

文章编号:1001-9014(2004)06-0418-05

# 一种新的多类模式识别方法

方勇, 戚飞虎

(上海交通大学 计算机科学与工程系, 上海 200030)

**摘要:**提出了一种基于支撑向量机的多类分类器,用 $N-1$ 个支撑向量机组合构成一个具有二叉树结构形式的 $N$ -多类分类器.讨论了该多类分类器的泛化推广能力,同时还提出了该多类分类器的基于特征空间的BTSVM学习算法. BTSVM算法使用核函数转换的方式计算特征空间的样本距离;采用类间最小距离最大化作为聚类准则,在每个决策结点产生两个最优子集;然后采用支撑向量机学习算法学习两个最优子集,确定决策结点的最优分类面.理论和实验结果表明,本文提出的基于支撑向量机的多类分类器在整体性能上要优于其它类似的分类器系统.

**关键词:**模式识别;多类分类器;支撑向量机;BTSVM算法

**中图分类号:**TP391 **文献标识码:**A

## NOVEL APPROACH TO MULTI-CLASS CLASSIFICATION

FANG Yong, QI Fei-Hu

(Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract:** A new multi-classifier was proposed based on support vector machines for a  $N$ -class classification problem, which comprised  $N-1$  support vector machines in the form of a binary tree. The generalization performance of multi-classifiers was discussed, and a new learning algorithm, the BTSVM algorithm, was presented based on high-dimension feature spaces. The BTSVM algorithm evaluates example distances by kernel functions, employs the maximization of minimum distances as clustering criteria to obtain two optimal subsets, and generates the optimal classification functions with support vector machine learning algorithms at each decision node. Theoretical analysis and experimental results show that the BTSVM algorithm is superior to other competitive multi-classifiers.

**Key words:** pattern recognition; multi-classifier; support vector machine; BTSVM algorithm

### 引言

支撑向量机是20世纪90年代在统计学习理论的基础上发展起来的一种完全不同的解决分类和回归问题的方法,它以结构风险最小化为目标,因此支撑向量机具有良好的泛化推广能力,并且与样本的具体分布无关.但是,目前为止支撑向量机只能解决二类分类问题而不能解决多类分类问题.为了用支撑向量机解决多类分类问题,一般的做法是使用多个支撑向量机来构成一个多类分类器系统.

基于支撑向量机的多类模式识别与分类系统的构造是一个值得研究的问题.构造一个有效的基于支撑向量机的多类分类器,应该要解决以下一些问题:一、选择一个合适的多类分类器结构,合理安排

多类分类器的节点与分支;二、确定决策节点的决策规则 and 使用的特征;三、使用尽可能少的决策节点,以得到最少的学习时间和最快的识别速度;四、要有一种快速的学习算法.目前,基于支撑向量机的多类模式识别与分类系统主要有以下类型:一对一方法(one versus one),为了分类 $N$ 类,需要训练 $N(N-1)/2$ 个支撑向量机,在这些支撑向量机中,总有一个支撑向量机将任意的二类区别开来,该方法有二种实现机制,一种是文献<sup>[1]</sup>提出的按多数决策方法,另一种是文献<sup>[2]</sup>提出的有向非循环图方法:一对多方法(one versus rest),在这种方法中,为了分类 $N$ 类,需要训练 $N$ 个支撑向量机,每一个支撑向量机都将某一类与其它的 $N-1$ 类区别开来<sup>[3]</sup>;纠错码方法,这种方法将输出进行编码,得到判断结

收稿日期:2003-05-29, 修回日期:2004-03-29

基金项目:国家自然科学基金资助项目(60072029, 60271033)

作者简介:方勇(1974-),男,湖南武冈人,博士研究生,主要从事图像处理、视频工程与多媒体通信研究.

Received date: 2003-05-29, revised date: 2004-03-29

果.

在上述方法中,都存在着一些不足之处:需要训练的支撑向量机个数太多,或分类未知样本时使用的支撑向量机过多,或者存在着过拟合的问题.因此,本文提出了一种基于二叉树结构的多类分类器,对于  $N$  类分类问题,它只需要  $N-1$  个支撑向量机;其次,在识别未知样本时只需使用较少的支撑向量机;最后,提出该多类分类器的学习算法.

## 1 多类分类器的构造及泛化能力分析

### 1.1 多类分类器的构造

分类树在多类模式识别中得到了广泛的应用,单个支撑向量机只适合两类分类问题,因此,要将支撑向量机用于多类分类问题,采用二叉树的结构来构造一个多类分类器是可行的.

