支持向量机多类目标分类器的结构简化研究

王立国 张 晔 谷延锋
(哈尔滨工业大学信息工程系,哈尔滨 150001)

摘 要 由于支持向量机(SVM)在模式识别和回归分析中有着独特优势,因此成为近来研究的热点,其优势主要体现在处理非线性和高维数据问题方面。最初的 SVM 特别适合解决两类目标分类问题,而对于多类目标分类,则需将其转化为多个两类目标分类问题,相应地即可构造多个两类目标子分类器,但由于这种情况导致了分类器结构的过于复杂,从而导致判决速度的降低。为了快速地进行分类,提出了一种简化结构的多类目标分类器,其不仅使得子分类器数目大大减少,而且使分类速度明显提高;同时对其分类精度和复杂度进行了对比分析。实验结果证明,该分类器是有效的。

关键词 支持向量机 多类目标分类器 核函数 模式识别
中图法分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2005)05-0571-04

The Research of Simplification of Structure of Multi-class Classifier of Support Vector Machine

WANG Li-guo, ZHANG Ye, GU Yan-feng
(*Deptartment of Information Engineering, Harbin Institute of Technology, Harbin* 150001)

Abstract Because of the unique property in pattern recognition and in regression analysis, support vector machine(SVM) has become the topic of research recently. The advantages of SVM mainly lie in its capabilities of processing non-linear and highly dimensional data problems. Unextended SVM is very suitable for solving two-class classification problems. For multi-class classification, however, it should be converted into many of two-class classification problems, and can be constructed many of two-class classifier correspondingly. But the case results in more complexity of classifier structure, and so leads to decrease of decision speed. In order to get a fast classification, a new multi-class classifier with simplified structure is put forward so that the number of subclassifiers and decisive time are reduced greatly. The accuracy and complexity are also contrastively analyzed here. The validity of the new classifier is proved by simulated experiments.

Keywords support vector machine (SVM), multi-class classifier, kernel function, pattern recognition

1 SVM 简介

1960 年以来,Vapnik 等学者研究建立了统计学习理论,并提出了结构风险最小化原理[1]。众所周知,SVM 是统计学习理论中较新的一门技术,已成为近年来统计学习机器解决模式识别和回归分析等实际问题的热点。目前,该技术已被广泛地应用到诸如手写字识别、人脸识别、指纹识别、头部姿态识别、文本自动分类、超谱图像分类、数据挖掘、非线性系统控制、故障诊断、函数模拟等领域,其发展趋势方兴未艾。与现有的诸如神经网络、遗传算法等智能学习机相比,由于 SVM 有着坚实的理论基础和较强的推广能力,因此在处理非线性问题和高维数据问题上显示出优良的性能。

SVM 的工作机理可概括为:寻找一个分类超平面来使得训练样本中的两类样本点能被分开,并且距离该平面尽可能地远;而对线性不可分的问题,则

基金项目:国家自然科学基金资助项目(69972013)
收稿日期:2003-05-16;**改回日期**:2004-12-17

第一作者简介:王立国(1974 ~),男。2002 年获哈尔滨工业大学理学硕士学位,现为该校信息工程系信号与信息处理专业博士研究生。研究方向为高光谱图像处理、支持向量机理论等。E-mail:wlg74327@hit.edu.cn

可通过核函数将低维输入空间的数据映射到高维空间,以便将原低维空间的线性不可分问题转化为高维空间上的线性可分问题。

对于线性可分的两类分类问题,其关键技术之一是寻找最优分类超平面,即寻找最优线性判别函数。设 $\boldsymbol{x}_i \in \boldsymbol{R}^d$ 为样本数据,$y_i \in \{+1,-1\}$ 为相应的类别标号,$i=1,\cdots,n$。线性判别函数的一般形式为 $g(\boldsymbol{x})=\boldsymbol{w}\cdot\boldsymbol{x}+b$,其相应的分类面为 $\boldsymbol{w}\cdot\boldsymbol{x}+b=0$。这样,为了使待分样本尽可能好地分开,则要求分类间隔(可表示为 $2/\|\boldsymbol{w}\|$)最大,这相当于使 $\|\boldsymbol{w}\|$ 最小[2]。这样,寻找最优分类面就转化为下面的优化问题:

