# 支持向量机解决多分类问题研究

郑勇涛　刘玉树
(北京理工大学计算机科学与工程系,北京 100081)
E-mail:zhengyt@sina.com.cn

**摘　要**　支持向量机(SVM)是建立在统计学习理论基础上的一种小样本机器学习方法,用于解决二分类问题。但在解决实际问题中遇到的多为多分类问题,通过研究现有提出的一些支持向量机多分类的方法,并进行分析比较,在一对一分类方法基础上提出具有容噪声的分类方法,通过标准数据集实验加以验证。

**关键词**　支持向量机　多类分类　统计学习理论

**文章编号** 1002-8331-(2005)23-0190-03　**文献标识码** A　**中图分类号** TP391.6

# An Analysis of Multi-class Support Vector Machines

**Zheng Yongtao　Liu Yushu**
(Department of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081)

**Abstract**: Support Vector Machines(SVM) are developed from the theory of limited samples Statistical Learning Theory (SLT) by Vapnik et al., which are originally designed for binary classification. However, many practical problems are multi-class. How to extend it for multi-class is a research issue. This paper analyzes and compares some provided solutions of multi-class SVM. And the paper puts forward the algorithm of a noise insensitive SVM multi-class classifier and the algorithm is verified through international standard databases.

**Keywords**: Support Vector Machine, multi-class, Statistical Learning Theory

## 1　引言

统计学习理论(Statistical Learning Theory,SLT)是一种专门研究小样本情况下机器学习规律的理论。V.Vapnik 等人从六七十年代开始致力于此方面研究,到 90 年代中期,随着其理论的不断发展和成熟,统计学习理论开始受到越来越广泛的重视。统计学习理论是建立在一套较坚实的理论基础之上的,为解决有限样本学习问题提供了一个统一的框架。同时,在这一理论基础上发展了一种新的通用学习方法——支持向量机(Support Vector Machine,SVM),它已初步表现出很多优于已有方法的性能[1]。

SVM 最初是为解决二分类问题而设计的,不能直接用于解决多分类问题。而在实际应用中遇到的多为多分类问题,目前已经有许多算法将 SVM 推广到多分类问题应用中,如文本(超文本)分类、图像分类、生物序列分析和手写字符识别等[2]。

## 2　支持向量机

支持向量机主要思想是建立一个超平面作为决策曲面,使得正例和反例之间的隔离边缘被最大化。如图 1 所示,实心点和圆圈分别代表两类样本,$H$ 为分类线,$H_1$,$H_2$ 分别代表各类中离分类线最近的样本且平行于分类线 $H$ 的直线,它们之间的距离称为分类间隔(margin)。所谓最优分类线就是要求分类线不但能将两类正确分开且分类间隔最大。同理,在多维空间假定训练数据可以被一个超平面分开,如果这个向量集合能被超平面没有错误地分开,并且离超平面最近的向量与超平面之间的距离最大,则称这个向量集合被这个最优超平面(Optimal Separating Hyperplane,OSH)最大分现开[1,3]。

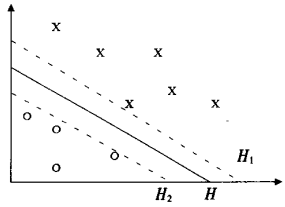


**图 1　线性可分条件下的最优分类线**

给定样本集:

$x_i \in R^n, y_i \in \{-1,1\}, i=1,\cdots,l$

$y_i[(w \cdot x_i)+b]-1 \geq 0 \quad i=1,\cdots,n$ (1)

分类间隔等于 $2/\|w\|$,$H_1$、$H_2$ 上的训练样本点就称作支持向量。利用 Lagrange 优化方法可以把上述最优分类面问题转化为其对偶问题,即在约束条件 $\sum_{i=1}^{n} y_i \alpha_i = 0$ 和 $\alpha_i \geq 0, i=1,\cdots,n$ 下对 $\alpha$ 求解下列函数的最大值:

$$Q(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) \quad (2)$$

$\alpha$ 为与每个样本对应的 Lagrange 乘子,求解对应的样本就是支持向量,得到最优分类函数:

$$f(x)=\mathrm{sgn}\{(w^* \cdot x)+b^*\}=\mathrm{sgn}\left\{\sum_{i=1}^{n} a_i^* y_i (x_i \cdot x)+b^*\right\} \quad (3)$$