定义 1:如果一棵二叉树具有  $N$  个叶结点且叶结点符合某种分布  $P_i, i=1, 2, \dots, N$ , 称树的期望高度为  $H(T) = \sum_{i=1}^N H(i)P_i$ , 其中,  $H(i)$  为叶结点  $i$  的高度. 如果该分布为均匀分布, 即  $P_i = \frac{1}{N}$ , 则树的期望

$$\text{高度为 } H(T) = \frac{1}{N} \sum_{i=1}^N H(i).$$

定理 1:如果一棵二叉树具有  $N$  个叶结点且符合均匀分布,在最佳情况下,树的期望高度为  $\log_2^N$ ; 在最坏情况下,树的期望高度为  $\frac{N}{2} - \frac{1}{2N}$ ; 一般情况下,树的期望高度处于二者之间.(证明略)

二叉树相对于其它的基于支撑向量机的多类分类器体系结构来说,具有结构简单,所需的支撑向量机个数少的优点;其次,期望高度小,识别速度快.

定义 2:已知空间  $H$ 、二值函数集合  $F = \{f: H \rightarrow \{-1, +1\}\}$ , 如果我们称一棵二叉树为多类分类器,则多类分类器的假设函数集就可以用一棵二叉树来实现,其中,二叉树的每个内结点用  $F$  中的一个元素来标识,叶结点用一个样本集类别来标识.

在多类分类器中,内结点又称为决策结点.设多类分类器的决策函数为  $D(X)$ , 为了计算一个样本  $X$  的决策函数值,从根结点开始,计算当前结点的二值函数,结果函数值为  $+1$ , 则将当前结点的左子结点设为当前结点;否则,将当前结点的右子结点设为当前结点,一直进行下去,直到到达一个叶结点为止,叶结点的标识值即为决策函数  $(D(X))$  的值.

### 1.2 多类分类器的泛化能力分析

分类器的泛化能力一般是用一致收敛性的界来

衡量的,使实际风险最小化.常用的概念就是“容量控制”,具体的衡量参数就是函数集所具有的 VC 维. VC 维越小,结构风险就越小,经验风险与实际风险就越趋向一致.

定理 2:假设多类分类器具有  $K$  个决策结点,且能够将  $m$  个样本正确地分为  $K+1$  类,设每个决策结点的分类间隔为  $\gamma_i, i=1, 2, \dots, K$ , 则多类分类器的泛化误差以大于  $1-\delta$  的概率小于

$$\frac{130R^2}{m} \left\{ D' \log(4em) \log(4m) + \log \frac{(4m)^{K+1} \binom{2K}{K}}{(K+1)\delta} \right\} \quad (1)$$

其中,  $D' = \sum_{i=1}^K \frac{1}{\gamma_i}$ ,  $\delta$  为 0 到 1 之间的一个常数,  $R$  为包含样本的超球半径<sup>[4]</sup>.

从定理 2 可以得知,对于特定的训练样本集,多类分类器的实际风险只与决策结点个数  $K$ 、分类间隔平方倒数和  $D'$  有关. 决策结点个数越少,多类分类器的泛化误差越小;分类间隔平方倒数和越小,多类分类器的泛化误差越小.因此,在设计多类分类器时,在正确分类的前提下,使用尽可能少的决策结点;其次,在每个决策结点,分类间隔要最大化;最后,要使用一种“最优”的二叉树结构,使分类间隔平方倒数和  $D'$  最小化.对于具体的问题,样本类别数  $N$  是确定的,因此,多类分类器的决策结点数也是确定的,没有选择的余地,且  $K=N-1$ . 在确定多类分类器的决策结点数后,要考虑的问题就是怎样使分类间隔平方倒数和  $D'$  最小化,也就是怎样来确定多类分类器的最优结构及最大化每个决策结点的分类间隔.

## 2 多类分类器的学习算法

### 2.1 特征空间的样本距离计算

支撑向量机将样本映射到特征空间时,是一种非线性映射,该映射并不能保证样本之间的马氏距离不变,也就是说,样本之间的距离关系在特征空间中不一定成立.因此,样本之间的距离度量应该在特征空间进行,而不能在样本空间中进行.在特征空间中,设任意两个样本之间的距离为  $M(X_i, X_j)$ , 则

$$\begin{aligned} M^2(X_i, X_j) &= \|\Phi(X_i) - \Phi(X_j)\|^2 \\ &= (\Phi(X_i) - \Phi(X_j))^T (\Phi(X_i) - \Phi(X_j)) \\ &= \Phi(X_i) \cdot \Phi(X_i) + \Phi(X_j) \cdot \Phi(X_j) - 2\Phi(X_i) \cdot \Phi(X_j) \\ &= K(X_i, X_i) + K(X_j, X_j) - 2K(X_i, X_j). \end{aligned} \quad (2)$$

通过式(2),可以把任意两个样本在特征空间中的距离转化为核函数的计算,不需要显式计算映射  $\Phi$ . 在具体计算时,可以预先计算每个样本在特征空间的点积  $K(X_i, X_j)$ ,并保存下来. 在计算样本的特征空间距离时,要多次用到这些数据,这样可以节省计算量.