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 \tag{1}$$

$$\text{s.t.}\quad y_i[(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b]-1\geqslant 0, i=1,2,\cdots,n$$

在处理非线性可分问题方面,SVM 的一个显著优势在于下面的两点延拓:

(1) 通过引入核函数 φ 解决非线性问题,即将 $\boldsymbol{x}_i$ 替换为 $\varphi(\boldsymbol{x}_i)$,将内积 $\boldsymbol{x}_i\cdot\boldsymbol{x}_j$ 替换为 $K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_j)$。

(2) 通过引入松弛变量 $\xi_i, i=1,2,\cdots,n$,用以解决不可分问题,即针对错分误差引入惩罚因子 γ。

这样,式(1)可重新描述为

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2+\gamma\sum_{i=1}^{n}\xi_i \tag{2}$$

$$\text{s.t.}\quad y_i[(\boldsymbol{w}\cdot\varphi(\boldsymbol{x}_i))+b]-1+\xi_i\geqslant 0$$

$$\xi_i\geqslant 0; i=1,2,\cdots,n; \gamma>0$$

对于这个凸优化问题,可引入拉格朗日乘子 $\alpha_i, i=1,2,\cdots,n$,将其转化为下面的对偶形式:

$$\max Q(\boldsymbol{\alpha})=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{3}$$

$$\text{s.t.}\quad 0\leqslant\alpha_i\leqslant\gamma; i=1,2,\cdots,n; \sum_{i=1}^{n}\alpha_i y_i=0$$

其中,$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm{T}}$。

这是一个带有不等式约束的二次优化问题,并存在唯一的最优解 $(\hat{\boldsymbol{\alpha}},\hat{b})$,其对应最优判决函数为

$$f(x)=\operatorname{sgn}\left\{\sum_{i=1}^{n}\hat{\alpha}_i K(\boldsymbol{x}_i,\boldsymbol{x})+\hat{b}\right\} \tag{4}$$

近年来,学者们较多地研究了核函数的构造及各种算法的优化来提高 SVM 的性能,使得 SVM 技术在理论和应用方面都得到了极大发展。然而,对于分类器结构优化问题的研究却相对较少。如今对于两类目标的分类问题,可以利用 SVM 技术直接解决,并且只需构造一个分类器;而对于多类目标分类问题,则需将其转化为多个两类目标分类问题,且需相应地构造多个两类目标子分类器。由于这种情况导致了分类器结构的过于复杂,从而使得 SVM 在多类问题的应用上受到了限制。基于上述原因,开展分类器结构的研究有其重要的现实意义和应用价值。

2 SVM 多类目标分类器的构造

基于 SVM 分类的训练(学习)过程就是为了得到一个分类器用来解决相应的分类问题,但不同的分类器结构影响着程序设计的难度、分类训练时间和测试时间。

2.1 现有多类目标分类器的构造方法

基于 SVM 的分类,无论是算法还是理论,其研究主要集中在两类分类问题上,而对于多类分类问题则往往通过转化为多个两类分类问题去解决,其典型的处理方法有以下 3 种[3]:一种是由以 Weston 为代表提出的多值分类算法,它是通过重构多值分类模型,并通过对其目标函数进行优化来实现分类,其缺点是算法选择的目标函数过于复杂而难以实现,同时计算量也很大;另外的两种方法则分别为 1-a-r(1-aginst-rest)法和 1-a-1(1-aginst-1)法,这是目前最为常用的两种方法。前者虽设计简单,但分类器的结构较为复杂,计算量也很大,后者改进了传统的一对多算法,虽易于构造,但计算量很大,分类器结构也过于复杂。现以 N 类问题为例来说明,算法 1-a-r 是去构造 N 个两类目标子分类器,其中,第 k 个子分类器用第 k 类中的训练样本作为正的训练样本,其余的作为负的训练样本,对于某个输入样本,其分类结果则为各子分类器输出值为最大的相应类别。算法 1-a-1 是由 Knerr 提出的,由于它是将 N 类中的每两类构造一个子分类器,从而需要构造 $N(N-1)/2$ 个子分类器,然后组合这些子分类器,即可利用投票法确定分类结果。由于这两种方法的共同缺点是推广误差无界,且分类器数目多,从而导致决策速度慢[3]。王建芬等提出一种结合决策树方法的分类器结构[4],该方法对于 N 类问题需构造 $N-1$ 个子分类器,可见子分类器数目减少的幅度依然不大。

