# 摘　要

90年代初，一种新型的学习算法在原有统计学习理论的基础上被提了出来，即支持向量机——Support Vector Machine (SVM)。本文对SVM的方法与性质进行了详细的研究，同时针对多类识别问题的SVM算法提出了一些改进，并分别提出了两种基于两类与多类SVM方法的新型流量检测系统。

论文的主要研究内容包括理论研究与实际应用两部分：

1. 文章的理论研究分成了两个部分：
   1) 第一部分首先对统计学习理论进行了全面的回顾，然后对两类SVM的方法与性能进行了详细的阐述，同时对其相关知识，例如核方法，最优化问题等也进行了较全面的讨论；
   2) 第二部分则对多类SVM方法进行了深入的研究，并且在现有多类SVM方法的基础上提出了两种改进，同时也对改进的合理性以及可行性进行了说明。
2. 在实际应用部分里，文章创造性地将SVM方法与P2P流量检测系统进行了有效的结合，分别将两类与多类SVM方法运用到实际的P2P流量检测问题当中，取得了比现有的P2P流量检测方法更快的速度以及更高的精度，并用实验数据证明了SVM方法与该系统的成功融合。

**关键词：统计学习理论，支持向量机（SVM），核方法，多类问题，模式识别，流量检测**

# Abstract

In the early 90's, a new learning method had been proposed based on the statistical learning theory, called support vector machine (SVM). This thesis gives a detailed review of the SVM and proposes some improvements of the multiclass SVM. Then it proposes two new systems for traffic identification based on the binary SVM and multiclass SVM respectively.

The main research contents of this thesis include two parts: the theory part and the application part.

1. The theory part of this thesis also includes two parts:

   1) The first part gives a comprehensive survey of statistical learning theory and binary SVM. Then it gives a detailed discussion for some related knowledge like kernel methods and optimization problems.

   2) The second part makes an exhaustive review of the multiclass SVM and proposes two improvements based on the traditional methods. Then it demonstrates the rationality and the feasibility of the proposed methods.

2. In the application part, the thesis combines SVM and the P2P traffic identification system innovatively and efficiently. It uses binary SVM and multiclass SVM to the practical P2P traffic identification problems respectively and gains faster speed and higher accuracy than the traditional methods. The experimental results show the successful combination of SVM and the traffic identification system.


**Key Words: Statistical Learning Theory, Support Vector Machine (SVM), Kernel Methods, Multiclass Problems, Pattern Recognition, Traffic Identification.**

# 独 创 性 声 明

本人声明所呈交的学位论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表和撰写过的研究成果，也不包含为获得国防科学技术大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文题目：支持向量机的多类识别及其在流量检测问题中的应用

学位论文作者签名：刘涉　　日期：2006 年 11 月 28 日

# 学位论文版权使用授权书

本人完全了解国防科学技术大学有关保留、使用学位论文的规定。本人授权国防科学技术大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档，允许论文被查阅和借阅；可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密学位论文在解密后适用本授权书。）

学位论文题目：支持向量机的多类识别及其在流量检测问题中的应用

学位论文作者签名：刘涉　　日期：2006 年 11 月 28 日

作者指导教师签名：[illegible]　　日期：2006 年 11 月 28 日

# 第一章 绪 论

## 1.1 选题依据及课题产生的意义

统计学习理论的基本内容起源于二十世纪六、七十年代，它与传统统计学相比有着本质上的区别。传统统计学的核心思想是渐近理论，研究的是当样本数目趋于无穷大时事物的统计特性；而统计学习理论研究的是如何利用有限的样本及经验数据进行学习的一种理论[1,2]，因而更具有实效性。

九十年代初，一种新型的学习算法在原有统计学习理论的基础上被提了出来，即支持向量机——Support Vector Machine (SVM)[3,4]。该学习算法表现出了很多优于传统机器学习方法的性能。许多研究人员认为，SVM 已经成功的取代了神经网络，成为了机器学习领域最新的研究热点。

支持向量机方法的几个主要优点如下：

1. 它的基础是统计学习理论，针对的是有限样本的情况，其目标是得到有限信息下的最优解而不是样本数目趋于无穷大时的最优值，因此更具有实际应用价值；

2. 算法利用核函数将实际问题通过某种非线性变换映射到高维特征空间(Feature Space)之中，然后在高维特征空间中构造线性判别函数来替代输入空间中的非线性判别函数。这一点保证了该方法有较好的泛化性能，同时通过直接计算核矩阵巧妙的解决了“维数灾难”问题，使算法复杂度与样本维数无关；

3. 通过核变换以及最优化方法，算法最终将转化为一个二次凸函数求极值问题，从理论上说，得到的必将是全局最优点，解决了其他机器学习方法，例如神经网络中无法避免的局部极值问题。

目前，SVM 算法在模式识别、回归分析、函数估计、信号处理等领域都有着广泛的应用[5-6]。尤其是最近几年，在对于如何有效的解决手写体字符识别[7]、图像识别[8-10]、语音/音频识别[11,12]、人脸识别[13-15]、文本分类[16,17]、车辆追踪与检测[18,19]及其他[20-23]许多实际应用问题上，SVM 算法已经开始全面体现它的优势。

从以上提及的文献中可以了解到，国际上目前对 SVM 方法的研究正在不断深入，而国内研究人员在该领域的工作才起步不久。因此我们需要及时学习并熟练掌握有关理论，同时开展全面、有效的研究工作，使我们能够在这一机器学习的重要领域尽快赶上甚至超过国际先进水平。由于 SVM 是一门新兴学科，某些方面还有待完善，例如：许多理论目前还不能在实际问题中得以实现；而某些实际算法在理论上的解释也并非完美。因此，在这些方面可做的工作还有很多。

论文从理论和实际两个方面对 SVM 方法进行了改进和创新。在理论部分，本文提出

了两种基于现有多类 SVM 方法的改进；在实际应用当中，文章创造性地将 SVM 思想与 P2P 流量检测系统进行了有效的结合，分别将两类与多类 SVM 方法运用到实际的 P2P 流量识别与应用级分类当中，取得了很好的效果。本论文在学术上有一定的独创性，同时上述改进与创新也能应用于自主车等其他实际系统。

## 1.2 支持向量机的国内外研究历史、现状及发展趋势

基于经验数据的机器学习方法是现代智能系统中最重要的研究内容之一，它研究的是如何从经验数据中寻找规律，并利用这些规律对未知数据或无法观测的样本进行快速、准确的预测。目前机器学习的方法主要可以分为以下三种[78]：

第一种是经典的参数估计法。该方法的理论基础是传统统计学。在这种方法中，参数的形式必须是已知的，接下来需要做的工作仅仅是利用训练样本估计参数的值。不过该方法有相当大的局限性：首先，它需要已知样本的分布形式；其次，传统统计学所需要的前提条件是样本数目趋于无穷大，而这两点在现实问题中都很难办到。因此一些在理论上很完美的学习方法在实际应用中可能并不理想。

第二种方法是非线性方法。该方法利用经验数据建立了一个非线性模型，它在一定程度上克服了第一种方法的困难。但是，该方法显得过于随意，缺乏一种统一的数学理论。

第三种则是基于统计学习理论的方法。与传统统计学相比，统计学习理论是一种专门研究小样本情况下机器学习规律的理论，以它为基础建立的学习体系比基于传统统计学的方法更加全面。因为它在考虑了渐近性能的同时更加强调了如何在有限的信息下得到最优的结果。

V. Vapnik 等人从六、七十年代起开始致力于统计学习理论的研究。到九十年代中期，随着其理论的不断发展与成熟，同时也由于神经网络等传统机器学习方法在理论上缺乏实质性的进展，统计学习理论开始受到越来越广泛的重视。SVM 正是基于统计学习理论的一种新的机器学习方法，是由 Vapnik 及其合作者发明，在 1992 年计算学习理论的会议上介绍并进入机器学习领域[3]。1995 年，Vapnik 在文献[1]中完整的阐述了 SVM 方法。1997 年，Vapnik，Gokowich 和 A.Smola 在文献[5]中全面的介绍了 SVM 方法在回归估计与信号处理中的应用，而 Drucker 等人也在文献[24]中对 SVM 回归分析进行了讨论。就在同一年，B. Scholkopf 的博士论文[25]也对支持向量学习进行了完整而深入的研究。

随着研究的不断深入，SVM 方法潜在的研究价值吸引了众多国际著名学者的关注，近些年出现了许多新的关于 SVM 的文章，同时也有大批专家对 SVM 方法作出了全面的综述。

1998 年，C. Burges 在文献[26]中以模式识别为主要背景，对 SVM 的发展与核心内容进行了清晰的回顾。Cristianini 和 Taylor 于 2000 年出版的书籍[27]也对 SVM 及其相关知识进行了简要而全面的介绍。而 A.Smola 分别于 1996 年和 1998 年完成的硕士论文[28]和博士论文[29]也是两篇关于 support vector 回归和核方法学习的经典文献。

与此同时，许多对已有经典方法的改进以及新的 SVM 学习方法也陆续出现[30-38]。文献[30]通过对RBF核函数的改进提高了SVM的分类性能。文献[31]以模式识别问题为例子，通过与问题本身不变性的结合，提高了 SVM 的泛化能力与分类速度。[32]提出了一种新的通过在线回归算法训练 SVM 的思想。[33]-[36]都是对同一个主题：$v$-SVM 展开的研究。该方法通过对支持向量个数的约束，得到了一种新的 SVM 分类器。[33]与[34]对$v$-SVM 的理论基础和算法实现进行了详细的剖析，[35]与[36]则对参数$v$的选择进行了探讨。文献[37]与[38]研究的又是另一方面的内容。它们研究的都是核函数参数选取的问题。而[37]更是提出了一种自动选择最优核函数模型的方法，避免了人为选择的主观性。

由于 SVM 表现出来的优越性能，许多研究人员都希望能将其应用到各种各样的实际问题当中去。而事实上，SVM 在这许多的应用领域当中也的确收到了相当出色的效果[7-23]。与此同时，在研究如何利用 SVM 解决大规模数据集问题方面也取得了可喜的成果[39-44]。

就目前来说，SVM 是统计学习理论中最新的内容，也是最实用的部分，可用于模式识别、回归估计或函数逼近等方面，不但是当前人工智能和模式识别领域研究的热点，而且也已成为机器学习和数据挖掘领域的标准工具。现在对 SVM 的研究主要集中在以下几个方面：

1. 缩短训练时间——主要研究如何提高 SVM 的学习效率；
2. 参数选择——SVM 本身以及核函数映射中一些自由参数的选择问题；
3. 多类 SVM 研究——基于 SVM 的多类问题分类算法；
4. 应用研究——主要研究 SVM 在一些具体领域的应用。

综上所述，从统计学习理论中发展出的 SVM 技术已经发展成为机器学习中一个独立的子领域，在理论和实践两方面都有着光明的前景。

## 1.3 论文研究内容与组织结构

### 1.3.1 论文主要研究内容

本论文主要研究内容如下：

(1). 统计学习理论与两类 SVM 回顾

该部分将传统统计学与统计学习理论进行一个简要的比较，突出了统计学习理论的优势。同时对两类 SVM 的概念与性质进行了详细的研究，并对与其密切相关的核方法和最优化理论也进行了必要的讨论。

(3). 多类 SVM 算法的研究

该部分对多类 SVM 的理论进行了深入的研究，之后针对传统方法提出了两种改进的算法，并证明了改进算法的合理性以及可行性。

(4). 实际应用研究。

该部分对 SVM 在实际中的应用进行了尝试性的研究，分别利用两类和多类 SVM 提出

了两种新的 P2P 流量检测模型。

### 1.3.2 论文总体组织结构

根据文章所研究的主要内容，总体组织结构安排如下：

第一章　绪论。综述课题的选题依据与研究意义，国内外关于 SVM 的研究现状及发展趋势，并介绍本文的主要研究内容和总体组织结构。

第二章　统计学习理论及两类 SVM 方法研究。介绍 Vapnik 的统计学习理论并将其与传统统计学进行简单对比，突出了统计学习理论在实际应用中的优势。回顾了两类 SVM 方法的概念与性质，同时也用必要的篇幅介绍了核函数及最优化技术。

第三章　多类 SVM 及其改进算法研究。本章深入研究了多种多类 SVM 的方法，并且针对传统的 1-vs-1 和 1-vs-all 方法分别进行了有效的改进，最后通过仿真结果与真实数据证明了改进的合理性。

第四章　基于 SVM 算法的新型 P2P 流量检测系统设计。本章将两类与多类 SVM 方法运用到实际的 P2P 流量检测与应用级分类问题当中，建立了两个新的流量检测模型，取得了比现有的 P2P 流量检测方法更快的速度以及更高的精度，并通过实验数据证明了该结论。

第五章　结论与展望。对本课题所做的研究工作与所得结论进行总结，并对以后工作进行展望。

# 第二章 统计学习理论及两类SVM方法研究

让机器拥有人类的学习能力，并且通过机器所独有的运算速度与精度来完成人类力所不能及的事情，是许多科学家的梦想。因此，机器学习一直是一个非常热门的话题。而机器学习领域中的核心问题就是如何让机器有效的模仿人类的学习以及推理能力。换句话说，机器学习最主要的任务就是研究如何从已知数据（即样本集）中寻找规律，并利用这些规律对未知的或者无法观测的数据进行有效和准确的预测[78]。

对绝大多数已有的机器学习方法而言，传统统计学是它们的重要基础。从理论上来说，传统统计学有着一套严密且完备的理论体系，应该能够使得这些方法取得完美的结果。但遗憾的是，在众多实际应用问题当中，这些基于传统统计学的机器学习方法并没有取得令人信服的结果。主要原因有两个：

第一，由于传统统计学有一个大前提，那就是假定样本的分布形式是已知的，接下来需要做的工作仅仅只是利用已知数据对给定形式的参数进行简单的预测。但是在许多实际学习问题中，样本的分布形式是未知的，因此基于传统统计学的预测与真实值之间可能存在着相当大的偏差；

第二，传统统计学的另一个核心思想是渐近理论，它需要足够多的样本数目，而这在实际当中往往也难以办到，所以进一步导致了预测结果的不甚理想。

以上两个原因导致了以传统统计学为基础的机器学习方法在实际应用中的效果并不尽如人意。因此，一种新的理论——统计学习理论[1,2]诞生了。虽然它仍然建立在统计学的基础上，但是其核心思想却不再是渐近理论，恰恰相反，它是一种专门研究有限样本，甚至是小样本情况下机器学习规律的理论。同时它也不再需要先假定样本的分布形式。正是由于这两个特点，使得统计学习理论在许多实际问题中表现出了优于传统统计学的性能，从而迅速成为了研究有限样本情形下机器学习问题的最主要工具。

随着统计学习理论的不断发展，一种新的机器学习算法在此基础上被提了出来，即支持向量机—Support Vector Machine (SVM)[3,4]。该方法最早的出现是为了解决模式识别中的二值分类问题，而它在这类问题中表现出来的卓越性能迅速使其成为了当前研究的热点。

本章下面的几节将分别从学习问题的数学模型，结构风险极小化原则（SRM 原则），两类SVM方法以及核函数的相关性质等几个方面来展开详细的讨论。

## 2.1 学习问题的数学模型

一般的机器学习问题可以用如下的语言来进行阐述：已知一组输入和输出之间存在着某种未知的联系，让机器通过对其中一些输入输出对的学习，建立一个该未知联系的模型，能对新的输入做出尽量准确的输出预测。在统计学习理论中，这种联系被看作是某个未知