**基金项目**:国家"十五"部委预研项目资助

**作者简介**:郑勇涛(1979-),男,汉族,山东菏泽人,硕士研究生,主要研究方向:数据挖掘,支持向量机。刘玉树(1941-),男,汉族,山东临沂人,教授,博士生导师,主要研究方向:人工智能、多媒体技术、分布式信息系统。

对于非线性问题，可以通过非线性变换转化为某个高维空间中的线性问题，在变换空间最优分类面。选用合适的核函数 $K(x_i,x_j)$ 满足 Mercer 条件变换到高维空间，目标函数(2)变换为：

$$Q(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j) \tag{4}$$

相应的分类函数变为：

$$f(x)=\mathrm{sgn}\{(w^*\cdot x)+b^*\}=\mathrm{sgn}\left\{\sum_{i=1}^{n}\alpha_i^* y_i K(x_i,x)+b^*\right\} \tag{5}$$

SVM 有以下几个特点：

(1)构造 SVM 能够解决一些其他方法无法解决的问题；

(2)构造 SVM 可加快学习过程；

(3)构造决策规则的同时，得到了支持向量；

(4)新的决策函数只通过改变定义特征空间的核函数即可实现。

## 3 现有多分类支持向量机算法

将 SVM 推广解决多分类问题有两类方法，第一种方法是将多分类看作二分类的组合，最终将多分类问题转化为二分类问题，第二种方法是通过修改目标函数，从根本上解决 SVM 处理多分类问题。由于后者代价过高，只适用于小规模问题，目前多采用第一类方法。对于 $k$ 类问题，给定样本集 $(x_1,y_1),\cdots,(x_l,y_l),x_i\in R^n,i=1,\cdots,l,y_i\in\{1,\cdots,k\}$。

### 3.1 一对多 SVM 分类(One-against-the rest)

一对多 SVM 分类是最为简单，也是最为普通的实现方案。对于 $k(k\geq 2)$ 类 SVM 分类问题，把类 1 作为一类，其余 $k-1$ 类视为一类，自然地将 $k$ 分类问题转化为二分类问题。这种分类方法在训练过程中，每个分类函数都需要所有的样本参与。分类函数为：

$$f(x)=\arg\max_{j\in\{1,\cdots,k\}}\sum_{i=1}^{n}(\alpha_i^j y_i^j K(x,x_i^j)+b^j) \tag{6}$$

上标表示第 $j$ 个 SVM 分类器的决策函数，$\alpha_i^j$ 和 $y_i^j$ 分别为第 $j$ 个支持向量的参数和类别标号，$b^j$ 为偏移量。对待测样本，若：

$$f^l(x)=\max_{j\in\{1,\cdots,k\}}f^j(x) \tag{7}$$

则输入样本属于第 $l$ 类。这种方法的训练时间与类别的数量成正比，并未考虑多个分类器对测试错误率的影响，当训练样本较大时，训练较为困难[6]。

### 3.2 一对一 SVM 分类(One-against-one)

一对一的解决方法是在 $k$ 类问题中进行两两组合，构造 $C_k^2$ 个分类器，该方法也称作 Pairwise Method。这种方法的确定是对于类别 $k$ 过大时，产生的子分类器过多，相对于一对多分类子分类器明显增加，训练时间更长。由于测试时要对任意两类进行比较，训练速度随着类别的增加成指数倍降低。

### 3.3 有向无环图 SVM 分类(Directed Acyclic Graph)

有向无环图 SVM 分类在训练阶段也是采用一对一 SVM 的任意两两组合训练方式，同样也需要构造 $C_k^2$ 个子分类器，但是在分类过程中，DAG 将所用子分类器构造成有向无环图如图 2，包括 $C_k^2$ 个节点和 $k$ 个叶子，其中每个节点是一个子分类器。当对未知样本训练时，从根节点开始分类，只需 $k-1$ 步即可完成分类[5]。和一对一 SVM 分类相比，在分类过程中减少了重复操作，大大提高了分类速度。这种分类方法的缺点是未考虑样本不平衡数据对分类速度的影响，而且没有考虑分类错误传递对后续产生影响。