针对子分类器数目过多,致使分类器结构复杂的问题,本文提出一种简化多类目标分类器结构的方法。

2.2 一种简化多类目标分类器的构造

为了说明的方便,可先以 8 类问题为例来对分类器结构的构造进行说明。设全部样本集合为 $\boldsymbol{P}$,

分类器的构造过程如下:

(1) 先将 $\boldsymbol{P}$ 按类别等分为两个样本集合,分别记为 $\boldsymbol{P}_1$、$\boldsymbol{P}_2$。又记 $\boldsymbol{A}_1=\boldsymbol{P}_1$、$\boldsymbol{B}_1=\boldsymbol{P}_2$,并分别重置 $\boldsymbol{A}_1$、$\boldsymbol{B}_1$ 相应的类别标号为 +1、−1,然后将 $\boldsymbol{A}_1$、$\boldsymbol{B}_1$ 作为两类目标进行训练即可得到第 1 个两类目标子分类器 C_1。

(2) 先将 $\boldsymbol{P}_1$ 按类别等分为两个样本集合,分别记为 $\boldsymbol{P}_{11}$、$\boldsymbol{P}_{21}$;然后将 $\boldsymbol{P}_2$ 按类别等分为两个样本集合,分别记为 $\boldsymbol{P}_{11}$、$\boldsymbol{P}_{21}$。同时记 $\boldsymbol{A}_2=\boldsymbol{P}_{11}\cup\boldsymbol{P}_{12}$(∪为集合取并运算符号)、$\boldsymbol{B}_2=\boldsymbol{P}_{21}\cup\boldsymbol{P}_{22}$,并分别重置 $\boldsymbol{A}_2$、$\boldsymbol{B}_2$ 相应的类别标号为 +1、−1;最后将 $\boldsymbol{A}_2$、$\boldsymbol{B}_2$ 作为两类目标进行训练即可得到第 2 个两类目标子分类器 C_2。

(3) 先将 $\boldsymbol{P}_{11}$ 按类别等分为两个样本集合,分别记为 $\boldsymbol{P}_{111}$、$\boldsymbol{P}_{211}$;然后将 $\boldsymbol{P}_{21}$ 按类别等分为两个样本集合,分别记为 $\boldsymbol{P}_{121}$、$\boldsymbol{P}_{221}$;接着将 $\boldsymbol{P}_{12}$ 按类别等分为两个样本集合,分别记为 $\boldsymbol{P}_{112}$、$\boldsymbol{P}_{212}$,再将 $\boldsymbol{P}_{22}$ 按类别等分为两个样本集合,分别记为 $\boldsymbol{P}_{122}$、$\boldsymbol{P}_{222}$,同时记 $\boldsymbol{A}_3=\boldsymbol{P}_{111}\cup\boldsymbol{P}_{121}\cup\boldsymbol{P}_{112}\cup\boldsymbol{P}_{122}$、$\boldsymbol{B}_3=\boldsymbol{P}_{211}\cup\boldsymbol{P}_{221}\cup\boldsymbol{P}_{212}\cup\boldsymbol{P}_{222}$,并分别重置 $\boldsymbol{A}_3$、$\boldsymbol{B}_3$ 相应的类别标号为 +1、−1,最后将 $\boldsymbol{A}_2$、$\boldsymbol{B}_2$ 作为两类目标进行训练即可得到第 3 个两类目标子分类器 C_3。

(4) 将 3 个子分类器 C_1、C_2、C_3 组合成为一个多类目标分类器 C。

至此,通过 3 个子分类器的判定交集就可以将待判定样本判定为唯一的一类。图 1 给出了 8 类问题分类中 3 个子分类器的构造示意简图,其中,矩形框表示原类别集合及被分裂的情况,圆形框表示在某步中被归为一类的原始类别标号集。值得注意的是,在上面的分类器构造过程中,每一步的类别等分方式都是任意的。

