将其放入下一次的训练集中.

3.3 算法描述

结合改进的 KKT 条件和错误驱动策略, 提出一种新的 KKT 条件下的错误驱动增量学习算法.

假设初始样本集为 Ω_0 , 新增样本集为 Ω_1 , 用原始样本集通过 SVM 训练得到的分类器为 Φ_0 , 增量学习之后得到的分类器为 Φ_1 , 原始样本中经训练得到的 SV 集为 Ω_{0SV} 、非 SV 集为 Ω_{0NSV} , 非 SV 集中满足改进 KKT 条件的为 Ω_{0NSV}^k 、满足错误驱动策略的为 Ω_{0NSV}^e , 新增样本中, 不满足 KKT 条件的为 Ω_1^k , 被误分类且

满足 KKT 条件的为 Ω_1^e .

具体算法描述如下:

(1) 用 Φ_0 训练得到分类器 Φ_0 ;

(2) 用 Φ_0 来检测 Ω_1 , 若不存在 Ω_1^k 和 Ω_1^e , 则算法结束, Φ_0 为增量学习后的分类器; 否则, 根据改进的 KKT 条件和错误驱动策略, 分别得到对应的 Ω_{0NSV}^k 和 Ω_{0SV} ;

(3) 将 Ω_{0SV} 、 Ω_{0NSV}^k 、 Ω_{0NSV}^e 、 Ω_1^k 和 Ω_1^e 作为下一次增量学习的训练集, 训练得到增量学习后的分类器 Φ_1 ;

(4) 更新 Ω_{0SV} 、 Ω_{0NSV} , 作为下一次增量学习时的初始样本, 并用当前样本集来更新类间距离 θ_{in} .

4 实验数据仿真

4.1 实验一

为直观显示本文增量学习算法对分类器的优化效果, 本实验采用二维数据做训练样本, 每次学习 10 个样本点, 并将优化效果同普通基于 KKT 条件的增量算法作比较, 做如下优化图.

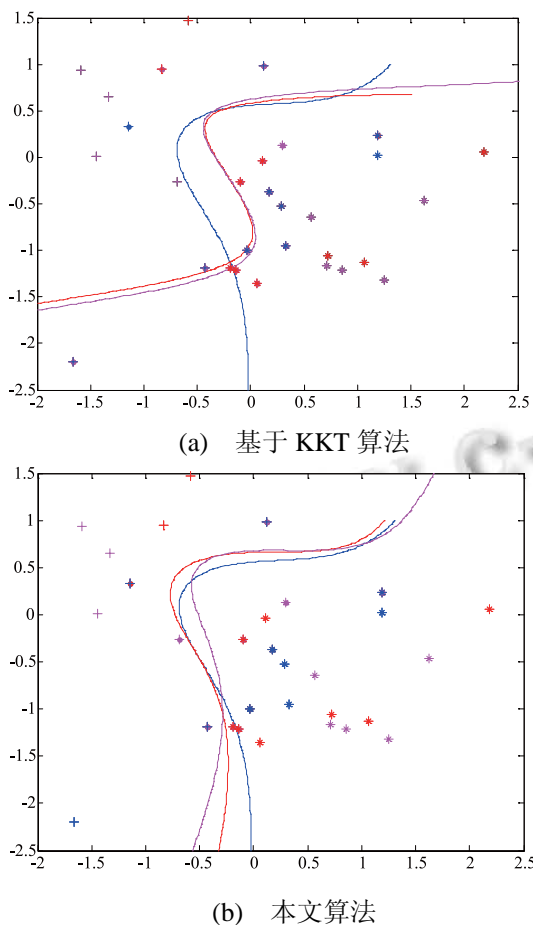


图 5 分类器优化效果图

图中, 蓝色、红色、粉色分别表示第一次、第二次、第三次训练得到的分类器. ‘+’和‘*’表示两类不同样本.

由图分析, 由于初始第一次训练时, 并没有涉及到增量学习, 因此本文算法和基于 KKT 的算法在第一次训练后得到的分类器是一样的. 但当第二、三次增量学习时, 基于 KKT 的算法舍弃了原始样本中的非 SV 集, 即上图中左下角的蓝色‘+’, 因此在增量学习时, 否定了该样本点对分类信息的贡献, 导致最终分类错误. 而本文算法, 考虑到了原始样本中非 SV 集对分类信息的影响, 保留了有可能成为 SV 集的非支持向量, 因此最终学习得到的分类器很大程度上反映了数据样本的真实分布.

4.2 实验二

本实验采用 UCI 标准数据集里的 Balance 数据集来验证本文算法的性能, 将该数据集分成独立的两部分, 一部分作为训练集 TR, 另一部分作为测试集 TE. 实验过程中, 进一步将训练集和测试集分为独立的 10 组数据, 来模拟增量学习过程, 每增量学习一次, 对该次的测试集进行测试, 记录其识别率.