## 2.2 最优分类器结构的确定

在 BTSVM 算法中,多类分类器结构由算法根据样本集模式类的分布来自动构造. 这就有两个问题需要解决:首先,如何在决策结点将多个模式类转换成两个子集;其次,如何评价分类器结构的优劣.

设在决策结点  $i$ ,输入样本集具有  $N$  个模式类:  $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_N\}$ ,为了适合支撑向量机处理,我们必须把  $N$  个模式类分成两子集:  $\Omega_{+1}$  和  $\Omega_{-1}$ ,其中,  $\Omega_{+1} \cap \Omega_{-1} = \Omega$ ,  $\Omega_{+1} \cap \Omega_{-1} = \phi$ ,且  $\Omega_{+1}$  和  $\Omega_{-1}$  都为  $\Omega$  的非空真子集. 当获得两个子集  $\Omega_{+1}$  和  $\Omega_{-1}$  后,就可用来训练一个支撑向量机. BTSVM 算法构造分类器时,决策结点的不同的聚类结果,对应着不同的分类器结构形式,这就需要一种分类器结构优劣的评价准则.

在特征空间中,样本是线性可分的. 所以,在 BTSVM 算法中,我们用两个子集中相距最近的样本的距离来衡量两个子集的相异程度:

$$\gamma(\Omega_{+1}, \Omega_{-1}) = \min_{X_i \in \Omega_{+1}, X_j \in \Omega_{-1}} M(X_i, X_j). \quad (3)$$

这个相似性度量与支撑向量机学习算法是相一致的,因为支撑向量机只与支撑向量有关,而与样本分布无关,也就是说,只与边界样本有关,因此采用两个子集之间的最小样本距离来衡量子集的相异程度,是合理的、可行的. 在这种情况下,寻求最优分类器结构的问题,就可转化为如下的一个优化问题:

$$\arg \min_D \sum_{i=1}^A \frac{1}{\gamma^2(\Omega_{i,+1}, \Omega_{i,-1})}. \quad (4)$$

其中,  $D$  为多类分类器,  $i$  为决策结点编号,  $\Omega_{i,+1}$  和  $\Omega_{i,-1}$  为该结点对应的两个子集,最小值对应的分类器结构即为最优结构. 对于小规模模式类问题,我们可以把所有可能的分类器结构形式找出来,计算每个分类器结构形式的目标函数值,进行评价;但是,对于大规模模式类问题,这种方法不适用. 因此,在本文中,我们提出了一种近似的最优分类器结构生成算法. 该算法是通过在每个决策结点  $i$  寻求  $\gamma(\Omega_{i,+1}, \Omega_{i,-1})$  最大化来实现的. 这是一种亚优算法,可能得到的是一个亚优分类器结构.

在采用式(3)作为相似性度量的情况下,在每个决策结点,我们用聚类的方法来获得两个近似最

优子集. 其基本过程如下:

算法 1:

1) 给定  $N$  个模式类的样本集,最初把每个模式类看成一个聚类  $\omega_i = \{\Omega_i\}$ ,设聚类数为  $c = N$ ;

2) 当  $c > 2$  时,重复以下操作:

2.1 根据最小距离确定两个最相近的聚类  $\omega_i$  和  $\omega_j$ ,

2.2 合并  $\omega_i$  和  $\omega_j$ ,得到一个新的聚类  $\omega_y = \{\omega_i, \omega_j\}$ ,从而得到一个聚类数为  $c - 1$  的聚类解,

2.3 将  $c$  值减 1.

根据以上算法,可以将  $N$  个模式类聚类成两个子集. 在具体实现时,由于任意两个模式类之间的最小距离要多次使用,我们可以在聚类算法开始之前,将任意两个不同模式类之间的最小距离计算出来并保存,以减少计算量.