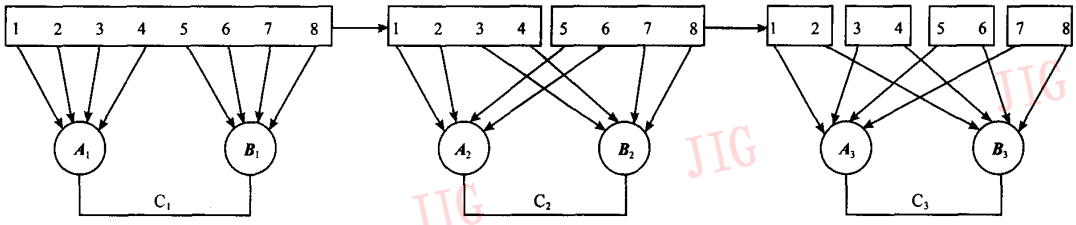

图 1　8 种类别的 3 个子分类器的构造

Fig. 1　The construction of 3 subclassifiers for 8 classes

一般对于 2^N 类问题来说,可通过 N 步完成。具体过程如下:

(1) 将原始样本中的 2^{N-1} 类别样本合为一类样本集合,余下为另一类样本集合,由此进行训练得到第 1 个两类目标子分类器。原始样本被按类别分裂为 2 个集合;

(2) 从上一步被分裂的 2 个集合中,各取 2^{N-2}(共 2^{N-1})类别样本合为一类样本集合,余下的为另一类样本集合,由此进行训练即可得到第 2 个两类目标子分类器,而原始样本则被按类别分裂为 4 个集合。

在第 k 步,从 $k-1$ 步得到的 2^{k-1} 个分裂类别集合中各取 2^{N-k}(共 2^{N-1})类别样本合为一类样本集合,余下的为另一类样本集合,由此进行训练即可得到第 k 个两类目标子分类器,而原始样本则被按类别分裂为 2^k 个集合。

这样重复进行,将得到 N 个子分类器,并可将它们组合成一个多类目标分类器,然后通过 N 个子分类器的输出值就可以将每个样本判定为唯一的一类。

如果类别数在 2^{N-1} 至 2^N 之间,则可以通过添加虚拟类别的方法将其类别数转化为"2^N",然后在最后构造出的 N 个子分类器中再去掉虚拟类别样本。事实上,所添加的虚拟类别只是形式上的参与而无实际参与。例如,当类别数为 6 时,则可以添加第 7、8 类,这样就获得了与图 1 相同的分类器构造形式。所不同的是,设计结果要在最后的 $\boldsymbol{A}_1$、$\boldsymbol{B}_1$、$\boldsymbol{A}_2$、$\boldsymbol{B}_2$、$\boldsymbol{A}_3$、$\boldsymbol{B}_3$、$\boldsymbol{A}_4$、$\boldsymbol{B}_4$ 中去掉第 7、8 类样本。

不难分析,本文方法得到的子分类器数目远少于 1-a-1、1-a-r 两种典型方法。表 1 为不同类别数下 3 种方法所需子分类器数目的比较(其中,$K=2^N$)。

表 1　3 种方法所需子分类器数目的比较

Tab. 1　The comparative of the number of needed subclassifiers for 3 methods

类别数	1-a-1	1-a-r	本文方法
4	6	4	2
16	120	16	4
K	$K(K-1)/2$	K	N

2.3　精度与复杂度的对比分析

通过分析可知,最终正确分类的样本是指在每级子分类器中都能正确分类的样本,而最终错分的样本

则为各级错分样本的交集。设一分类器共有 N 个子分类器,其分类精度依次为 $a_1,a_2,\cdots,a_N$,其中最小的设为 $a_{\min}$,则由上面的分析,分类精度 a 的范围为

$$1-[(1-a_1)+(1-a_2)+\cdots+(1-a_N)]\leqslant a\leqslant a_{\min}$$

即

$$a_1+a_2+\cdots+a_N+1-N\leqslant a\leqslant a_{\min} \quad (5)$$

而对于 1-a-r 和 1-a-1 两种传统方法,其分类精度将高于各子分类器的最小分类精度。

如果不考虑设计的复杂度,则可以用分类的时间来衡量分类器的复杂度。对于训练而言,由于不同结构下各子分类器的复杂度可能不同,而相同结构中各子分类器间还存在着特定的联系,所以不能仅以子分类器的数目来衡量分类器的复杂度。相比之下,由于分类的测试过程不受上面因素的制约,其耗时主要用在核函数的运算上,因而可以以其所需处理核函数运算的次数来作为复杂度的衡量指标。在相同条件下,对于相同的 $K=2^N$ 类训练样本,则可以计算出不同结构的分类器在测试中所需处理的核函数运算次数的相对值(表 2)。