的联合概率分布 $F(\mathbf{x},y)$，而机器学习的目的就是通过一个独立同分布的样本集：

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),...,(\mathbf{x}_n,y_n) \tag{2.1}$$

在一个函数集 $H$：$\{f(\mathbf{x},\omega)\}$ 中选择一个“最优”的函数 $f(\mathbf{x},\omega_0)$ 进行预测，使得期望风险

$$R(f)=\int L(y,f(\mathbf{x},\omega_0))dF(\mathbf{x},y) \tag{2.2}$$

最小。其中 $H$：$\{f(\mathbf{x},\omega)\}$ 为预测函数集，$\omega$ 为函数中的待求参数，而 $L(y,f(\mathbf{x},\omega_0))$ 则为用 $f(\mathbf{x},\omega_0)$ 对 $y$ 进行预测时的损失函数。这里 $L(y,f(\mathbf{x},\omega_0))$ 是损失函数的一种广义形式，在针对具体问题的时候它都有着不同的表达式。例如在一般的模式识别问题和回归预测问题当中，损失函数有两种常见的形式：一种是

$$L(y,f(\mathbf{x},\omega_0))=|y-f(\mathbf{x},\omega_0)| \tag{2.3}$$

被称为线性损失函数；另一种是

$$L(y,f(\mathbf{x},\omega_0))=(y-f(\mathbf{x},\omega_0))^2 \tag{2.4}$$

被称为最小二乘损失函数。而两类模式识别问题中最简单的 0—1 损失函数可以算做是线性损失函数的一种特例。

在机器学习的过程当中，最终目的是使得期望风险(2.2)最小，但联合概率分布 $F(\mathbf{x},y)$ 的具体形式是未知的，已知的只有样本集(2.1)。而想仅仅通过样本集(2.1)中的 $n$ 个样本点就能找出使期望风险最小的 $f(\mathbf{x},\omega_0)$ 是不可能的，因此只能退而求其次。由于已知样本集(2.1)，因此可以计算 $f(\mathbf{x},\omega_0)$ 在这 $n$ 个样本点上的偏差，通过样本集上所谓的“经验风险”来对 $f(\mathbf{x},\omega_0)$ 的优劣进行评价。

**定义 2.1**（经验风险）设给定样本集(2.1)，并且给定损失函数 $L(y,f(\mathbf{x},\omega_0))$，函数 $f(\mathbf{x},\omega_0)$ 的经验风险是指

$$R_{emp}[f]=\frac{1}{n}\sum_{i=1}^{n}f(\mathbf{x}_i,\omega_0) \tag{2.5}$$

定义了经验风险之后，随之而来的就是经验风险极小化原则，简称为 ERM 原则：

**定义 2.2**（ERM 原则）[45]给定样本集(2.1)与函数集 $H$：$\{f(\mathbf{x},\omega)\}$，ERM 原则就是在 $H$ 中，选择使经验风险 $R_{emp}[f]$ 达到最小的函数 $f$。

ERM 原则强调的是将经验风险 $R_{emp}[f]$ 降到最小，但是经验风险毕竟和真实风险有所不同，过分地强调 ERM 将有可能导致所谓的“过学习”问题。而解决“过学习”问题的

关键就在于限制函数集$H$的范围，在限定了函数集$H$范围的前提下再通过ERM原则来求取合格的函数。

## 2.2 VC维与结构风险极小化原则

通过前面的介绍可以知道，必须对侯选函数集$H$的大小进行严格的限制。那么究竟应该通过一个什么样的指标来对在$H$的大小进行限制呢？下面将会引入这个指标，即VC维的定义。

在介绍VC维的概念以及函数集$H$的大小与VC维之间的关系之前，首先需要了解"打散"的概念。这里考虑的是两类模式识别问题，即$f(\mathbf{x},\omega)\in\{-1,1\}$。

**定义2.3**（打散）假设有$n$个输入点，分别用+1和-1来标记这些点，那么将有$2^n$种不同的标记法。如果对于每一种标记法，都能从函数集$H$中选出合适的函数来正确划分这种标记（正确划分是指将标记为+1的点和标记为-1的点完全分开），那么则称这$n$个输入点可以被函数集$H$打散。

有了打散的定义之后，VC维的定义就很自然地可以描述如下：

**定义2.4**（VC维）[45]如果一个函数集$H$能打散某个点集的点的最大数量为$n$，则称函数集$H$对应这个点集的VC维为$n$。

需要注意的是，如果一个函数集的VC维为$n$。那么至少存在某一个$n$个点的集合能被该函数集打散，但是并不是任意的$n$个点都能被该函数集打散。

下图2.1给出的是一个打散和VC维概念的具体例子。在$R\times R$平面上任意选取不共线的三点，由图可知，平面上的直线集可以将这三点完全打散。与此同时，很容易验证平面内任意四点都不可能被平面上的直线集打散。因此$R\times R$平面上的直线集能打散的最大点数为3，从而该直线集对应平面上点集（不共线）的VC维是就是3。

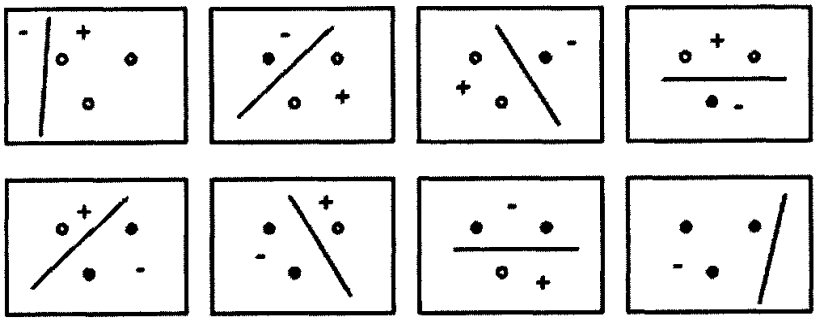


图2.1 $R^2$空间中的三个点（不共线），被平面上的直线集打散

在有了VC维的概念之后，接下来的问题自然是如何通过VC维这个指标来对侯选函数集$H$的大小，也就是容量进行严格的限制。首先不加证明的给出如下定理（定理的具体

证明参见文献[2]，[45]以及[46]）：

**定理 2.1** 设函数集$H$的VC维为$h$，训练样本的个数为$n$，若$h<n$成立，则对于任意的联合概率分布$F(\mathbf{x},y)$和任意的$\delta\in(0,1)$，函数集$H$中的所有函数$f$都可以使得下面的不等式至少以$1-\delta$的概率成立

$$R[f]\le R_{emp}[f]+\phi\left(\frac{h}{n},\frac{\log(\delta)}{n}\right) \tag{2.6}$$

其中$\phi$被称为置信范围，也可以称做置信区间，定义如下：

$$\phi\left(\frac{h}{n},\frac{\log(\delta)}{n}\right)=\sqrt{\frac{1}{n}(h(\log\frac{2n}{h}+1)-\log(\delta/4))} \tag{2.7}$$

不等式(2.6)的左边是在 2.1 节中提到过的期望风险，右边的第一项为经验风险，第二项为置信范围，这两项之和则被称为**结构风险**。

由(2.6)可以看出，结构风险是期望风险$R[f]$的上界。因此考虑通过结构风险来估计期望风险是合理的。当函数集$H$的较大的时候，即便可以选取到使$R_{emp}[f]$较小的$f$，但是由于$H$的VC维较大，从而使得置信范围也比较大，达不到尽量使结构风险减小的要求。反之，如果一味缩小函数集$H$，虽然会使得$H$的 VC 维较小，但是很可能导致较大的$R_{emp}[f]$，同样达不到预期的效果。所以应该两者兼顾，在其中取得一个合理的平衡。

与此同时，由式(2.7)可以看出，当且仅当$h/n$足够小的时候，$\phi$可以忽略不计，此时用经验风险来估计期望风险是合理的。但是当样本数目较少时，即使能选取到 VC 维较小的函数集$H$，也不能保证此时的$h/n$足够小，则此时的$\phi$不能忽略不计，如果仍然用经验风险来估计期望风险的话，效果将会大打折扣。

由于以上几方面的原因，统计学习理论中出现了一种新的准则——结构风险极小化原则（SRM 原则）[45,46]：

**定义 2.5**（SRM 原则）SRM 原则是在函数集$H$中选择一个函数$f$，使得不等式(2.6)的右端，也就是经验风险与置信范围的和达到最小。

图 2.2 形象地描述了 SRM 原则：图中的$S_i$ $(i=1,2,3)$表示一系列的函数集，并且满足$S_1\subset S_2\subset S_3$，在每一个$S_i$中都能够找到使得经验风险最小的函数$f_i$，随着$i$的增加，经验风险$R_{emp}[f_i]$单调递减而$S_i$的VC维单调递增。若选取的目标过小，则学习的力度不够，经验风险将会很大；若选取的目标过大，则容易造成过拟合现象，使得学习的结果不可信。因此当前的目标就是要寻找一个使得经验风险与置信范围之和最小的函数集$S_i$，并求得相

应的 $f_l$。

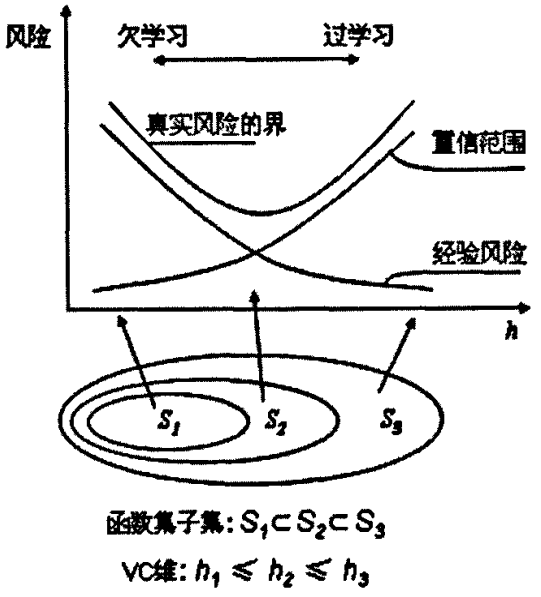



图 2.2 结构风险极小化原则（SRM 原则）

## 2.3 两类 SVM 与核函数研究

随着统计学习理论的不断发展，一种新的机器学习算法在原有基础上被提了出来，即支持向量机—Support Vector Machine (SVM)。这种新方法的核心思想就是上一节所提到的 SRM 原则，从几何学的角度理解的话，它是通过构造最大间隔超平面来同时满足分类精度与泛化能力。本节将从直观的几何角度入手，并结合最优化方法对 SVM 的核心思想进行分析，然后进一步讨论广义 SVM 以及由其衍生出来的核函数的内容。

### 2.3.1 最大间隔超平面与线性可分 SVM

首先引入一个最简单的两类模式识别问题：

如图 2.3，给出两类不同颜色的点，黑色与白色各代表一类样本点。考虑一条适当的直线，将这两类不同的点完全正确的分开。

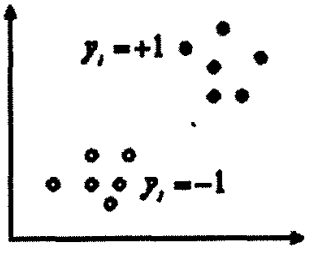
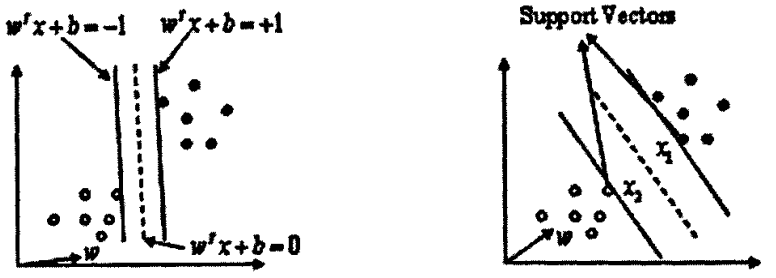



图 2.3 简单的两类分类问题

图 2.4 中的虚线代表构造的分类超平面。那么从图中可以明显的看出，无论是图 2.4(a)还是图 2.4(b)中的虚线都可以将图中的两类样本点完全正确的分开。然而根据 SRM 原则，在正确率相同的前提下，泛化能力成为了评判方法好坏的重要标准。那么究竟哪个分类平面有着更好的推广性能，也就是泛化性能呢？从直观上来说，样本点和分类平面隔得越远的话，“安全系数”就越大。基于这样的理解，图 2.4(b)的分类平面显然要比图 2.4(a)的分类平面离样本点更远，因此也更“安全”，即泛化性能更好。




(a) 小间隔超平面　　(b) 大间隔超平面

图 2.4 不同间隔超平面示意图

接下来的内容将通过数学语言来描述这种构造最大间隔超平面思想，以便对这种直观的理解给出严格的理论依据。

给定样本集$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),...,(\mathbf{x}_n,y_n)$，其中$\mathbf{x}_i\in R^n$，$y_i\in\{-1,+1\}$。求满足以下条件的超平面：能将$y_i=-1$的样本点和$y_i=+1$的样本点完全分开，并且使得两类样本点与超平面之间的距离达到最大。首先给出如下定义：

**定义 2.6** （线性可分）如果存在$\mathbf{w}\in R^n$与$b\in R$，使得对于样本集中所有的样本，都满足如下约束条件：

$$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geq 1 \tag{2.8}$$

则称样本集是线性可分的。

该不等式很容易通过对$\mathbf{w}$与$b$的调整而得到。这样的$\mathbf{w}$与$b$决定了一个分类超平面$\mathbf{w}\cdot\mathbf{x}+b=0$。图 2.4 中的两条虚线都是满足要求的超平面。根据解析几何中点到直线的距离公式，样本集中任意一点$\mathbf{x}_i$到该超平面的距离为：

$$d_i=\frac{\mathbf{w}\cdot\mathbf{x}_i+b}{\|\mathbf{w}\|} \tag{2.9}$$

联立(2.8)与(2.9)可得：

$$y_i d_i\geq\frac{1}{\|\mathbf{w}\|} \tag{2.10}$$

由不等式(2.10)可得知，$1/\|\mathbf{w}\|$是点到超平面的距离的下界。接下来的定义将在超平面与其参数之间建立一个一一对应的关系。

**定义 2.7** （超平面的标准形式）给定一个样本集与超平面$\mathbf{w}\cdot\mathbf{x}+b=0$，这里所有的超

平面都可以表示成一种统一的形式：即使得样本集中所有的点到超平面的距离中的最小值为 $1/\|\mathbf{w}\|$，那么在样本集中必定存在某些样本点使得不等式(2.10)能取到等号。该形式被称为超平面的标准形式。

在定义了超平面的标准形式以后，接下来自然应该给出最大间隔超平面的定义：

**定义 2.8** （最大间隔超平面）在所有标准形式的超平面中，使得最短距离，即 $1/\|\mathbf{w}\|$ 最大的超平面被称为最大间隔超平面。

根据上述定义，求取最大间隔超平面的问题将被很自然地转化为如下最优化问题。

**问题 1**：已知 $y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geq 1$，$i=1,2,...,n$，求 $\frac{1}{2}\|\mathbf{w}\|^2$ 的最小值。