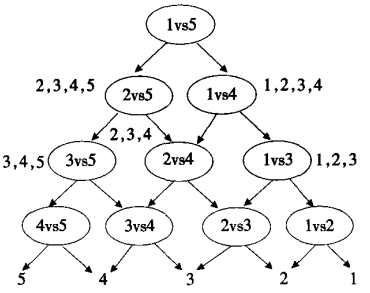


图 2 采用 DAG 5 分类决策过程

### 3.4 多分类 SVM 分类(Multi-class SVM)

该方法通过修改目标函数，把多分类问题转换为解决单个优化问题，从而建立 $k$ 分类支持向量机。由二分类支持向量机推广可得：

$$\min_{w,\xi,b}\Phi(w,\xi)=\frac{1}{2}\sum_{m=1}^{k}\|w\|^2+C\sum_{i=1}^{l}\sum_{m\neq y_i}\xi_i^m\geq \omega_{y_i}^T\phi(x_i)+b_{y_i}\geq \omega_m^T\phi(x_i)+b_m+2-\xi_i^m \tag{8}$$

$$\xi_i^m\geq 0,I=1,\cdots,l\quad m\in\{1,\cdots,k\}\backslash y_i$$

最终得到目标决策函数为：

$$f(x)=\arg\max_{m=1,\cdots,k}[\omega_m^T\phi(x)+b_m] \tag{9}$$

利用 Lagrange 优化方法将求最优分类面问题转换为其对偶问题最终得以解决[4]。这种方法可以一步完成，训练方法多采用块算法或 SMO 算法。缺点是训练时间相对前面几种方案较慢，适用于样本数量规模较小的问题的求解。

## 4 实验及分析

支持向量机是一种处理二分类问题的方法。上节讨论了目前处理多分类问题的方法。本文以一对一方案为基础，依据 VC 维置信范围设计了多值分类器。设训练样本集共有 $N(N>2)$ 种类别，这 $N$ 类训练样本两两组合，则共创建 $M=N(N-1)/2$ 个训练集，使用具有容噪性能的支持向量机二值分类算法对 $M$ 个训练集进行学习，生成 $M$ 个子分类器，产生 $M$ 个分类输出 $a_{ij}$，$M$ 个分类输出构成了输出矩阵 $A[a_{ij}]_{N\times N},a_{ij}=-a_{ij},a_{ii}=0$，因此在求解输出矩阵 $A$ 的各个元素时，只需要求解上三角或下三角阵即可。在支持向量机理论中，类别决策函数应满足经验风险最小(值为 0)和 VC 置信范围最小。由于每个子分类器训练集不同，造成决策函数分类精度也会不同，致使它们的 VC 维置信范围也相应地会不同。VC 维置信范围越小，决策函数误差也越小。实际操作过程中就是求得函数 $\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\xi_i+\gamma\sum_{j=1}^{2}d_i$ 最小。

用 Iris 标准数据集进行实验，其数据是用来测试机器学习的 UML 标准数据库 (http://www.ics.uci.edu/~mlearn/MLRepository.html)。Iris 数据集中共有四个属性：sepal length，sepal width，petal length，petal width；包括三类数据：Versicolor (类 1)，Virginica(类 2)，Setosa(类 3)。在这三个类别中，类 1 与其

余的两类之间分类较为复杂。而类2与类1、3间,类3与类1、2间的分类比较容易,因为它们之间的分类界限较清楚,因而着重考虑第一种情况。取核函数为高斯核函数,参数为0.2,$C$=1 000,采用SMO方法进行训练。训练得到的支持向量数为15,训练的正确率为97.3%。采用BSVM算法(Chih-Jen Lin,BSVM 2.04)和SVM_light(Thorsten Joachims,SVM_light 3.5)用Iris数据集进行测试,核函数用高斯核函数,$C$=1 000。结果如表1所示,在这里主要考虑了准确率问题,迭代次数、运行时间和支持向量数没有在表1中体现。

**表1 BSVM与SVM_light测试结果**

| | BSVM | | SVM_light | |
|---|---|---|---|---|
| 训练样本数 | 75 | 25 | 75 | 25 |
| 测试样本数 | 75 | 125 | 75 | 125 |
| 准确率/% | 89.3 | 71.2 | 89.3 | 70.4 |

BSVM和SVM_light这两种分类器在训练集数很多时,均能够取得比较好的分类结果。当训练样本数为25时,这25个数据是从三个类中特意选择出来的,从表1中可以看出,它们的性能将大大下降。采用了具有容噪性能的支持向量机多值分类器,构造了三个分类器,分别是类1对类2、3,类2对类1、3,类3对类1、2。用上面相同的Iris样本进行测试,测试结果如表2所示。在样本很多的情况下,本算法和其它算法准确率方面没有什么明显的改善。但当样本数相对少,并且分布不够好的时候,本算法会好于其它的算法。