表 2 3 种方法在测试中核函数运算次数的比较

Tab. 2 The comparative of times for computing kernel function for 3 methods

类别数目	1-a-r	1-a-1	本文方法
4	4	3	2
16	16	15	4
K	K	$K-1$	N

究其实质,相对于两种传统方法,由于本文方法舍弃了各子分类器间大量的冗余信息及其较小的纠错补偿,因此可在牺牲较小的分类精度的同时,获得分类器结构的简化和分类速度的较大提高。

3 仿真实验及相关分析

本文分类实验利用的是 1992 年 6 月拍摄的美国印第安纳州西北部印第安遥感试验区的典型 AVIRIS 高光谱图像。实验时,选取真实图中的 4 类地物进行分类实验。训练样本 400 对,测试样本 320 对。实验采用高斯核函数的最小二乘支持向量机[5]和高效的 SMO 算法[6],且迭代运算中不存储核函数,同时采用 1-a-r 法和 1-a-1 法作为参考。表 3 给出了在相同迭代终止标准、不同方法下所用的训练时间和测试时间,时间单位为 s。本文方法得到的分类精度为 93.75%,两种参考方法的分类精度分别为 94.69%、94.37%。实验结果表明,基于本文提出的分类器结构所构建的算法无论是训练时间,还是测试时间都远小于两种传统方法,而分类精度却降低不到 1%。实验结果完全地证实了前面的理论分析。

表 3 3 种方法下的训练时间和测试时间比较

Tab. 3 The comparative of training time and test time for 3 methods

	1-a-r	1-a-1	本文方法
训练时间(s)	81.7500	59.7970	37.5780
测试时间(s)	14.4702	11.21870	7.4288

4 结 论

本文提出了一种简化结构的分类器,它可以大大地降低分类器的复杂性。其优点是多方面的,如可减少训练时间、减少测试时间、降低编程复杂度以及子分类器数目的减少使得各判决函数中的参数单独调节成为可能等。但必须指出,本文方法获取的优点是以牺牲较小的分类精度为代价的。由于在目标分类问题中,分类精度和分类速度常为一对互相矛盾的指标,因此在解决实际问题中究竟采用哪种分类器,应根据用户的要求而定,其中对于分类速度要求较高的问题(如 SVM 的实时应用),本文提出的方法就极为有效。如果需将分类精度与分类速度综合考虑,则可以将传统分类器与本文提出的分类器组合成混合的分类器来协调二者间的需求矛盾。因此进一步的研究将集中在提高该分类器的分类精度上,并将其应用到实际问题中。

参考文献(References)

1 Vapnik V N. The nature of statistical learning theory[M]. New York: Springer Press, 1995.

2 BIAN Zhao-qi, ZHANG Xue-gong. Pattern recognition (Second edition)[M]. Beijing: Tsinghua University Press, 2000. [边肇祺,张学工. 模式识别(第二版)[M]. 北京:清华大学出版社,2000.]

3 LIU Jiang-hua, CHENG Jun-shi, CHEN Jia-pin. Support vector machine training algorithm: A review[J]. Information and Control, 2002, **31**(1): 45~50. [刘江华,程君实,陈佳品. 支持向量机训练算法综述[J]. 信息与控制,2002,**31**(1):45~50.]

4 WANG Jian-fen, CAO Yuan-da. The application of support vector machine in classifying large number of catalogs[J]. Journal of Beijing Institute of Technology, 2001, **21**(2): 225~228. [王建芬,曹元大. 支持向量机在大类别数分类中的应用[J]. 北京理工大学学报,2001,**21**(2):225~228.]

5 Suykens J A K, Brabanter J D, Lukas L, *et al*. Weighted least squares support vector machines: Robustness and sparse approximation[J], Neurocomputing, 2002, **48**(1): 85~105.

6 Keerthi S S, Shevade S K. SMO algorithm for least squares SVM [A]. In: Proceedings of the International Joint Conference on Neural [C], Portland, Oregon, USA. 2003, **3**: 2088~2093.