注意到参数 $b$ 仅仅出现在约束条件中，而并没有出现在目标函数中。

最初的线性 SVM 思想就是来自于最大间隔超平面的方法。它首先找到两类样本点中距离最近的点，然后将这两类不同的点等距的分开。这种方法就被称为线性 SVM 分类法。

### 2.3.2 线性不可分 SVM 与核函数

在许多实际问题当中，简单的线性分类器往往达不到预期的效果，那是因为在这些问题当中的样本集常常是线性不可分的。因此应该根据这类问题针对线性分类器进行有效的改进，使之能够适应样本集线性不可分的问题。

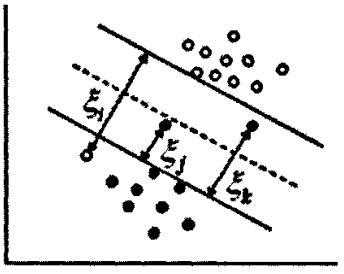


**图 2.5 用于线性不可分样本的分类器**

如图 2.5，不同颜色的两类样本点是线性不可分的。在该问题中，再想寻找上一小节节中的最大间隔超平面是不可能的，因为不存在能够正确划分所有样本点的最优超平面。于是我们自然而然地想到放松对准确率的要求，即允许有不满足约束条件(2.8)的样本点存在。通过引入松弛变量 $\xi_i\geq 0$，$i=1,2,...,n$，可以得到“放宽的”约束条件：

$$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geq 1-\xi_i \tag{2.11}$$

由该约束条件可以看出，只要 $\xi_i$ 取得足够大，一定能够找到满足约束的 $\mathbf{w}$ 和 $b$，即能够找到符合要求的分类超平面。但是显然不能将 $\xi_i$ 取得过大，那样的话分类的错误率将急剧增加。因此为了避免取得过大的 $\xi_i$，应该将其加入最优化问题 1 中成为尽量最小化的一部分，也就是通常所说的进行惩罚。于是问题 1 应该改成如下形式：

问题 2：已知 $y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geq 1-\xi_i$，$\xi_i\geq 0$，$i=1,2,...,n$，求 $\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{n}\xi_i$ 的最小值。

其中第一项 $\frac{1}{2}\|\mathbf{w}\|^2$ 为正则化参数，为的是避免过学习，而第二项 $C\sum_{i=1}^{n}\xi_i$ 体现的是错误率的限制，为的是使经验风险尽可能的小。其中 $C$ 被称为惩罚参数，用来在置信范围与经验风险之间取得合理的折衷，使得结构风险达到最小，从而满足 SRM 准则。

但是这种方法并不适用于所有的线性不可分情况。当不满足约束条件(2.8)的样本点超过一定数量的时候，虽然能够通过选取足够多并且足够大的松弛变量 $\xi_i$，使所有的样本点都能够满足“放宽的”约束条件(2.11)，但是那样做就失去了分类的意义，因为此时的错误率已经大大超出了允许的范围。因此需要通过其他的途径来解决该问题，但最大间隔超平面的思想仍然值得借鉴。由于能够正确划分样本集的最优超平面已经不存在，于是考虑是否有满足要求的最优“超曲面”能够代替它。可以通过如下途径寻找（如图 2.6）：

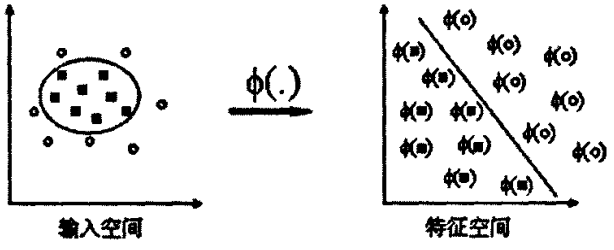



图 2.6 低维空间到高维空间的映射

引入从输入空间 $R^n$ 到更高维 Hilbert 空间 $H$ 的非线性映射 $\phi$，通过这样的映射变换，原来的样本集变成了 $(\phi(\mathbf{x}_1),y_1),(\phi(\mathbf{x}_2),y_2),...,(\phi(\mathbf{x}_n),y_n)$ 的形式，而分类超平面也变成了 $\mathbf{w}\cdot\phi(\mathbf{x})+b=0$。这个分类面虽然在输入空间 $R^n$ 中是非线性的，但是在高维 Hilbert 空间 $H$ 中却可以被看作是线性的，因此可以认为它是一种广义线性超平面。

与之前类似，如果在 Hilbert 空间 $H$ 中仍然有少数线性不可分的样本的话，仍然可以引入松弛变量 $\xi_i\geq 0$，那么所求的最优化问题将变成如下形式：

问题 3：已知 $y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i)+b)\geq 1-\xi_i$，$\xi_i\geq 0$，$i=1,2,...,n$，求 $\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{n}\xi_i$ 的最小值。

其中 $\frac{1}{2}\|\mathbf{w}\|^2$，$C\sum_{i=1}^{n}\xi_i$ 与参数 $C$ 的意义和问题 2 中完全相同。

该最优化问题的解对应着如下拉格朗日函数的鞍点：

$$L(\mathbf{w},b,\alpha)=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{n}\alpha_i[y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i)+b)-1] \tag{2.12}$$

$\alpha_i$表示拉格朗日乘子。通过求解该函数的鞍点，可以将问题3化为一个更简单的对偶问题：

问题4：已知$\sum_{i=1}^{n}\alpha_i y_i=0$，$0\le\alpha_i\le C$，$i=1,2,...,n$，求$\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)$的最大值。

在许多问题当中，真正的求出从$R^n$到$H$的非线性映射$\phi$并不是一件容易的事情，因此通过$\phi$来计算内积$\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)$在很多时候都是难以实现的。而往往在求解问题4的时候，所需要知道的仅仅是内积$\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)$，而不是$\phi(\mathbf{x}_i)$和$\phi(\mathbf{x}_j)$。于是探索一种不通过映射$\phi$而直接求取内积$\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)$的方法是很有必要的。Boser等人于1992年在计算学习理论的会议上首次提出了这样的观点[3]，即只要知道了Hilbert空间$H$中的内积定义，则可以通过SVM的算法，计算出该空间中的最优超平面。核函数的方法正是基于这样的观点，通过将非线性映射和内积计算合为一步，实现所期待的隐式映射与计算。

**定义2.9**（核函数）核函数$K$是一个关于$\mathbf{x}_i$和$\mathbf{x}_j$的函数，它对所有的$\mathbf{x}_i,\mathbf{x}_j\in R^n$，均满足：$K(\mathbf{x}_i,\mathbf{x}_j)=\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)$。其中$\phi$是从输入空间$R^n$到高维Hilbert空间$H$的非线性映射。

在定义了核之后，便可以不再显式的计算该特征映射，而通过直接计算核来达到目的。那么该方法的关键就在于能够找到一个高效而合理的核函数。接下来的内容将对核函数的一些必须具备的条件及基本性质进行研究。

首先由$K(\mathbf{x}_i,\mathbf{x}_j)=\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)$可以看出，核函数$K$是Hilbert空间$H$中两个向量的内积，因此它必须是对称的，同时还需要满足下面的Cauchy-Schwarz不等式：

$$K(x,y)^2\le K(x,x)\cdot K(y,y) \tag{2.13}$$

但是以上这两点并不能成为充分条件，因为一个对称且满足不等式(2.13)的二元函数并不一定能表示成该Hilbert空间中的内积，它还需要满足另外一个条件，那就是著名的Mercer核定理[27,46]：

**定理2.2**（Mercer核定理）在$L_2$赋范线性空间中，核函数$K(x,y)$能展开成如下形式：

$$K(x,y)=\sum_{i=1}^{\infty}\lambda_i\varphi_i(x)\varphi_i(y)，\ \lambda_i>0 \tag{2.14}$$

的充要条件是：对于使得$\int f^2(x)dx<\infty$且不为0的任意函数$f(x)$，条件

$$\iint K(x,y)f(x)f(y)dxdy>0 \tag{2.15}$$

成立。该条件通常被称为Mercer条件。

由定理 2.2 可以看出，满足 Mercer 条件的 $K(x,y)$ 必定能够表示成某个特征空间中的内积，因此可以看作是一个合理的核函数。

在成功的定义了核函数之后，原来的最优化问题 4 可以转化为如下问题：

**问题 5**：已知 $\sum_{i=1}^{n}\alpha_i y_i=0$，$0\le\alpha_i\le C$，$i=1,2,...,n$，求 $\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)$ 的最大值。

由于该问题是一个二次最优化问题，因而一定存在全局的，即唯一的最优解。由 Karush-Kuhn-Tucker（KKT）条件可知，问题的最优解必定满足：对于任意的 $i=1,2,...,n$，

$$\alpha_i[y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i)+b)-1]=0 \tag{2.16}$$

成立。结合 KKT 条件(2.16)，可以得到最优分类面的参数 $\mathbf{w}$ 与 $b$ 分别为：

$$\begin{aligned}\mathbf{w}&=\sum_{i=1}^{n}\alpha_i y_i\phi(\mathbf{x}_i)=\sum_{i=SV}\alpha_i y_i\phi(\mathbf{x}_i)\\ b&=y_j-\sum_{i=1}^{n}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x}_j)=y_j-\sum_{i=SV}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x}_j)\end{aligned} \tag{2.17}$$

其中下标 $j$ 的选取必须满足条件 $\alpha_j>0$。同时可以得到最优分类面的决策函数：

$$f(\mathbf{x})=\mathbf{w}\cdot\phi(\mathbf{x})+b=\sum_{i=1}^{n}\alpha_i y_i\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x})+b=\sum_{i=SV}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x})+b \tag{2.18}$$

由式(2.16)-(2.18)可以看出，只有位于超平面上或者位于超平面内的极少数样本点才会对应着非零的 $\alpha_i$，而绝大多数样本点对应的 $\alpha_i$ 值是 0。由于最优分类面的决策函数只与那些使得 $\alpha_i$ 值非零的少数样本点有关，因此这些样本点被称为支持向量（SV），同时这种算法也被称为支持向量机（SVM）算法。

对于上述非线性支持向量机，有两个互相矛盾的目标：即间隔最大化与错误率最小化。而其中的系数 C 起到的正是一个平衡这两个目标的作用。从定性的角度来看，当 C 值较大的时候，算法对错误率的要求较为严格；而当 C 值较小的时候，算法更为注重的是间隔，也就是泛化能力。然而从定量的角度来看的话，C 值本身并没有明确的含义，因而在 C 值的选取上也没有合适的标准，造成了选取的时候一定的主观性和随意性。Scholkopf 等人在文献[33]中提出了一种新的 SVM 的算法，即 $v$-支持向量机，简称 $v$-SVM。该方法引入了一个新的变量 $v$，它其实是系数 C 的一种变形，目的是为了能够通过调整该参数来控制支持向量（SV）的个数与错误率，从而指定一种合适的系数选取标准。具体来说，参数 $v$ 满足如下两条性质：

1．假定错分的样本个数为 $p$，则有 $v\ge\frac{p}{n}$，即 $v$ 是分类错误率的上界；

2．假定支持向量的个数为$q$，则有$v \le \frac{q}{n}$，即$v$是支持向量占总样本数比例的上界。因此可以根据对错误率与支持向量的要求，合理的选择参数$v$。

设样本集与前文提到过的完全相同，那么基于$v$-SVM 的原始最优化问题为：

问题 6：已知$y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i)+b) \ge \rho-\xi_i$，$\xi_i \ge 0$，$i=1,2,...,n$，$\rho \ge 0$，求$\frac{1}{2}\|\mathbf{w}\|^2 - v\rho + \frac{1}{n}\sum_{i=1}^{n}\xi_i$的最小值。

通过对偶理论，问题 6 可变为如下形式：

问题 7：已知$\sum_{i=1}^{n}\alpha_i y_i = 0$，$0 \le \alpha_i \le \frac{1}{n}$，$i=1,2,...,n$，$\sum_{i=1}^{n}\alpha_i \ge v$，求$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)$的最大值。

同理，根据 KKT 条件，可以解出参数$b$以及最优分类面的决策函数。

## 2.4 本章小结

本章对统计学习理论，两类 SVM 以及相关的知识进行了系统的研究。

第一节首先给出了一般学习问题的数学描述，然后引出了经验风险极小化原则（ERM 原则），这也是最开始处理分类问题时用到的原则。

随着学习理论的不断深入，人们意识到仅仅依靠 ERM 原则来处理问题是不够的，于是开始探索更加合理的方法。统计学家 Vapnik 在 20 世纪 60 年代开始了泛化能力估计的研究，并提出了 VC 维的概念[47,48]。随即出现了许多估计泛化能力的方法，进而产生了一种新的风险估计原则——结构风险极小化原则（SRM 原则），它的目标是寻找一个使得经验风险与置信范围之和最小的估计函数$f$。这种准则也是当今学习能力估计最主要的手段。

第三节通过一个最简单的两类模式识别问题，引出了最大间隔超平面的思想，同时也引出了基于该思想的线性 SVM 分类器。然而在样本集线性不可分的情形下，原来的线性 SVM 分类器显然不再适用，因此需要针对不可分的情形对原有分类器进行改进。于是便出现了线性不可分的 SVM，其中的关键技术就是引进的松弛变量$\xi$。通过松弛变量$\xi$，SVM 分类器在准确率上稍有降低，而泛化能力却得到了很大的提高。但是当样本集中线性不可分的样本点过多的时候，仅仅采用松弛变量$\xi$已经不能保证哪怕是最基本的准确率，显然需要另寻他法。通常所采用的方法是将原来$R^n$空间中的所有样本点通过某个非线性函数$\phi$映射到更高维的 Hilbert 空间$H$中去，使样本点在空间$H$中线性可分，然后再在该空间中进行线性分类器的设计。然而求取这样一个非线性映射$\phi$是一件很困难的事情，它会产生所谓的“维数灾难”，使得实际计算很困难甚至完全不可能实现。核函数在这个时候体现

了它的优越性。它通过直接计算样本在特征空间 $H$ 中的内积来确定最优分类面的决策函数，避免了直接计算映射 $\phi$ 带来的麻烦。关于核方法全面而深入的研究可以参见文献[45]。本节的最后还提到了常规 SVM 的一种变形：$v$-SVM，它通过参数 $v$ 成功的对支持向量（SV）的个数以及分类错误率进行了控制。

另外，本章对相关的最优化方法也进行了必要的概述，其中涉及到的主要内容有约束最优化，拉格朗日乘子，对偶理论与 KKT 条件。首先把原始的分类问题转化为一系列的约束最优化问题，然后利用拉格朗日乘子与对偶理论将问题进一步的简化，最后利用 KKT 条件对问题进行求解。关于拉格朗日乘子与对偶理论更深入的阐述，可以参考文献[45]与[49]；关于 KKT 条件的进一步讨论，可以参考文献[49]与[50]。

# 第三章 多类 SVM 及其改进算法研究

支持向量机的出现最初是为了解决两类模式识别问题。当它在两类问题中展现出其卓越的性能之后，人们自然而然地想到了利用它来解决模式识别与机器学习等领域中的其他难题。对于我们所处的客观世界来说，许多问题所需要面对的事物类别远远不止两类，例如语音识别，手写体数字识别问题等。因此如何将 SVM 方法有效的应用于多类模式识别问题迅速成为了 SVM 研究中的热点。针对多类模式识别问题的经典 SVM 算法主要有一对一方法（1-vs-1），一对多方法（1-vs-all），决策树方法（DAGSVM）等几种。本章将对这些主要的多类 SVM 算法以及相关方面的研究成果进行全面的讨论，并提出了两种分别基于一对多和一对一 SVM 的改进算法。