实验时, 将 Balance 的第三类作负类, 其他类作正类. 同时将本文算法跟非增量(unincremental)学习法以及基于 KKT 的算法作比较, 做如下识别率曲线:

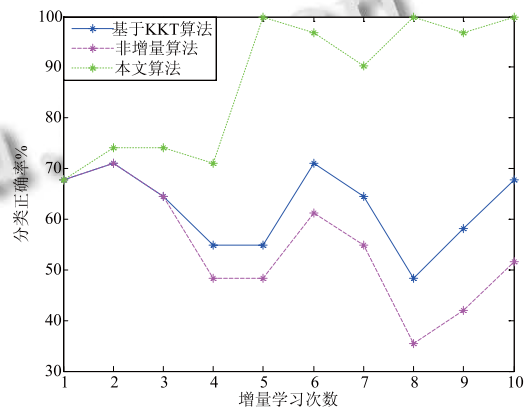


图 6 学习效果比较图

为更直观显示各方法之间识别率的区别, 将图 6 中的数据进行了统计, 做如表 1 所示.

由表 1 分析可得, 由于初始第一次训练时, 并没有涉及到增量学习, 因此三种算法性能一样; 但随着新增样本的加入, 非增量算法的泛化能力已远远不能反应新样本的信息, 导致其识别率的低下; 而基于

KKT 的算法,虽然识别率较非增量算法有一定提高,但其没考虑样本集中非 SV 集对分类信息的影响,因此其分类精度提高的不快;而本文算法,在增量到 4、5 步的时候,分类器的性能已远远优于其他算法,并能在接下来的学习中一直保持较好的性能.

表 1 各算法识别率的比较表

算法 学习数	非增量算法	基于 KKT 算法	本文算法
1	67.74%	67.74%	67.74%
2	70.97%	70.97%	74.19%
3	64.52%	64.52%	74.19%
4	48.39%	54.84%	70.97%
5	48.39%	54.84%	100%
6	61.29%	70.97%	96.77%
7	54.84%	64.52%	90.32%
8	35.48%	48.39%	100%
9	41.94%	58.06%	96.77%
10	51.61%	67.74%	100%

5 结语

本文提出了一种新的基于 KKT 条件的错误驱动增量学习算法.考虑了在增量学习过程中有可能成为 SV 的非支持向量,以及新增样本中满足 KKT 条件但被错分类的样本,并将新增样本做改进的 KKT 条件和错误驱动策略的筛选.实验结果显示,该算法在优化分类器效果,提高分类器性能方面上有良好的作用.

参考文献

1 王作为.具有认知能力的智能机器人行为学习方法研究[学

位论文].哈尔滨:哈尔滨工业大学,2010.

- Coppock H, Freund J. All or None versus Incremental Learning of Errorless Shock Escapes by the Rat. *Science*, 1962, 135: 318–319.
- Epifanio B, Takio K, Masakatsu H, Hiroaki I. Adapting SVM Image Classifiers to Changes in Imaging Conditions Using Incremental SVM: An Application to Car Detection, *Computer Vision-AACV2009*.
- Werner B. Incremental Regularization to Compensate Biased Teachers in Incremental Learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010: 1–8.
- Syed NA, Sung K. Handling Concept Drifts in Incremental Learning with Support Vector Machines. *Proc. of the 5th ACM SIGKDD international conference*, 1999: 316–321.
- Christopher P, Johns H. SVM Incremental Learning Adaptation and Optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003: 2685–2691.
- Zhao Y, Wang K, Zhong P. Incremental support vector machine based on border samples. *Computer Engineering and Design*, 2010, 31: 161–164.
- Luo J, Pronobis A, Caputo B. Incremental Learning for Place Recognition in Dynamic Environments. *IEEE International Conference on Intelligent Robots and Systems*. 2007. 721–729.
- 王珏,石纯一.机器学习研究.广西师范大学学报,2003,21: 1–15.
- 李祥纳,艾青,秦玉平.支持向量机增量学习算法综述,渤海大学学报,2007,9(4):187–189.

(上接第 201 页)

2012,7(29),375–378.

- 马耀兰.基于支持向量机方法的股票预测模型.鸡西大学学报,2013,4(13):124–125.
- 胡文军,王士同等.适合大样本快速训练的最大夹角间隔核心集向量机.电子学报,2011,39(5):1178–1184.
- 朱孝开,杨德贵.基于推广能力测度的多类 SVDD 模式识别方法.电子学报,2009,37(3):464–469.
- 林雄,冯海.基于 SVDD 多类分类新方法的研究.信息技术, 2008,7:100–103.
- 谢志强,高丽,杨静.基于球结构的完全二叉树 SVM 多类分类算法.计算机应用研究,2008,25(11):3268–3271.

- Ban T, Abe S. Implementing Multi-class Classifiers by One-class Classification Methods. *International Joint Conference on Neural Networks Vancouver*. BC, Canada. Sheraton Vancouver Wall Centre Hotel. 2006. 3272–332.
- 吴强,贾传炎,张爱锋,刘爽.球结构支持向量机的改进算法及仿真研究.系统仿真学报,2008,20(2):345–348.
- 刘小平.一种新的不平衡数据 v-NSVDD 多分类算法.南京大学学报(自然科学),2013,2(49):150–157.
- 汪东.基于支持向量机的选时和选股研究[学位论文].上海:上海交通大学,2007.