## 2.3 BTSVM 算法

在 BTSVM 算法中,当确定了分类器的结构后,就应该确定每个决策结点的最优分类面. 正如前面所讨论的,每个决策结点是用一个支撑向量机实现的. 使用支撑向量机,可以保证在既定的分类器结构下,单个决策结点的分类间隔最大. 基于上面的分析,我们给出 BTSVM 算法的具体过程:

算法 2:

1) 令多类分类器  $D$  的初始状态只含有一个树根,即一个新结点( $\Omega$ ),其中  $\Omega$  为全体训练样本的集合;

2) 若  $D$  的所有新结点( $\Omega'$ )只含有一个模式类,即所有的训练样本都属于同一个类,则用这些新结点所对应的模式类标志这些新结点为叶结点,并停止执行学习算法,即得学习结果;

3) 否则,选取一个不具备步骤(2)所述状态的新结点( $\Omega'$ ),并标志该新结点为决策结点;

4) 应用算法 1 将该决策结点的训练样本集  $\Omega'$  聚类成两个子集  $\Omega'_{+1}$  和  $\Omega'_{-1}$ ;

5) 用支撑向量机的学习算法求得该决策结点的最优分类面;

6) 从该决策结点伸出 2 个分叉,每个分叉代表该决策结点的一个不同决策函数值,从而形成两个新结点:  $(\Omega'_{+1})$  和  $(\Omega'_{-1})$ ;

7) 转步骤(2).

下面,我们来分析 BTSVM 算法的时间复杂度,包括训练时间和识别时间. 在文献[5]报道了一个经验法则:支撑向量机的训练时间与训练样本集的大小成超线性关系,  $T = cn^a$ ,其中,  $c$  为比例常数,  $a$

为一常数,对于基于分解方法的支撑向量机学习算法来说, $\alpha \approx 2$ .

设样本集大小为  $n$ , 模式类别数为  $N$ , 对于标准的一对多方法, 需要训练  $N$  个支撑向量机, 则总的训练时间为

$$T_{1 \rightarrow r} = cNn^\alpha. \quad (5)$$

对于一对一方法, 需要训练  $N(N-1)/2$  个支撑向量机, 假设每个模式类所包含的训练样本数目相等, 则每个支撑向量机的训练样本数为  $2n/N$ , 总的训练时间为

$$T_{1 \rightarrow 1} = c \frac{N(N-1)}{2} \left(\frac{2n}{N}\right)^\alpha \approx 2^{\alpha-1} cN^{2-\alpha} n^\alpha. \quad (6)$$

对于 BTSVM 算法, 一般情况下,  $n \gg N$ , 故分类器构造时间会远远少于支撑向量机的学习时间, 因此分类器构造时间忽略不计. 假设分类器的结构

为完全二叉树结构形式, 每个模式类包含的训练样本数目相等, 支撑向量机的个数为  $N-1$ , 高度  $H(i)$  的决策结点的支撑向量机的训练样本数为  $n/2^{H(i)-1}$ , 总的训练时间为

$$T_{\text{BTSVM}} = \left(\frac{1 - 2^{\log_2(1-\alpha)}}{1 - 2^{1-\alpha}}\right) cn^\alpha. \quad (7)$$

从训练时间来看, BTSVM 算法要明显优于一对多方法, 而与一对一方法大致相当.

从识别时间来看, 在计算一个未知样本时, 一对多方法使用  $N$  个支撑向量机, 一对一方法使用  $N(N-1)/2$  个支撑向量机; BTSVM 算法使用的支持向量机数目与多类分类器的期望高度一致; 其次, 由于 BTSVM 算法采用了最优分类器结构, 在一定程度上降低了支撑向量机的复杂度.

表 1 USPS 手写数字数据集实验参数及实验结果

Table 1 Experimental results of USPS handwritten digit data set

算法	$\sigma^2$	C	准确率(%)	支撑向量(个)	训练时间(s)	测试时间(s)
一对多方法	10.00	100	94.9	3 365	2429.31	47.61
投票表决方法	83.33	100	95.1	2 472	261.15	39.03
有向非循环图方法	83.33	100	95.1	2 472	262.03	38.53
BTSVM 算法	500.00	100	95.1	2 178	152.42	23.89
神经网络方法			94			

表 2 UCI 英文字母数据集实验参数及实验结果

Table 2 Experimental results of UCI letter data set

算法	$\sigma^2$	C	准确率(%)	支撑向量(个)	训练时间(s)	测试时间(s)
一对多方法	0.31	100	97.6	9 391	399.66	40.16
投票表决方法	0.31	100	97.7	8 724	112.70	29.67
有向非循环图方法	0.33	100	97.6	8 603	114.84	19.69
BTSVM 算法	1.00	100	97.6	2 224	13.65	18.91
神经网络方法			95.7			