## 3.1 经典多类 SVM 方法及研究现状

### 3.1.1 多类模式识别问题

模式识别领域中的许多问题，例如语音识别，汉字识别，文本分类等，都属于多类模式识别的问题。与两类模式识别类似，多类识别问题的目的同样也是试图通过经验数据（样本），对未知的事物或者信息进行分类。所不同的只是类别的个数由两个变为了多个。那么同样，多类模式识别问题也有其相应的数学模型：

给定样本集$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),...,(\mathbf{x}_n,y_n)$，其中$\mathbf{x}_i \in R^n$，$y_i \in \{1,2,...,k\}$，$i=1,2,...,n$。目标是寻找一个决策函数$f$，使得对于所有$i=1,2,...,n$，都满足$f(\mathbf{x}_i)=y_i$。也就是根据不同的$y_i$值，将所有样本点正确的分成$k$类。其中$y_i$值相同的样本为同一类。图 3.1 就是一个简单多类识别的模型，其中$k=3$。

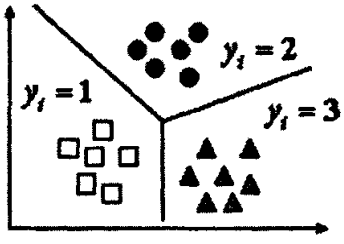



图 3.1 多分类问题模型

下面的小节将对几种经典的多类 SVM 方法：一对一（1-vs-1），一对多（1-vs-all），决策树方法（DAGSVM）等分别进行讨论，并对这些方法以及一些最近的研究成果进行较为全面的分析与总结。

### 3.1.2 一对一支持向量机（1-vs-1 SVM）

顾名思义，一对一支持向量机（1-vs-1 SVM）[51,52]是利用两类 SVM 算法在每两类不同的训练样本之间都构造一个最优决策面。如果面对的是一个 $k$ 类问题，那么这种方法需要构造 $k(k-1)/2$ 个分类平面（$k>2$）。这种方法的本质与两类 SVM 并没有区别，它相当于将多类问题转化为多个两类问题来求解。具体构造分类平面的方法如下：

从样本集中取出所有满足 $y_i=s$ 与 $y_i=t$ 的样本点（其中 $1\le s,t\le k$，$s\ne t$），通过两类 SVM 算法构造最优决策函数：$f_{st}(\mathbf{x})=\mathbf{w}_{st}\cdot\phi(\mathbf{x})+b_{st}=\sum_{i\in SV}\alpha_i^{st}y_iK(\mathbf{x}_i,\mathbf{x})+b_{st}$。用同样的方法对 $k$ 类样本中的每一对构造一个决策函数，又由于 $f_{st}(x)=-f_{ts}(x)$，容易知道一个 $k$ 类问题需要 $k(k-1)/2$ 个分类平面。

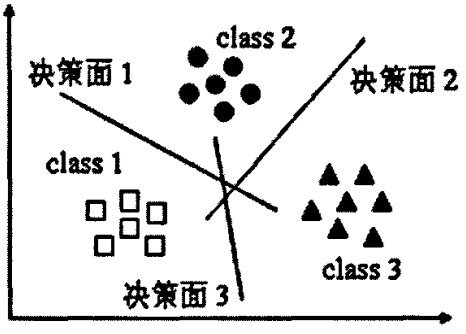



图 3.2 一对一支持向量机（1-vs-1 SVM）示意图

根据经验样本集构造出决策函数以后，接下来的问题是如何对未知样本进行尽量准确的预测。通常的办法是采取投票机制[51,53]：给定一个测试样本 $\mathbf{x}$，为了判定它属于哪一类，该机制必须综合考虑上述所有 $k(k-1)/2$ 个决策函数对 $\mathbf{x}$ 所属类别的判定——有一个决策函数判定 $\mathbf{x}$ 属于第 $s$ 类则意味着第 $s$ 类获得了一票，最后得票数最多的类别就是最终 $\mathbf{x}$ 所属的类别。

1-vs-1 SVM 方法的优点在于每次投入训练的样本相对较少，因此单个决策面的训练速度较快，同时精度也较高。但是由于 $k$ 类问题需要训练 $k(k-1)/2$ 个决策面，当 $k$ 较大的时候决策面的总数将过多，因此会影响后面的预测速度。这是一个有待改进的地方。在投票机制方面，如果获得的最高票数的类别多于一类时，将产生不确定区域；另外在采用该机制的时候，如果某些类别的得票数已经使它们不可能成为最终的获胜者，那么可以考虑不再计算以这些类中任意两类为样本而产生的决策函数，以此来减小计算复杂度。

### 3.1.3 一对多支持向量机（1-vs-all SVM）

同样可以顾名思义，一对多支持向量机（1-vs-all SVM）[52]是在一类样本与剩余的多类样本之间构造决策平面，从而达到多类识别的目的。这种方法只需要在每一类样本和对应的剩余样本之间产生一个最优决策面，而不用在两两之间都进行分类。因此如果仍然是一个$k$类问题的话，那么该方法仅需要构造$k$个分类平面（$k>2$）。该方法其实也可以认为是两类 SVM 方法的推广。实际上它是将剩余的多类看成一个整体，然后进行$k$次两类识别。具体方法如下：

假定将第$j$类样本看作正类（$j=1,2,...,k$），而将其他$k-1$类样本看作负类，通过两类SVM 方法求出一个决策函数：$f_j(\mathbf{x})=\mathbf{w}_j\cdot\phi(\mathbf{x})+b_j=\sum_{i\in SV}\alpha_i^j y_i K(\mathbf{x}_i,\mathbf{x})+b_j$。这样的决策函数$f_j(\mathbf{x})$一共有$k$个。给定一个测试输入$\mathbf{x}$，将其分别带入$k$个决策函数并求出函数值，若在$k$个$f_j(\mathbf{x})$中$f_s(\mathbf{x})$最大，则判定样本$\mathbf{x}$属于第$s$类。

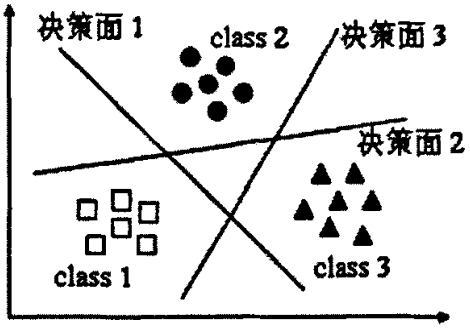



图 3.3 一对多支持向量机（1-vs-all SVM）示意图

1-vs-all SVM 方法和 1-vs-1 SVM 相比，构造的决策平面数大大减少，因此在类别数目$k$较大时，其预测速度将比 1-vs-1 SVM 方法快许多。但是由于它每次构造决策平面的时候都需要用上全部的样本集，因此它在训练上花的时间并不比 1-vs-1 SVM 少。同时由于训练的时候总是将剩余的多类看作一类，因此正类和负类在训练样本的数目上极不平衡，这很可能影响到预测时的精度。另外，与 1-vs-1 方法类似，当同时有几个$j$能取到相同的最大值$f_j(\mathbf{x})$时，将产生不确定区域。

### 3.1.4 决策树算法（DAGSVM）

决策树算法（DAGSVM）[54]与前面两种方法均不太一样，1-vs-all SVM 和 1-vs-1SVM 通过一系列的决策函数来依次确定样本的类别，它们可以被认为是“肯定型”的算法，而 DAGSVM 却是通过在每层节点处对不符合要求的类别进行排除，最后得到样本所属的类

别，应该算是一种“否定型”的算法。具体算法如下：

在训练阶段 DAGSVM 和 1-vs-1 SVM 方法的步骤一样，它们的差别主要体现在预测阶段。DAGSVM 首先从 $k(k-1)/2$ 个分类决策面中任意选取一个，不妨设为 $f_{st}$，然后将未知样本 $\mathbf{x}$ 代入该决策函数进行判定：若在此决策函数中 $\mathbf{x}$ 被判定为第 $s$ 类，那么将所有与第 $t$ 类样本相关的决策函数全部删除，然后从剩下的与第 $s$ 类样本相关的分类决策面中任取一个重复以上步骤；若是被判定为第 $t$ 类，方法也是完全类似。依此类推，直到决出样本 $\mathbf{x}$ 的最终类别。图 3.4 是使用决策树算法解决一个四类问题的完全示意图。

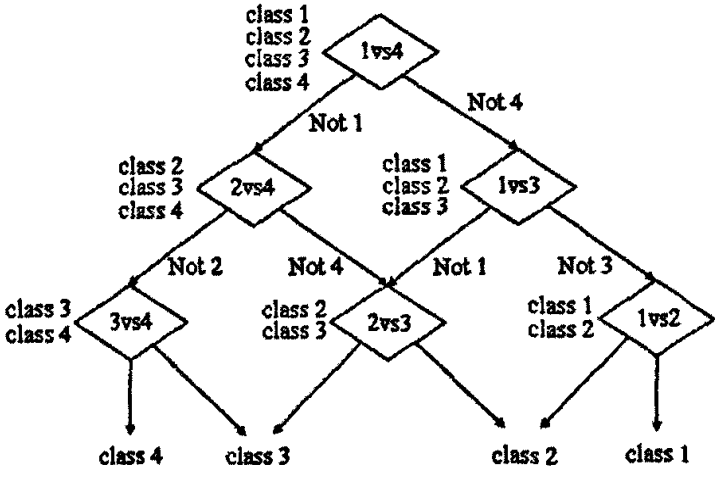



图 3.4 决策树算法（DAGSVM）示意图

该方法和 1-vs-1 SVM 一样，训练的时候首先需要构造 $k(k-1)/2$ 个分类决策面。然而和 1-vs-1 SVM 方法不同的是，由于在每个节点预测的时候同时排除了许多类别的可能性，因此预测的时候用到的总分类平面只有 $k-1$ 个，比 1-vs-1 SVM 要少很多，预测速度自然提高不少。但 DAGSVM 算法也有其不足之处。正由于它采取的是排除策略，那么最开始的判定显得尤为重要，如果在开始阶段就决策错误的话，那么后面的步骤都没有意义了。因此如何选取判定的顺序和得开始阶段的判定令人信服是值得进一步研究的问题。

### 3.1.5 分析与总结

在经典的多类 SVM 算法当中，1-vs-1，1-vs-all 和 DAGSVM 是三种最为常见的方法。1-vs-1 与 DAGSVM 都需要在训练的时候构造 $k(k-1)/2$ 个分类平面，而 1-vs-all 却只需要 $k$ 个。然而 1-vs-all 方法训练每个决策面的时候都需要用到所有的训练样本，而 1-vs-1 与 DAGSVM 方法只需要用到其中两类样本，因此在训练的时间上 1-vs-all 方法未必比 1-vs-1 和 DAGSVM 方法要少。在预测的时候，1-vs-all 和 DAGSVM 方法所需要用到决策面分别为 $k$ 个和 $k-1$ 个，而 1-vs-1 却要用到 $k(k-1)/2$ 个，因此在预测时间上 1-vs-all 和 DAGSVM 都要优于 1-vs-1。而在精度方面，C. W. Hsu 和 C.J. Lin 在文献[55]中提出 1-vs-1 和 DAGSVM

方法的精度要比 1-vs-all 高，这个观点似乎也得到了大多数人的认同。但 R. Rifkin 等人却在文献[56]中表达了相反的观点，他提出并验证了 1-vs-all 算法的性能不会像人们通常认为的比基于 1-vs-1 和 DAGSVM 策略的算法差，相反，1-vs-all 策略的性能完全可以达到甚至优于其他多类算法的性能。

其他较为常见的多类 SVM 方法还有被称为整体策略（all-together method）和纠错输出编码（ECOC: error-correcting output codes）的算法[57,58]。整体策略并不需要构造多个决策面，它将所有的样本看作一个整体，然后将多类识别问题转化为一个大的整体最优化问题，只需要通过对该最优化问题求解就能判断出测试样本所属类别。但是正由于它要同时处理所有样本，因此其计算量相当大，在实际操作中几乎不可行。该方法的详细论述见文献[57]。ECOC 算法每次根据需要将 $k$ 类样本分成适当的两类，且这两类都可以包含不同数目的多类样本，然后构造分类平面。通过计算可以知道，$k$ 类样本用该方法进行分类时所需决策平面个数最少可以达到$[\log_2 k]$或$[\log_2 k]+1$个，是目前所有方法中最少的。可是这样的分类机制虽然更加快速灵活，但如何将 $k$ 类样本适当的分成两类是一个很困难的问题，因此在实际应用中该方法并不常见。对 ECOC 算法的深入讨论可以参考文献[58]。

近两年针对多类 SVM 方法的研究还有一些有意思的成果。例如将 SVM 与无监督方法结合[59]，与最近邻方法结合[60]，或者是引入多类模糊 SVM 的概念[61,62]。除此之外，还有一些其他文献[63,64]也提出了新的多类 SVM 算法。由此可见，多类 SVM 算法正是目前 SVM 研究中的热点。接下来的两节分别提出了基于 1-vs-all 与 1-vs-1 方法的两种改进。通过实验数据可以表明，改进的 1-vs-all 算法同时提高了原有算法的精度与速度。而改进的 1-vs-1 算法在保证了精度的同时，大大的提高了原有算法的速度。

## 3.2 基于 1-vs-all 策略的改进算法

### 3.2.1 改进 1-vs-all 算法

传统的 1-vs-all 方法将每一类都与其他类区分开来，虽然只构造了 $k$ 个分类平面，但是由于它每次构造决策平面的时候都需要用上全部的样本集，而且总是将剩余的多类看作一类，因此很有可能对训练速度和分类精度造成影响。本节针对这两个方面的不足对传统 1-vs-all 方法进行改进，希望能够获得比原来更加优越的性能。

首先从训练速度上考虑，既然要减少训练时间，那么必须减少分类平面个数或者每次参与决策平面计算的样本数。然后从分类精度方面考虑，也应该减少每次参与训练的样本数。又已知现有的分类平面数为 $k$，这已经是一个比较小的数，因此不太可能再大幅度的减少分类平面的个数，那么剩下的途径只有减少每次参与训练的样本数。基于以上的分析，改进的算法描述如下：

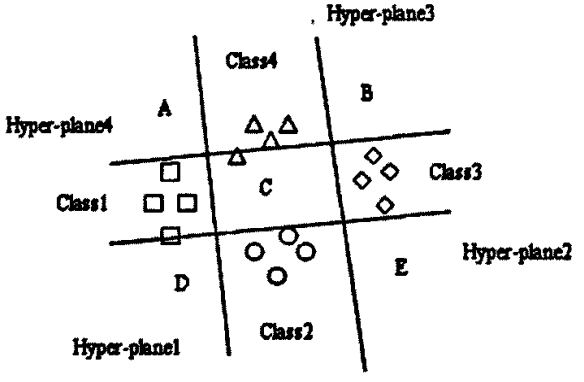
1. 从所有 $k$ 类样本中任意选取一类，记为第 1 类，然后将剩余的 $k-1$ 类作为整体看