**表2 Iris测试结果**

| 训练样本数 | 测试样本数 | 准确率/% |
|---|---|---|
| 75 | 75 | 89.3 |
| 25 | 125 | 81.3 |

## 5 结束语

将SVM应用到特定的实际问题需要解决大量的设计问题。一旦确定核函数和优化条件,系统的关键就确定了。目前SVM已经在包括文本(超文本)分类、图像分类、生物序列分析和生物数据挖掘、手写字符识别等领域的应用取得了成功。这些实际问题大都是些多分类问题。对于大量类别进行分类时比如文本或网页分类时,采用简单的一对多SVM分类训练速度和分类速度较低,有向无环图SVM分类具有理想的训练速度,但分类速度较慢。

由于SLT理论和SVM方法尚处在发展阶段,很多方面尚不完善,许多理论目前还只有理论上的意义,尚不能在实际算法中实现;SVM方法中如何根据具体问题选择适当的核函数也没有理论依据。因此,对于支持向量机的研究还有很多工作要做。(收稿日期:2005年1月)

## 参考文献


1.张学工.关于统计学习理论与支持向量机[J].自动化学报,2001;26(1):32~42

2.李国正,王蒙,曾华军译.支持向量机导论[M].北京:电子工业出版社,2004-03

3.Vladimir N Vapnik.An Overview of Statistical Learning Theory[J].IEEE Transactions on Neural Networks,1999;10(5):988~999

4.Hsu C-W,Lin C-J.A Comparison of Methods for Multiclass Support Vector Machine[J].IEEE Transactions on Neural Networks,2002;(13):415~425

5.Boonserm Kijsirkul,Nitiwut Ussivakul.Multiclass Support Vector Machines Using Adaptive Directed Acyclic Graph[C].In:IEEE/INNS International Joint Conference on Neural Networks(IJCNN-2002),2002:980~985

6.Vojtech Franc,Vaclav Hlavac.Multi-class Support Vector Machine[C].In:16th International Conference on Pattern Recognition(ICPR'02),2002:236~239


(上接158页)

就会越小。所以,用户可以根据应用中实际的媒体流的播放质量和自己的要求来动态调整$D_{play}$的大小,以获得理想的上溯层数阀值$m$。最后,接收者A通过计算冗余RTP包(超时到达的重传包)产生的概率、最终丢失RTP包的概率,可以揭示自己的多媒体流回放特性同$D_{wait}$、$D_{play}$的关系。

## 7 结论

本文介绍了一种多媒体组播应用中的RTP包丢失恢复方案。通过缓冲路由器的辅助,配合以缓冲播放机制,使得基于重传的RTP包丢失恢复成为可能。该方案采用了重传请求抑制和丢弃机制,大大提高了网络资源的利用率。

目前实时多媒体流的组播应用大多都还限于局域网之内,而本文所提出的基于分层缓冲路由器的包丢失恢复方案,对今后关于实时多媒体组播在广域网中的应用这类研究工作有着一定的参考价值。

然而本文提出的基于缓冲路由器的方案也有不足之处,它对用户的地理位置分布有一定的要求,需要用户相对集中地分布在若干个区域内。否则,当有用户加入或退出组播组时,有可能需要在组播树中进行手工添加或取消缓冲路由器的操作。(收稿日期:2004年11月)

## 参考文献


1.Schulzrinne H,Casner S,Frederick R et al.RTP:A Transport Protocol for Real-Time Applications[S].RFC 1889,1996

2.Basso A,Varakliotis S.Transport of MPEG-4 over IP/RTP[C].In:IEEE International Conference,2000;(2):1067~1070

3.Transport of MPEG-4 Elementary Streams[S].RFC 3640,2003-11

4.G Feng,C K Siew,K L Yeung.Active resource allocation for active reliable multicast[J].IEE Proc-Commun,2003;150(2):69~79

5.Jaehe Yoon,Azer Bestavros,Ibrahim.Adaptive Reliable Multicast[J].IEEE,2000;3(6):1542~1546

6.Philip A Chou,Alexander E Mohr,Albert Wang et al.Error Control for Receiver-Driven Layered Multicast of Audio and Video[J].IEEE Transactions on Multimedia,2001;3(1):108~122

7.Quji Guo,Qian Zhang,Wenwu Zhu et al.A Sender-Adaptive & Receiver-Driven Layered Multicast Scheme for Video Over Internet.IEEE ISCAS,2001:141~144

8.Jun Peng,An Efficient and Scalable Loss Recovery Scheme For Video Multicast.http://www.rpi.edu/~pengj2/tmm04.pdf