表 3 UCI 森林覆盖类型数据集实验参数及实验结果

Table 3 Experimental results of UCI covertype data set

算法	$\sigma^2$	C	准确率(%)	支撑向量(个)	训练时间(s)	测试时间(s)
一对多方法	1.00	10	72.2	10 085	1 443.50	2 651.67
投票表决方法	1.00	10	73.2	9 712	118.23	2 681.38
有向非循环图方法	1.00	10	73.3	9 712	135.89	2 350.31
BTSVM 算法	1.00	10	73.2	9 627	108.84	2 306.77
神经网络方法			70			

表 4 人脸数据集实验参数及实验结果

Table 4 Experimental results of face data set

算法	$\sigma^2$	C	准确率(%)	支撑向量(个)	训练时间(s)	测试时间(s)
一对多方法	125.00	100	98.6	447	28.44	2.52
投票表决方法	100.00	100	98.6	360	7.22	2.16
有向非循环图方法	100.00	100	98.6	360	7.20	2.17
BTSVM 算法	1 000.00	100	98.6	320	5.23	1.81

### 3 实验结果与讨论

本文中,我们采用4个数据集来评估算法的性能:USPS 手写数字数据集、UCI 英文字母数据集、UCI 森林覆盖类型数据集和一个小规模的人脸数据集。USPS 手写数字数据集包括10个模式类,即数字字符0到9;它包括7291个训练样本和2007个测试样本,每个样本作为一个256维的特征向量,每个特征被归一化为 $[-1, +1]$ 。UCI 英文字母数据集包括26个模式类,即英文字母A到Z;它包括16000个训练样本和4000个测试样本,提取每个样本的16个统计特征作为特征向量,输入是样本图像的统计值,同样,每个特征的值也被归一化为 $[-1, +1]$ 。UCI 森林覆盖类型数据集包括7个模式类,即有7种森林覆盖类型;它包括12600个训练样本和548612个测试样本,每个样本的输入特征向量包括10个被归一化为零均值、单位方差的连续特征和2个 $n$ 选1方式的离散特征,整个特征向量的维数为54。人脸数据集包括20个模式类,每个模式类包括大约32幅图像,对应着同一个人不同的表情、方向及是否戴眼镜等情形;我们随机的选取每个模式类的24幅图像作为训练样本集,其余的作为测试样本集,即训练样本集大小为480,测试样本集大小为144;每幅图像都被归一化为 $30 \times 32$ 大小且像素值为 $[-1, +1]$ 之间的图像。实验结果如表1~表4所列。

实验表明,BTSVM 算法的整体性能要优于其它的方法。分类准确率各种方法大致相当,但是,在训练时间与测试时间上,BTSVM 算法优于其它的方法。实验中,我们用识别相同测试集所需的时间来衡量分类器的识别速度,从实验结果可以看出,BTSVM 算法的识别速度在各个数据集上明显要优于其

它的方法;从训练时间来看,BTSVM 算法也优于其它的基于支撑向量机的方法。

### 4 结论

基于支撑向量机的多类模式识别问题是一个值得研究的问题,本文提出了一种新的基于支撑向量机的多类分类器结构形式。它把 $N-1$ 个支撑向量机通过二叉树的形式组合成一个多类分类器;同时,本文也提出了分类器的学习算法:BTSVM 算法,该算法不仅可以自动确定分类器的近似最优结构,而且还可以确定每个决策节点的最优分类面。实验结果表明,该算法具有训练时间短、识别速度快、分类准确率高、泛化能力好等优点,尤其是大规模模式类识别问题。但是,如何寻求一个全局最优解还有待深入研究。

### REFERENCES

- [1] Friedman J H. Another approach to polychotomous classification [R]. Technical report, Department of Statistics, Stanford University, 1997.
- [2] Platt J H, Cristianini N, Shawe-Taylor J. *Large Margin DAGs for Multiclass Classification* [M]. Advances in Neural Information Processing Systems 12, Cambridge MIT Press, 2000, 547—553.
- [3] Mayoraz E, Alpaydin E. Support vector machine for multiclass classification [C]. *Proceedings of International Workshop on Artificial Neural Networks*, Spain: Alicante, 1999.
- [4] Bennet K P, Cristianini N, Shaue Taylor J, et al. Enlarging the margins in perceptron decision trees [J]. *Machine Learning*, 2000, 41(3): 295—313.
- [5] Platt J C. *Fast Training of Support Vector Machines using Sequential Minimal Optimization* [M]. Advances in Kernel Methods: Support Vector Learning, Cambridge: MIT Press, 1999, 185—208.