作一类，用两类 SVM 算法构造出一个分类平面，并对应的记为决策平面 1；

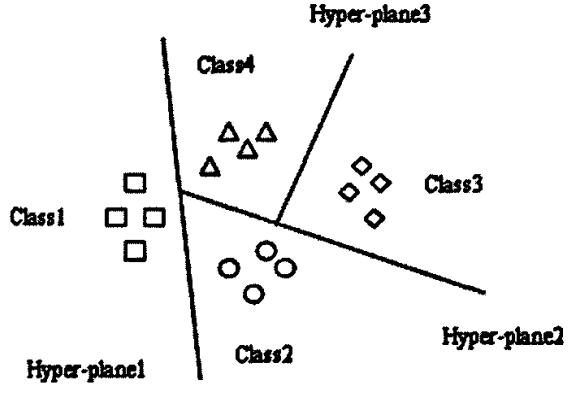
2. 忽略第 1 类样本，从剩下的$k-1$类样本中再任意选取一类，记为第 2 类，然后将剩余的$k-2$类看作一类，再用两类 SVM 算法构造出一个分类平面，并对应的记为决策平面 2；
3. 忽略第 1，2 类所有的样本，然后重复以上的步骤对剩余的$k-2$类进行分类，直到所有的类别都被区分开来为止。

到最后可以发现，实际构造的分类超平面只有$k-1$个，比预计的$k$个还减少了一个。这是因为最后第$k$类样本只需要用之前构造的$k-1$个分类平面就能和其他类样本完全分开，而不需要再另外构造一个超平面。

从以上叙述可以看出，该方法利用$k-1$个分类平面成功的完成了全部$k$类样本的分类，并且步骤越往后，某类样本用来和其他类样本分开的决策平面就更多，即用到的信息更多，因而精确度也将越高。




(a)传统 1-vs-all 算法




(b)改进 1-vs-all 算法

图 3.5 四类问题改进算法示意图

图 3.5 通过一个四类问题举例对该改进算法进行说明。如图所示，图(a)展示的是传统

1-vs-all 的分类方法，而图(b)则是改进的分类方法。首先任意确定一类样本并标记为 class1，然后构造超平面 1 将其和其他类分开；然后类似的确定 class2，并用超平面 2 将它与剩下的两类分开；接着同样的选取 class3，并用超平面 3 将它与 class4 分开。从图 3.5(b)中可以看出，class1 与其他类分开只用到了超平面 1，class2 与其他类分开用到了超平面 1 和超平面 2，而 class3 与其他类分开则用到了超平面 1，超平面 2 和超平面 3，class4 的情况与 class3 一样。明显的，通过信息的有效利用，不但使分类平面比传统方法少了一个，同时也减少了后面步骤中参与分类的样本数。

### 3.2.2 仿真结果

本小节将用一个仿真实验对改进 1-vs-all 方法的性能进行验证。用来进行仿真实验的测试图片是一幅飞机的照片，如图 3.6。

图 3.6 测试图片

仿真测试的目标是将飞机从图片的背景中识别出来。该实验将测试图片中的象素点根据其 RGB 值分成 4 类，每一类在图中取一定数量的样本点，然后根据这些样本点再对测试图片中的所有象素点进行分类，产生识别后的图片。传统的 1-vs-all 方法需要 4 个决策平面来对四类样本点进行划分，而改进的方法只需要三个。

(a)传统 1-vs-all 算法

(b 改进 1-vs-all 算法

图 3.7 测试结果

图 3.7 为仿真测试的结果。从图中可以明显看出，用传统算法进行识别的结果并不是特别理想，在飞机周围有许多比较模糊的地方没有和飞机清楚的分开；而用改进算法识别的结果则非常理想，飞机与背景能够很清楚的区分开来，甚至连飞机上靠近尾部的一个微小的五角星标记也显得十分清晰，这充分说明了改进算法在精度上的优越性。

本节提出的基于 1-vs-all 的改进算法主要优势体现在两个方面：第一，它比从前的方法少一个分类平面，而且每次训练的样本比传统方法要少，因此运算速度得到了提高；第二，由于每次的训练样本的适当减少以及先验信息的合理利用，使得分类的精度有所增加。此外，除了分类平面本身，该方法没有其它不确定区域。

下一节将会提出另一种基于 1-vs-1 策略的改进算法。

## 3.3 基于 1-vs-1 策略的改进算法

### 3.3.1 改进 1-vs-1 算法

传统的 1-vs-1 方法通过在每两类不同的样本之间构造决策平面来完成分类的任务，它一共需要构造 $k(k-1)/2$ 个超平面。因为分类平面的个数较多而且每次投入训练的样本集都较小，因此保证了该方法的高精确性；但也正因为分类平面多的原因，导致了该方法在整个分类过程特别是预测阶段需要耗费较多的时间。而对于很多实际应用问题来说，训练以及预测速度是衡量该方法成功与否的一个重要标志。象许多需要在线学习或即时预测的问题，例如入侵检测[65]，语音识别[12,66]，障碍识别[67]等，快速及时的算法是成功解决问题的保证。甚至在保证了一定精度的前提下，提高预测的速度比进一步的提高精度显得更为重要。基于这样的考虑，本节提出了一种改进的 1-vs-1 算法，使得在保证了合理精度的同时，大大提高了训练以及预测阶段的速度。它只需要构造 $[3(k-1)/2]$ 个决策平面（$[x]$ 函数表示对 $x$ 进行高斯取整运算），同时也保留着传统 1-vs-1 方法的优点——预测精度。具体步骤如下：

1. 将 $k$ 类样本进行编号，从第 1 类到第 $k$ 类，选取的顺序任意；
2. 把第 3 类到第 $k$ 类样本看作一个整体，也就是同一类，那么该 $k$ 类问题就转化为了一个 3 类问题，接下来用传统的 1-vs-1 方法计算出该三类问题的三个决策函数，这样第 1 类与第 2 类样本被成功的分离了出去。如果 $k=3$，那么算法到此为止；否则进行下一步；
3. 在剩余的 $k-2$ 类中，把第 5 类到第 $k$ 类样本看作一个整体，这又将问题转化为了一个 3 类问题，重复以上的步骤，可以将第 3 类与第 4 类样本成功的分离出去；
4. 依此类推，最后将剩下两类或者三类样本，用两类 SVM 对剩下的这两类或者三类样本进行分类，直到把所有类别成功的分开。

根据以上的步骤，我们发现该方法构造的分类平面总个数大大少于原来的 1-vs-1 方法。

原有的方法在面对一个$k$类问题的时候需要$k(k-1)/2$个决策平面，而在改进的算法中，第1类与第2类被三个决策面成功的从全部$k$类样本中分离出去。同样，第3类与第4类也被三个决策面分离出去。到最后如果剩下两类，那么只用再构造一个超平面，若剩下三类，也只需要再构造三个决策平面便能成功完成分类的任务。因此如果$k$是奇数的话，那么该方法需要构造的总决策平面个数为$3(k-1)/2$；如果$k$是偶数，则总的决策平面个数为$3(k-2)/2+1$。这两个表达式可以被写成一种统一的形式：$[3(k-1)/2]$。由此可以看出，改进的方法在决策平面的个数上与原方法的差为$(k-3)(k-1)/2$（$k$为奇数）或$(k-2)^2/2$（$k$为偶数）。当$k$值较大时，该差距将相当明显。

下表给出的是$k$值为3到10的时候两种方法构造的决策平面个数的比较。当$k$=3，4，5的时候，个数的差距并不明显；但是当$k$从6增加到10的时候，明显可以看出个数的差距在迅速增加。

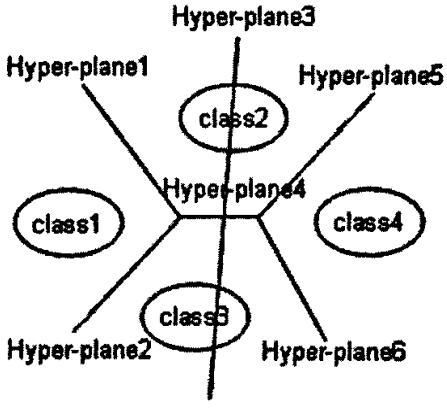
| 类别数目 | 决策平面数目 | |
|---|---|---|
| | 传统 1-vs-1 | 改进 1-vs-1 |
| 3 | 3 | 3 |
| 4 | 6 | 4 |
| 5 | 10 | 6 |
| 6 | 15 | 7 |
| 7 | 21 | 9 |
| 8 | 28 | 10 |
| 9 | 36 | 12 |
| 10 | 45 | 13 |

表 3.1 传统 1-vs-1 与改进方法构造的分类平面数比较

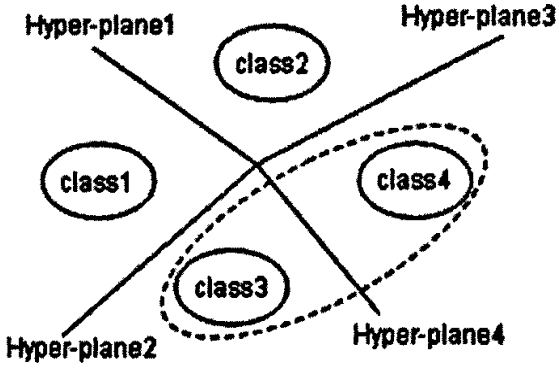
与前一节类似，接下来将用一个四类问题举例对该改进算法进行说明。首先对这四类进行编号，依次从 class1 到 class4。图 3.8(a)展示的是用传统的 1-vs-1 方法对该问题进行分类的结果。它需要的决策平面数为：4×3/2=6 个。从图中可以看出，构造用来分开 class1 和 class4 的超平面 3 是完全没有必要的，因为它们可以通过超平面 1 和超平面 2 正确的区分开。

图 3.8(b)展示了用改进的算法解决该问题的过程：首先将 class3 和 class4 看作一个整体，记作 class3'，然后根据 1-vs-1 方法对 class1，class2 和 class3'进行分类。该步骤完成之后，接下来的工作非常简单，只需要再对 class3 和 class4 构造一个分类平面即可。从图中明显可以看出，只需要四个分类平面就可以有效的将四类不同的样本识别出来。利用超平面 1 和超平面 2 可以将 class1 识别出来；利用超平面 1 和超平面 3 可以将 class2 识别出来；而 class3 和 class4 都可以利用超平面 2，超平面 3 和超平面 4 进行区分。与传统的方法相比，该算法减少了不必要的决策面，提高了训练速度，使用较少的分类平面达到了和

较多分类平面相似的分类效果，同时没有产生不确定性分类区域（分类平面本身除外）。




(a)传统 1-vs-1 算法




(b)改进 1-vs-1 算法

图 3.8 四类问题改进算法示意图

### 3.3.2 实验结果

本小节将用一个仿真实验和一个实际数据的测验来验证改进算法的性能，并将其与传统的 1-vs-1 和 1-vs-all 方法进行对比。实验采用了[68]提供的 SVM 工具包来实现传统的 1-vs-1 和 1-vs-all 算法，然后在此基础上用 matlab 语言实现了改进的 1-vs-1 算法，并在两个实验中分别展示了改进算法的性能。

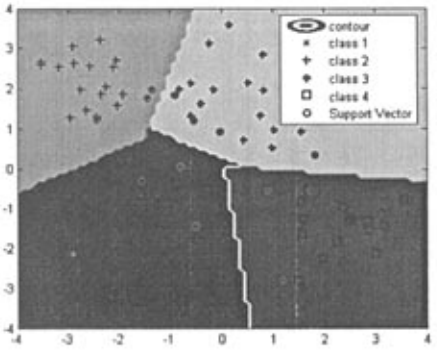
设计仿真实验的目的是为了从直观上比较以上所说的三种 SVM 算法的分类效果。和[68]类似，该实验给出了 8 组数据，分别是从 rank1 到 rank8，分成四类，每一组数据代表其中一类的 $x$ 值或者 $y$ 值。其中每一组数据都包括了 20 个浮点数，这些浮点数都是从一组服从标准正态分布的样本点之中随机抽取出来的。然后根据如下算式将四类样本点分开：

$$\begin{aligned}(x_1,y_1) &= (\sum \text{rank1}-1.4,\ \sum \text{rank2}-1.4)\\ (x_2,y_2) &= (\sum \text{rank3}-2.4,\ \sum \text{rank4}+2)\\ (x_3,y_3) &= (\sum \text{rank5}+0,\ \sum \text{rank6}+2)\\ (x_4,y_4) &= (\sum \text{rank7}+2,\ \sum \text{rank8}-1.4)\end{aligned} \tag{3.1}$$

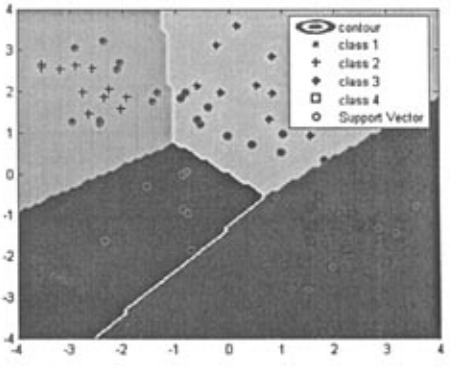
根据式(3.1)，四类样本点将坐落在 $R^2$ 平面内的四个不同区域，测试数据集的范围描述如下：

$$X=[-4:0.1:4],\ Y=[-4:0.1:4],\ (x,y)=X\times Y \tag{3.2}$$

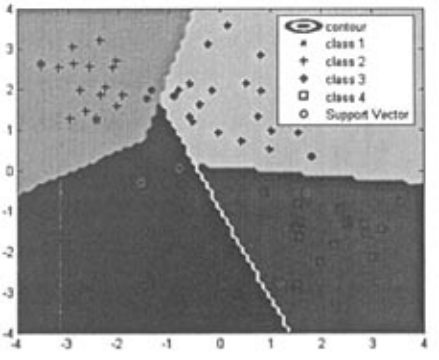
这意味着 X 和 Y 都将是 81 维的向量[-4, -3.9, -3.8, …, -0.1, 0, 0.1, 0.2, …, 3.9, 4]，而测试数据则是向量 X 与 Y 的笛卡尔积。




(a) 传统 1-vs-1 算法




(b) 传统 1-vs-all 算法




(c) 改进 1-vs-1 算法

图 3.9 仿真实验结果

仿真实验的结果如图 3.9。图中四种不同形状的点代表着四个不同的类别，而不同颜色的区域代表着分类的结果，被圆圈包围的样本点代表的是支持向量(Support Vector)。支持向量的个数越少，最优化问题的求解则越简单。三种方法：1-vs-1，1-vs-all，和改进 1-vs-1 的支持向量个数分别为 13，32 和 9 个。其中 1-vs-all 方法需要的支持向量个数远远多于其

他两种方法，原因是 1-vs-all 每次训练都需要用到所有的样本，因此产生的支持向量的个数也随之增多。而改进方法比传统 1-vs-1 支持向量要更少的原因在于它只需要构造 4 个分类平面，比传统方法少了两个。

第二个测验是用实际数据进行的实验，为的是展示改进算法在保证了较高精度的同时大大提高了预测的时间。用来进行比较的仍然是 1-vs-1，1-vs-all，和改进 1-vs-1 三种方法。

| 数据来源 | 特征 | 类别数目 | 训练样本 | 测试样本 |
|---|---|---|---|---|
| [69] | 53 | 7 | 300 | 2000 |

表 3.2 测试使用数据说明

该实验所用到的实际数据可以从[69]中下载。表 3.2 是关于实验的数据说明。该数据集分类 7 类，特征一共有为 53 个，本次实验用来训练和测试的数据分别为 300 和 2000 个。

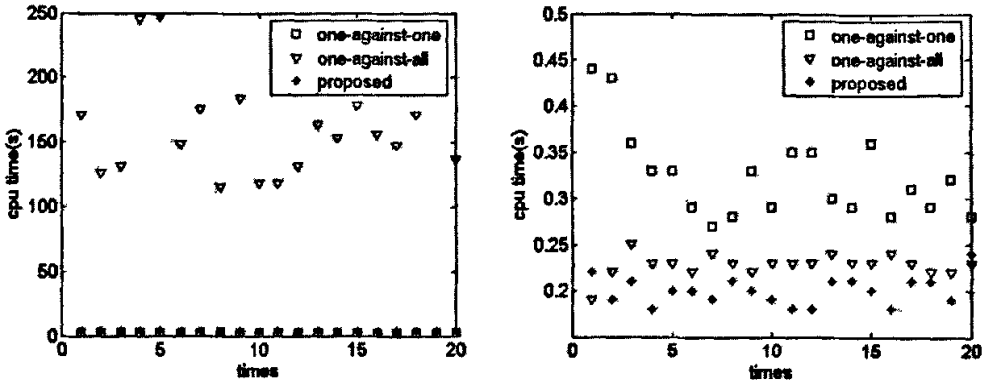
| | d=1 | d=2 | d=3 | d=4 | d=5 |
|---|---|---|---|---|---|
| 1-vs-1 | 735 | 779 | 730 | 727 | 728 |
| 1-vs-all | 873 | 855 | 817 | 876 | 874 |
| 改进 1-vs-1 | 705 | 864 | 817 | 806 | 805 |

表 3.3 分类结果：错分数据的个数

表 3.3 给出了基于多项式核 $K(x_i,x_j)=(x_i\cdot x_j+1)^d$ 的三种方法的实验结果。多项式的次数 $d$ 依次从 1 取到 5。从表中可以看出，当 $d$=1 时，改进 1-vs-1 算法的错分数目是三种方法中最少的；而当 $d$=2 时，改进算法的错分数目是三种方法中最多的；当 $d$=3，4，5 时，改进算法的错误率普遍比 1-vs-all 方法低，但是却比 1-vs-1 方法高。表格中的结果体现出改进的方法在精度上达到了 1-vs-1 和 1-vs-all 两种方法的平均水平，因此可以认为达到了合理的精度。

接下来的图 3.10 从训练时间以及预测时间两方面充分体现了改进方法的优越性。该部分实验对三种方法的训练和预测时间进行了 20 次测试，使用的数据集与参数完全一样。用来进行实验的机器配置为 1.7G Pentium 4 的 CPU 和 360 MB 内存。从图 3.10(a)中可以看出，由于每次训练都用上了所有的训练样本，因此 1-vs-all 方法所需要的训练时间要远远多于其他两种方法，20 次实验所需的平均时间为 161.38 秒。而 1-vs-1 和改进算法的平均训练时间分别为 2.7555 秒和 3.7904 秒，考虑到训练过程本身的耗时性及 1-vs-all 方法训练所需要的时间，这两种方法在训练中所花费的时间几乎可以忽略不计。图 3.10(a)展示的是三种方法在预测阶段所花费的时间。由于总共确定了 $k(k-1)/2$ 个决策平面并且在预测的时候都将投入使用，因此 1-vs-1 方法在预测过程中耗时最多。由图中可以看出，改进

1-vs-1 方法所花费的预测时间是所有方法中最少的，20 次测试的平均时间为 0.2003 秒，其预测速度达到了原有 1-vs-1 方法的两倍。




(a) 训练时间　　(b) 预测时间

图 3.10 分类时间对比结果

## 3.4 本章小结

本章对多类 SVM 方法进行了系统的讨论。第一节回顾了三种经典的多类 SVM 方法，分别是：1-vs-1 方法，1-vs-all 方法和 DAGSVM，并对这三种方法进行了全面的比较。以 $k$ 类问题为前提，1-vs-1 方法与 DAGSVM 在训练的时候都需要构造 $k(k-1)/2$ 个决策平面，而 1-vs-all 方法只需要构造 $k$ 个决策面。但是由于 1-vs-all 方法在每次构造决策面的时候，都需要用到所有的样本，而 1-vs-1 方法与 DAGSVM 只需要两类样本，因此在训练阶段的速度 1-vs-all 方法并不比其他两种方法快。在预测阶段，1-vs-all 方法和 DAGSVM 各需要用到 $k$ 个和 $k-1$ 个决策函数，而 1-vs-1 方法却需要进行 $k(k-1)/2$ 次决策，因此就预测速度来说 1-vs-all 方法和 DAGSVM 要优于 1-vs-1 方法。在分类精度方面，看法并不统一，甚至存在一些完全相反的观点。综合考虑时间以及效率等因素，在实际当中应用得最广的还是 1-vs-1 方法和 1-vs-all 方法。接下来的两节分别针对 1-vs-1 方法和 1-vs-all 方法提出了相应的改进策略。

在 3.2 节中提出的 1-vs-all 改进策略主要是基于一种“排除法”的思想。每进行完一次分类就排除一类样本，直到分出最终结果。该方法的优点在于多次利用先验分类平面，不但使得分类平面的个数减少了一个，而且步骤越往后参与构造分类平面的样本点越少，提高了分类的速度与精度。但是这种改进算法仍然有不少值得注意和进一步研究的地方，例如以何种顺序选取类别就是一个值得研究的地方。在该算法中，类别的选取是任意的，但是从分类机制中可以看出，越靠前的分类越重要，因为剩下的所有类别在进行预测的过程当中都将用到之前所有决策面的信息。因此如何按照类别的重要性和分类的难易程度进行合理的顺序选取与该算法的最终效果密切相关。

接下来的 1-vs-1 改进策略从本质上来说将原问题分解为了一系列的三类问题，然后针对这一系列的三类问题构造分类平面。最终构造的决策函数为$[3(k-1)/2]$个，比原有的 1-vs-1 方法少了$[(k-2)^2/2]$个。实验结果说明了该方法在保持了合理精度的同时大大的提高了预测的速度。但是和 1-vs-all 的改进一样，该方法虽然有自己的优点，但也有许多有待提高的地方。例如如何以合理的顺序选取类别仍然是值得关注的一个方面，还有如何在保持该方法速度优势的同时提高其预测精度也是下一步将要研究的问题。

# 第四章 基于SVM算法的新型P2P流量检测系统设计*

自从90年代初期SVM作为一种新的机器学习方法出现在模式识别领域以来，就一直受到广泛的关注与重视。到目前为止，SVM已经被公认为是精度最高的模式分类器之一，它在各种各样的实际应用领域当中都取得了相当出色的成果。本章的重点是将SVM算法与P2P流量检测模型相结合，设计了两种新的检测系统：第一个系统通过两类SVM以及一个平滑函数，对P2P流量以及非P2P流量进行识别；第二个基于多类SVM的系统不但可以对P2P流量进行检测，而且可以完成对P2P流量的应用级分类。

## 4.1 P2P流量检测问题简介

P2P是一种分布式网络，网络的参与者需要共享他们所拥有的部分硬件资源，而这些资源都能被其它对等节点直接访问而无需经过中间实体。作为网络的参与者既可以获取其他用户提供的资源，也需要给其他用户提供自己的部分资源。

虽然P2P技术的发展为网络资源的传播提供了方便，但是由于目前的技术还不够完善，因此将不可避免地带来许多问题，如侵犯知识产权，传播网络病毒与不健康内容，占用网络资源等。基于以上这些原因，加强对P2P流量的检测与管理势在必行。

针对传统流量的检测方法是基于端口的，而P2P技术的发展趋势则是尽量避免被这些流量检测方法所检测到，以下是它常采用的几种主要方法：

- 动态选定端口或使用可变端口；
- 伪装成其它流量；
- 对数据进行加密。

这些特征使得P2P流量比传统流量要难检测得多。目前针对P2P流量的检测方法主要有以下三类：

1) 基于payload特征的检测方法[70-73]：

该方法通过对数据包应用层协议的检测来识别P2P流量。它的特点是简单、实用，不仅可以区分P2P流量与非P2P流量，还可以对流量进行应用分类。但该方法无法检测对数据进行了加密的P2P流量。同时随着P2P流量的增加，特征串的数量也相应增加，使得该方法每次检测时所需要匹配的特征串越来越多，检测效率也将越来越低。

2) 基于统计特征的检测方法[74-76]：

该方法是利用网络流量的统计特征，例如IP地址来检测P2P流量的方法。与第一种方法相比，该方法易于检测对数据进行了加密的P2P流量以及payload特征未知的P2P流量。并且其检测效率不会随着P2P流量的增加而降低。但该方法也有其不足之处：首先，它不能对P2P流量进行应用级分类；其次，它的实现过于复杂，导致了实际应用的困难。

* 本章内容主要来源于作者与国防科技大学计算机学院硕士研究生王锐合作的三篇论文（附录中论文[2,3,4]），特此声明。

3) 跨层方法[77]：

该方法同时利用了流量中的 payload 特征与统计特征，试图将两种特征有机的结合，扬长避短，从而达到降低实现难度与提高检测精度的目的。

以上三种方法都有属于自己的优缺点。本章所提出的方法是将 SVM 与 P2P 流量检测问题相结合，从大类别上来说应该属于基于统计特征的检测方法。但据作者所知，利用统计学习理论中的最新成果——SVM 来解决 P2P 流量检测问题，目前尚无人在这方面开展系统的研究。接下来的两节将分别研究基于两类 SVM 的 P2P 流量识别问题与基于多类 SVM 的 P2P 流量应用级分类问题。

## 4.2 基于两类 SVM 的 P2P 流量检测

### 4.2.1 问题描述，数据采集与特征提取

本节所需要解决的问题是 P2P 流量识别，也就是只要判断出该流量是 P2P 流量还是非 P2P 流量即可。考虑到 SVM 在进行两类分类时的卓越性能，同时考虑到精确识别 P2P 流量的必要性与重要性，本节将使用两类 SVM 方法来完成流量检测的工作，并希望该系统达到以下四条标准：

1) 能够检测出主要的 P2P 流量；
2) 能够检测出未知的和加密的 P2P 流量；
3) 能够顺利完成流量较大时的识别工作；
4) 能够通过学习不断提高分类性能。

为了更好的展现 SVM 的卓越性能，本节设计了两组不同的实验并分别构建了对应的实验框架。两组实验的数据均来自于从校园网络上收集到的 netflow 数据。如图 4.1 所示，netflow 被安装在连接校园网与因特网的路由器上，因此可以通过它采集到从校园网到因特网的所有流量。

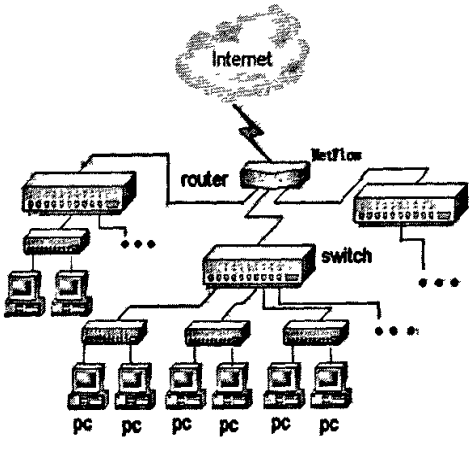



图 4.1 数据收集示意图

对于第一组实验，采集的数据分为 5 组，从 setA 到 setE，如表 4.1 所示。setA 与 setB 分别为纯 P2P 流量和非 P2P 流量，setC 是 P2P 与非 P2P 的混合流量，而 setD 与 setE 分别是花了一天和一周时间收集到的混合数据。每个数据集都是隔 15 分钟汇聚一次，每小时汇聚四次。所有数据集的大小以及混合流量中 P2P 所占比例都已在表中列出。

| | 开始时间 | 总时间 | 间隔时间 | PCs |
|---|---|---|---|---|
| **Set A** | 10:00 04/11 | 1 hour | 15 min | 1 |
| **Set B** | 10:00 04/12 | 1 hour | 15 min | 1 |
| **Set C** | 10:00 04/13 | 1 hour | 15 min | 1 |
| **Set D** | 10:00 04/14 | 1 day | 15 min | 12 |
| **Set E** | 10:00 04/15 | 1 week | 15 min | 30 |
| | **P2P 流量比例** | **P2P 流量类型** | **Flows** | **Bytes** |
| **Set A** | 100% | BitTorrent | 245.3k | 177.2M |
| **Set B** | 0% | -- | 300.1k | 35.8M |
| **Set C** | 59.6% | BitTorrent | 378.8k | 43.3M |
| **Set D** | 82.3% | pplive, eDonkey | 3.004M | 2.53G |
| **Set E** | 10.3% | pplive, eDonkey, BitTorrent | 29.67M | 23.77G |

表 4.1 实验 1 数据收集示意图

对于第二组实验，采集的数据一共分为 6 组，如表 4.2 所示。第一组数据用来作为训练数据集，其中包含 28.2%的 BT 流量。后面 5 组数据集：setA 到 setE 都是测试数据集，其中 setA、setB 和 setC 分别包括了 3 种不同的 P2P 流量以及一些非 P2P 流量。setD 是从 10 台电脑上采集得到的，其中包括了 4 种不同的 P2P 流量以及一些非 P2P 流量。而 setE 包含的全部是非 P2P 流量，采集该数据集的目的是为了对第二个实验中的虚警率进行测试。

| | 间隔时间 | P2P 流量比例 | P2P 流量类型 | Flows |
|---|---|---|---|---|
| **Training Set** | 15 min | 28.2% | BitTorrent | 2535 |
| **Testing Set A** | 15 min | 55.3% | eDonkey | 1348 |
| **Testing Set B** | 15 min | 70.2% | pplive | 2201 |
| **Testing Set C** | 15 min | 61.3% | baizhao | 1423 |
| **Testing Set D** | 15 min | 53.3% | BitTorrent, eDonkey, pplive, baizhao | 4827 |
| **Testing Set E** | 15 min | 0% | -- | 7305 |

表 4.2 实验 2 数据收集示意图

在采集完数据之后，接下来的工作是要对这些原始数据进行特征提取。提取的特征必须满足两点要求：第一，要适合在 SVM 机制下进行分类；第二，要能够体现出 P2P 和非 P2P 流量的区别。经过观察发现，P2P 与非 P2P 流量的主要区别体现在 IP 地址的差异和端

口选取的差异上。因此该实验选取一个三维向量来反映这些差异：

$$vector(flow) = < f(src,dst), g(src,spt), h(src,dpt) > \tag{4.1}$$

其中第一维特征，即函数 $f$ 体现了 P2P 和非 P2P 流量在 IP 地址上的差异，而第二维特征和第三维特征则体现了两种流量在端口选取上的差异。其中函数 $f$ 有如下定义：

$$f(src,dst) = dif(src,DST) / same(src,dst) \tag{4.2}$$

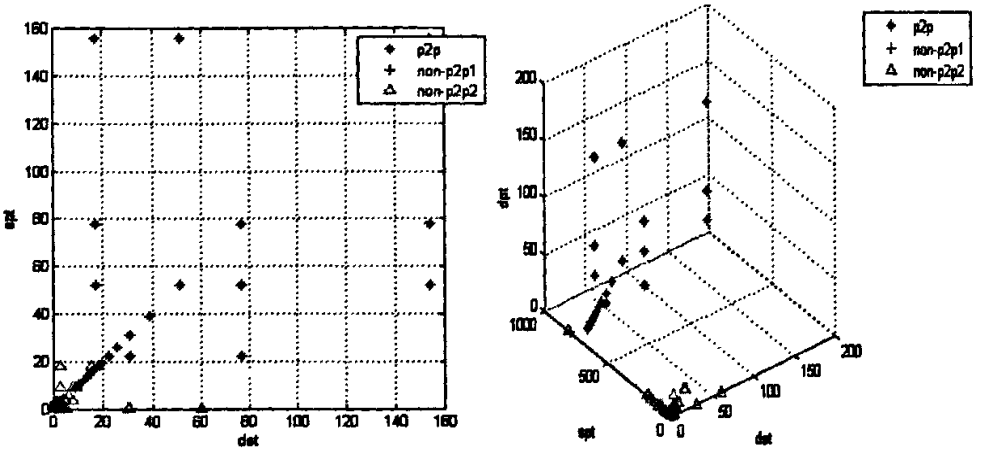
该函数有两个自变量：源 IP src 与目标 IP dst。分子 $dif(src,DST)$ 表示的是与源 IP 有连接的不同目标 IP 的个数，而分母 $same(src,dst)$ 则表示该目标 IP 与源 IP 建立的连接数。下面的例子将对该定义进行说明：

假设某源 IP 与 4 个不同的目标 IP 存在连接，它与前三个 IP 的连接数都是 1，与第四个 IP 的连接数是 2，则 $dif(src,DST)=4$，$same_1(src,dst)=same_2(src,dst)=same_3(src,dst)=1$，$same_4(src,dst)=2$，故 $f_1=f_2=f_3=4$，$f_4=4/2=2$。对于 P2P 流量，它总是 $dif(src,DST)$ 值较大，而 $same(src,dst)$ 值较小，因此 $f$ 值较大；非 P2P 流量则恰恰相反。

依照类似的方式可以给出对源端口和目标端口的函数定义：

$$g(src,spt) = dif(src,SPT) / same(src,spt) \tag{4.3}$$

$$h(src,dpt) = dif(src,DPT) / same(src,dpt) \tag{4.4}$$




(a) 二维特征效果图　　(b) 三维特征效果图

图 4.2　二维与三维特征效果比较

图 4.2 展示的是使用两维特征与三维特征进行识别的效果比较图。从图中可以看出，使用三维特征明显比二维特征更能区分 P2P 和非 P2P 流量。图 4.2(a)是前两维特征构成的二维图像，P2P 与非 P2P 流量在原点附近比较难以区分；而图 4.2(b)展示的是三维特征效果图，从图中可以清晰的看到，P2P 与非 P2P 流量可以通过该三维特征很好的区分开来。

在进行完特征提取之后，下一步应该是核函数的选择。该步骤与特征提取同样重要，因为一个好的核函数能够保证映射到高维空间的特征数据线性可分。目前最常见的核有如

下几种：

(1)多项式核：
$$K(x,x_i)=[(x\cdot x_i)+1]^d \tag{4.5}$$

(2)径向基函数核(RBF 核)：
$$K(x,x_i)=\exp\left(-\frac{|x-x_i|^2}{\sigma^2}\right) \tag{4.6}$$

(3)Sigmoid 核：
$$K(x,x_i)=\tanh[v(x\cdot x_i)+c] \tag{4.7}$$

本节出于以下几点考虑，选择 RBF 核做为 SVM 中的核函数[1,45,79]：

1) 一般来说选取正定核做为映射的核函数会比非正定核效果要好，而 Sigmoid 核是非正定的， RBF 核却是正定核；
2) RBF 核需要调整的参数只有一个，而多项式核需要调整三个参数，相比之下基于 RBF 核的模型更容易控制；
3) RBF 核的值域为(0, 1]，而多项式核的值域为(-∞, +∞]，相比之下 RBF 核的数值范围更小，更容易计算。

### 4.2.2 实验一及结果分析

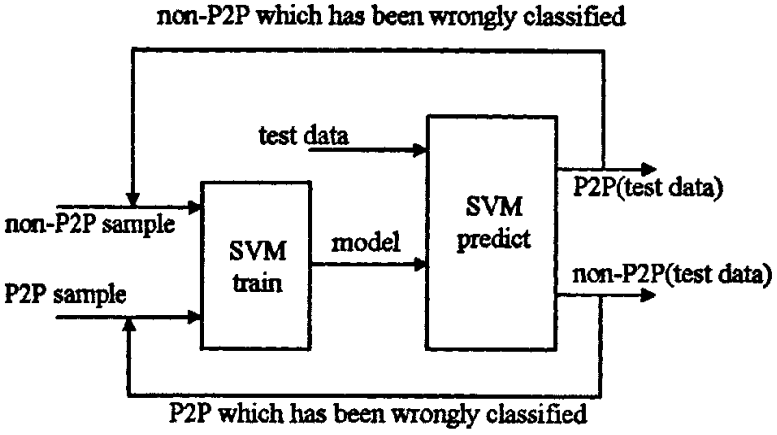



图 4.3 实验一框图

实验一的流程如图 4.3 所示。首先使用一定数量的 P2P 和非 P2P 数据作为训练样本输入到支持向量机之中，根据这些训练数据输出一个模型，这个模型实际上就是通过样本构造的一个决策函数。然后将测试数据输入该模型进行分类，如果非 P2P 流量被识别为 P2P 流量的比例较大，则说明虚警率比较高；相反，如果 P2P 流量被识别为非 P2P 流量的比例较大，则说明漏警率比较高。接下来应该对当前的情况进行反馈：将错分为 P2P 的非 P2P 流量重新归到非 P2P 一类；同时也将被错分为非 P2P 的 P2P 流量重新归到 P2P 一类。循环执行前面的步骤，直到获得理想的精度为止。

在该实验最开始的时候，投入训练的样本并不多，这是为了保证能够迅速的构造决策函数。接下来的反馈过程将错分的样本进行纠正，相当于在下一次训练中加入了先验知识，从而满足了前文中所提到过的四条标准之一：通过学习提高分类性能。随着迭代次数的增加，投入训练的样本越来越多，包含的数据特征也越来越全面；同时由于反馈次数的增多，该学习器的分类性能也随着加入的先验知识不断增强，最终将达到理想的效果。

在使用该框图中的迭代反馈法训练 SVM 的同时，还需要调整 SVM 本身的两个参数值 C 和$\gamma$，使之达到最优。本实验采用的是一种叫做 grid-search 的参数搜索方法[79]。它的处理过程分为两步：第一步，以一定步长确定 C 和$\gamma$需要遍历的值；第二步，遍历 C 和$\gamma$取值的任意一种组合，从中找出最优参数组合。

虽然还有许多其它更精细或者计算复杂度更小的方法可用，但是基于以下三个原因，本实验仍然选择了 grid-search 这一简单而实用的方法[79]：

1. 由于该搜索方法只涉及到两个参数，因此使用它寻找到最优参数组合的时间并不比其它更为精细的方法多很多；
2. 由于参数选取的过程与训练预测过程是相互独立的，因此参数选取所花时间的长短并不会影响到训练和预测的速度；
3. 由于每组参数可以单独计算，因此该方法可以实现并行计算，从而可以大大减少搜索时间。

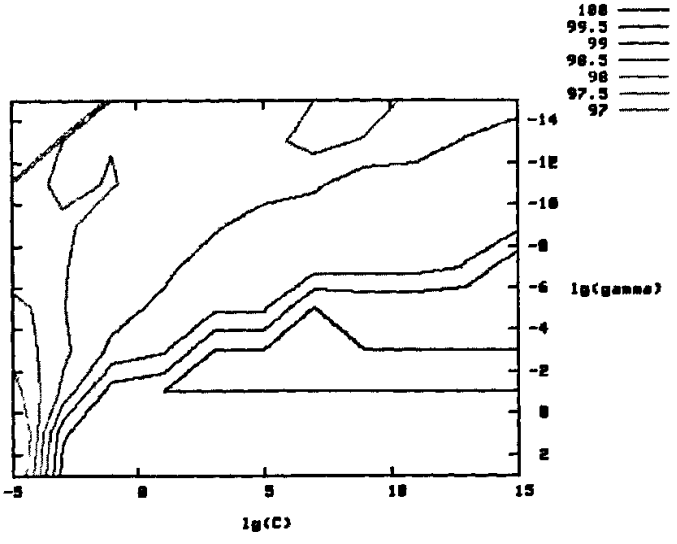



图 4.4 寻找最优参数示意图

图 4.4 是使用 grid-search 方法寻找最优参数的示意图。从图中可以看出，右下角这条颜色最深的曲线使得训练准确率达到 100%，它所对应的参数即为最优参数。

值得注意的是，如果第一次 grid-search 选取的参数不能达到令人满意的精度，则可以通过扩大搜索范围的方法或使用更小的步长来进行下一次搜索，直到符合精度要求的参数

组合出现为止。

表 4.3 是使用迭代反馈法进行流量检测时所获得的结果。使用如下定义确定表中虚警率（false positive）与漏警率（false negative）：

$f_1$为被正确分类的 P2P 流量；$f_2$为被错分到 P2P 中的非 P2P 流量；$f_3$为被错分到非 P2P 中的 P2P 流量。

虚警率 $= f_2/(f_1+f_2)\times 100\%$，漏警率 $= f_3/(f_1+f_3)\times 100\%$。

该实验一共分为 6 组。第 1 组、第 2 组、和第 5 组实验都只使用了前两维特征，其余的 3 组实验则使用了三维特征。第 1 组到第 4 组所使用的测试数据集为 SetD（表 4.1），其中第 1 组和第 3 组实验是没有通过迭代反馈直接得到的结果，从表中可以看出，这两次实验所得的虚警率和漏警率相对来说都比较高。而第 2 组和第 4 组实验都是经过了一次迭代反馈之后的结果，从表中可以看出，它们的虚警率和漏警率已经大大降低。第 5、6 组实验使用的是大测试数据集 SetE（表 4.1）。表中的数据表明通过迭代反馈之后的检测系统已经达到了较高的识别精度。

| | 向量维数 | P2P 样本 | non-P2P 样本 | 测试集 | 虚警率 | 漏警率 |
|---|---|---|---|---|---|---|
| 1 | 2 | Set A | Set B | Set D | 3.807% | 22.701% |
| 2 | 2 | Revised Set A | Revised Set B | Set D | 1.3% | 2.7% |
| 3 | 3 | Set A | Set B | Set D | 10.9% | 20.3% |
| 4 | 3 | Revised Set A | Revised Set B | Set D | 0.3% | 3.5% |
| 5 | 2 | Revised Set A | Revised Set B | Set E | 2.3% | 1.5% |
| 6 | 3 | Revised Set A | Revised Set B | Set E | 2.1% | 1.1% |

表 4.3 P2P 流量检测结果比较

图 4.5 是基于 SVM 的检测方法与 Application-signature 检测方法[71]的精度对比示意图。Application-signature 方法是一种基于 payload 的方法，它是由 S. Sen 等人在 2004 年的 world wide web(WWW)会议上提出来的。之所以选择该方法与本章提出的方法进行对比是因为该方法应用广泛并且易于实现。两种方法均选用了数据集 SetD 作为测试数据集，检测结果每隔一小时记录一次。从图 4.5 中可以看出，位于上方的深色曲线来自基于 SVM 的方法，该方法在所有时刻的检测精度都 Application-signature 方法要高。由此可以看出，基于 SVM 的检测方法在实际问题中获得了比原有流量检测方法更高的精度。

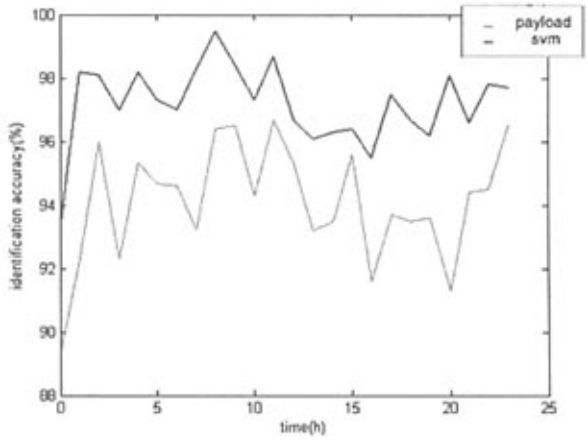



图 4.5 精度对比结果

### 4.2.3 实验二及结果分析

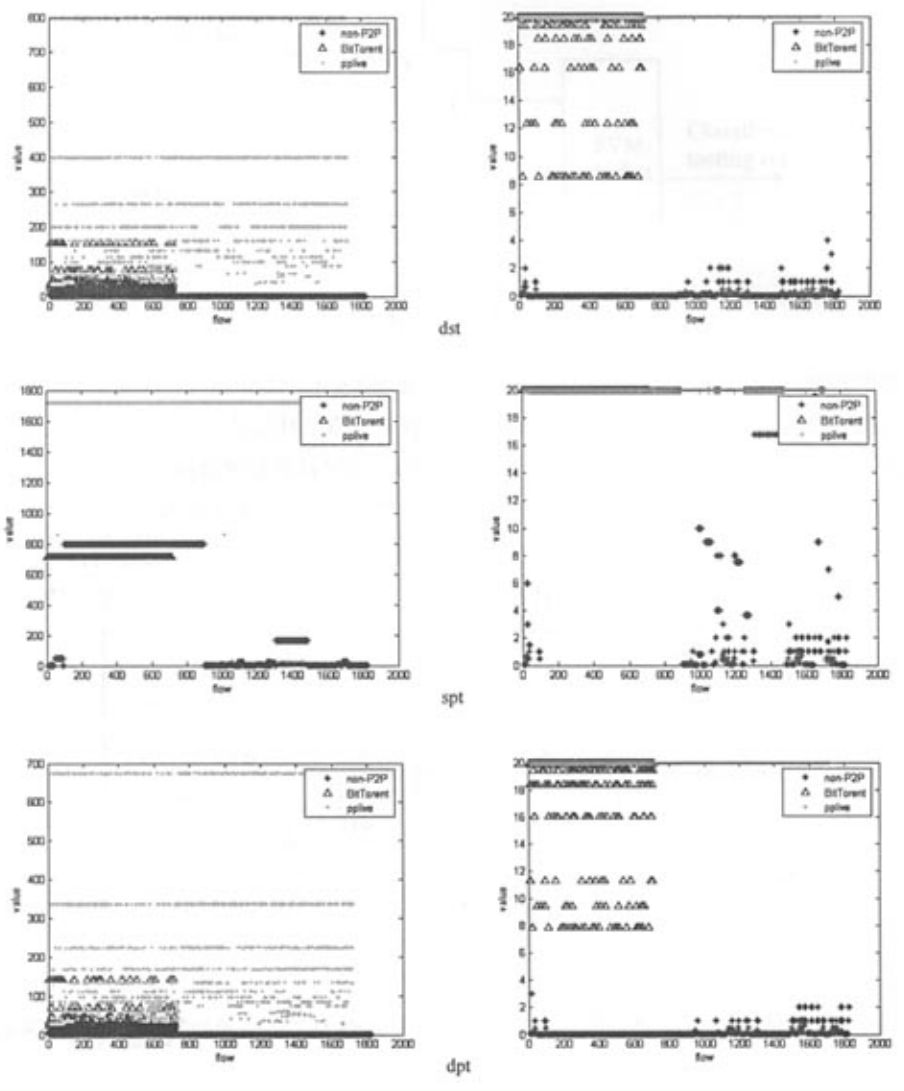
本小节将介绍第二个实验。

当 BT、pplive 两种 P2P 流量以及 web 非 P2P 流量的前三维特征被放在一起进行对比时，通过观察可以发现，虽然 BT 和 pplive 这两种 P2P 流量在这三维特征上都表现出不同于非 P2P 流量的性质，但是在第二维特征上三者之间的关系比较模糊，这有可能使得检测产生困难。为了消除不同 P2P 流量之间的数据模糊带来的困难，同时不增加系统的复杂度，该小节的实验不使用迭代反馈系统，而引入如下的平滑函数：

$$q(x)=\begin{cases} x, 1<x\le t_1 \\ t_2-\dfrac{t_2-t_1}{e^{x-t_1}}, x>t_1 \end{cases} \qquad t_2>t_1 \tag{4.8}$$

该函数中的自变量$x$是平滑处理前任意一维特征的值，而$q(x)$则是该维特征经过平滑处理之后的数值。从式(4.8)中可以看出，平滑后的值$q(x)$在 1 到$t_2$之间变化，若$1<x\le t_1$，则处理后的值不变；若$x>t_1$，则$q(x)$增加的速度呈指数级递减，最大值为$t_2$。因此只要能在检测之前为每一维特征选取合适的$t_1$和$t_2$，就能达到减小各类 P2P 流量之间差异的目的。

图 4.6 展示了平滑处理的效果。左边的三幅图(4.6(a))是平滑处理之前的数据，而右边的三幅图(4.6(b))则是平滑处理之后得到的数据，此次实验选取的$t_1$=10，$t_2$=20。从图中可以看出，经过平滑处理后两类 P2P 数据之间的差异明显减小，而它们与非 P2P 数据之间的差异基本保持不变，从而消除了不同 P2P 流量之间的数据模糊带来的困难，为更好的区分 P2P 与非 P2P 流量提供了条件。




(a) 平滑前的三组数据　　(b) 平滑后的三组数据

图 4.6　平滑处理的效果

图 4.7 为包含了对数据进行平滑处理的实验二框图。为了减小实验复杂度，该框图中不再含有迭代反馈过程。第一步先将训练集进行平滑处理，然后通过训练得到 SVM 分类器，接下来用平滑处理过的测试数据对该分类器的检测效果进行预测。由于平滑处理减小了 P2P 流量之间的特征差异，因此该实验将会获得一个较高的检测精度。

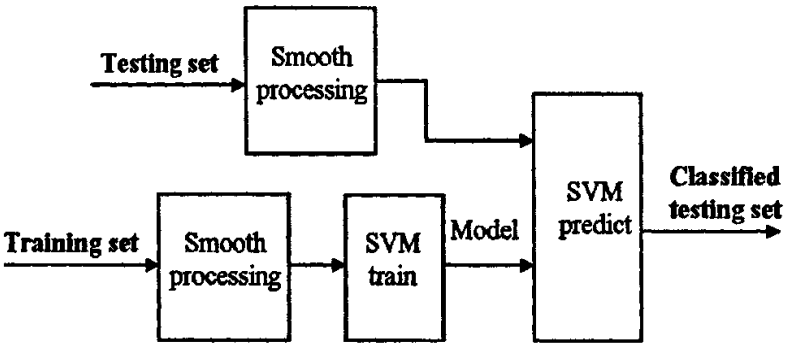



图 4.7 实验二框图

该实验用来进行训练的 Training Set 见表 4.2。与实验一相同，本节的实验也同样需要对 SVM 本身的两个参数值进行选取。通过 grid-search 方法可得：C=2, $\gamma$=0.0078125。表 4.4 是基于原始数据的检测结果与平滑处理后的检测结果之间的对比。从该表中可以看出，经过平滑处理后，虚警率大大减少，尤其是 SetB 与 SetE，虚警率已降低到 0。但该方法使得漏警率稍有增加，不过仍然维持在较低的比率上。因此我们可以认为，经过上述的平滑处理，该检测系统同时拥有了低虚警率与低漏警率的特点。

| | 虚警数 | | 漏警数 | |
|---|---|---|---|---|
| | original | smooth | original | smooth |
| Testing Set A | 481 | 12 | 5 | 29 |
| Testing Set B | 504 | 0 | 0 | 0 |
| Testing Set C | 215 | 14 | 11 | 13 |
| Testing Set D | 839 | 37 | 89 | 90 |
| Testing Set E | 1951 | 0 | 0 | 0 |

表 4.4 检测精度对比

以上两个实验表明，本节设计的基于两类 SVM 的 P2P 流量检测方法精度较高，达到了前文中提出的四个标准，再一次体现了 SVM 在实际分类与检测问题中的卓越性能，并为下一节多类 SVM 的成功应用打下了良好的基础。

## 4.3 基于多类 SVM 的 P2P 流量应用级分类

### 4.3.1 问题描述，数据采集与特征提取

本节所需要解决的问题是 P2P 流量的应用级分类。在很多时候，仅仅分辨出 P2P 与非 P2P 流量是不够的，还需要对 P2P 流量的类别进行更为详细的区分，这就涉及到多类分类的问题。考虑到两类 SVM 在 P2P 流量检测中的成功应用，本节使用多类 SVM 方法进行 P2P 流量的应用级分类。与上一节类似，对该系统也提出如下四条标准：

1) 能够检测出主要的 P2P 流量；
2) 能够检测出未知的和加密的 P2P 流量；
3) 能够顺利完成流量较大时的识别工作；
4) 能够顺利完成 P2P 流量的应用级分类。

在实验数据方面，和上一节一样，采集的数据仍然来自于从校园网络上的收集到的 netflow 数据。如表 4.5 所示，该实验收集了 5 组数据。前四组数据集 A，B，C，D 包含的分别是 BT，eDonkey，Kazaa 以及 pplive 流量，这些都属于 P2P 流量。而第 5 组数据集包含的是大量的非 P2P 流量。每组数据采集都花费 12 小时，并且每隔 15 分钟汇聚一次。

| | 开始时间 | 总时间 | 间隔时间 | PCs |
|---|---|---|---|---|
| **Set A** | 10:00 05/14 | 12 hour | 5 min | 1 |
| **Set B** | 10:00 05/15 | 12 hour | 5 min | 1 |
| **Set C** | 10:00 05/16 | 12 hour | 5 min | 1 |
| **Set D** | 10:00 05/17 | 12 hour | 5 min | 1 |
| **Set E** | 10:00 05/18 | 12 hour | 5 min | 10 |
| | **P2P 流量比例** | **P2P 流量类型** | **Flows** | **Bytes** |
| **Set A** | 100% | BitTorrent | 30.68k | 337.26M |
| **Set B** | 100% | eDonkey | 2.72k | 629.82M |
| **Set C** | 100% | Kazaa | 4.43k | 22.548M |
| **Set D** | 100% | pplive | 58.15k | 168.62M |
| **Set E** | 0% | -- | 460.24k | 799.92M |

表 4.5 实验数据采集

接下来的特征提取与上一节稍有不同。在原来三维向量的基础上又增加了一维特征，形成了一个 4 维特征向量，其表达式为：

$$vector(flow) =< f(src,dst), g(src,spt), h(src,dpt), bytes / pkts > \quad (4.9)$$

增加的第四维特征在进行 P2P 流量内部间的检测时起着重要的作用。

本次实验依然选用了 RBF 核做为 SVM 中的核函数。

### 4.3.2 实验设计与结果分析

一般说来，在 P2P 流量检测问题当中，虚警率和漏警率都是检验该系统精度的重要标准。但通过进一步的观察可以发现，如果检测方法的虚警率较高，那么在根据该虚警率严格控制 P2P 流量的同时将会导致许多非 P2P 流量也被错误的控制，从而对非 P2P 用户的网络活动造成较大的影响；另一方面，如果检测方法的漏警率较高，它造成的影响主要在于少控制了部分 P2P 流量，但由于大部分 P2P 流量还在其控制之下，因此系统仍然能够正常工作。与此同时，较高的漏警率也不会对非 P2P 用户的网络活动造成影响。因此该方法设

计的检测框架应该优先考虑检测系统的虚警率。

另外，P2P 流量可以被分为两个部分：数据流量与控制流量。数据流量包括了所有用户的数据传输，其它流量则都是控制流量。而一个检测算法的好坏主要取决于数据流量比例的大小。因此数据流量的精确识别显得更为重要。

基于上面的讨论，本节提出了如下两条设计原则：

- 优先考虑降低虚警率；
- 优先考虑精确识别数据流量。

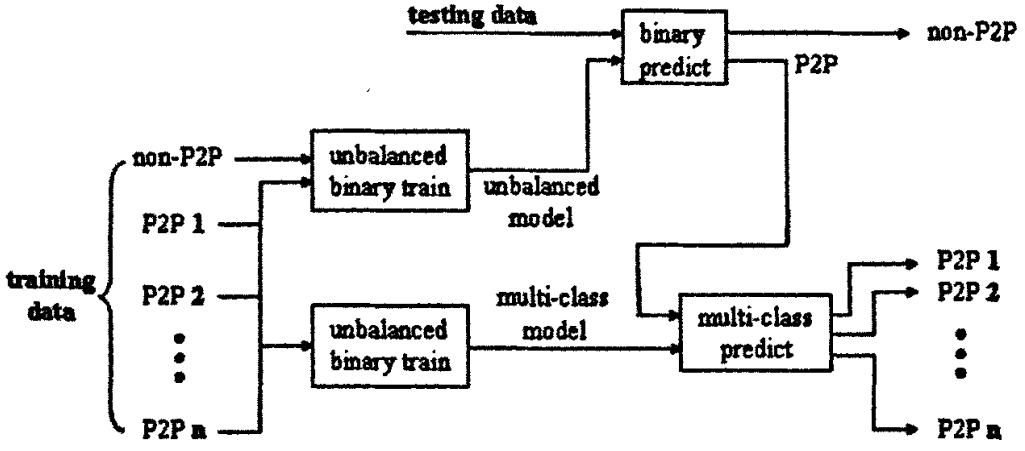



图 4.8 实验框图

根据以上的两条原则，本节构造了如图 4.8 所示的实验框图。具体流程如下：

首先将训练样本中的所有 P2P 流量与非 P2P 流量混合在一起，进行两类识别，通过该过程构造非平衡两类 SVM 分类器。基于简单和易于实现的原则，这里非平衡 SVM 分类器的构造只是根据 P2P 与非 P2P 流量的不同重要性对权值进行了设置。与此同时，将 $n$ 种不同的 P2P 流量通过加权多类 SVM 进行训练，完成应用级多类识别器的构造。在这两个过程中，需要根据不同流量的重要性对权值分别进行合理的设置。接下来将使用未知的预测数据对非平衡两类 SVM 模型进行性能测试，并输出测试结果。在输出测试结果的同时，再将被识别为 P2P 的流量作为输入，对加权多类 SVM 模型进行性能测试，并输出应用级分类结果。这样一来，输入的一组未知数据在输出的时候将分为 $n+1$ 组，其中包括 $n$ 组不同的 P2P 流量和一组非 P2P 流量。

该系统框架有如下两个优点：第一，它使用了非平衡两类 SVM 识别器，满足了第一条设计原则中优先减少虚警率的要求；第二，该框架通过加权多类 SVM 对数据流量与控制流量的重要性进行了不同程度的设置与处理，实现了第二条设计原则。

接下来将进行检测效果的对比。首先使用 Application-signature 方法和 Transport-layer 方法[75]跟本节提出的方法进行对比。如上一节所说，Application-signature 方法是一种基于

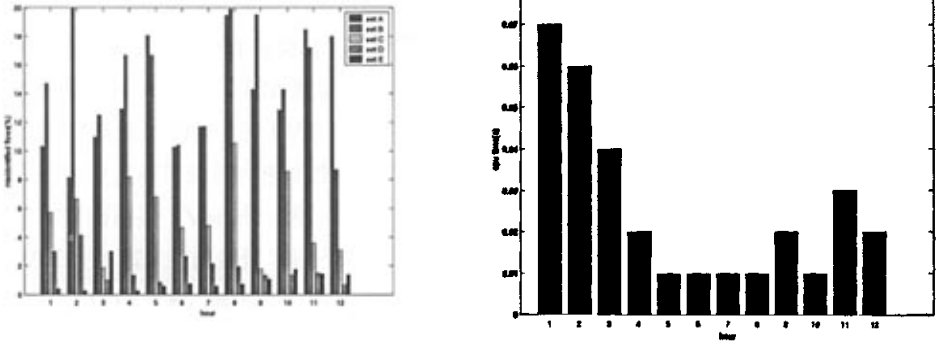
payload 特征的方法。而 Transport-layer 方法与基于 SVM 的方法一样，是基于流量统计特征的方法。选择这两种方法进行对比的原因与上一节类似，它们不但易于实现，同时在实际问题中也应用广泛。

| 方法类型 | 应用级分类 | 是否不需要私有信息 | 检测加密与未知 P2P 流量 |
|---|---|---|---|
| Signature | 是 | 否 | 否 |
| Transport Layer | 否 | 是 | 是 |
| Proposed method | 是 | 是 | 是 |

表 4.6 三种方法对比结果

比较结果在表 4.6 中给出。Application-signature 方法可以对流量进行应用级分类，但需要私有信息，同时不能检测加密与未知的 P2P 流量；而 Transport-layer 方法虽然不需要私有信息，而且能够检测加密与未知的 P2P 流量，但却不能对流量进行应用级分类；而基于 SVM 的方法综合了前两者的优点，拥有了较强的分类性能。

下面的实验对基于 SVM 方法的精度与效率进行了测试。首先将采集的数据集 A，B，C，D，E 依照式(4.9)进行特征选择，然后将处理过的每个数据集分成 12 个子集，每个子集包含一个小时的流量。接下来从这五个数据集中各提取一个子集。从数据集 E 中提取的子集用来做非 P2P 流量的训练数据，而从数据集 A，B，C，D 中提取的子集分别用来做第一种 P2P 流量到第四种 P2P 流量的训练集。将选出来的 5 个子集按照实验框图 4.8 的方法进行训练，并用全部 60 组子集作为输入数据进行测试，同时详细的记录输出数据以及整个过程花费的时间。



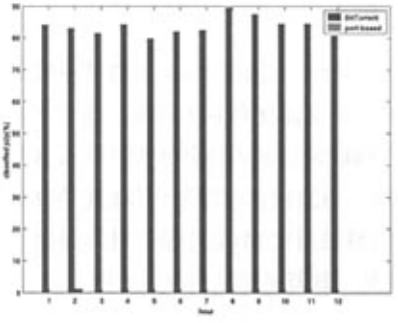
(a) 预测精度　　(b) 预测效率

图 4.9 两类预测结果示意图

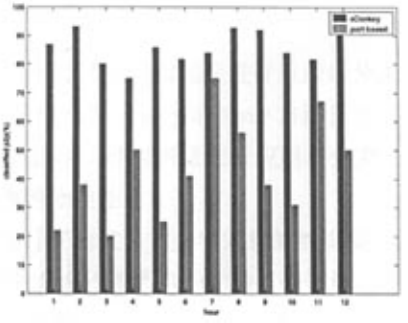
图 4.9 展示了非平衡两类 SVM 的识别精度与效率。图 4.9(a)记录了 5 个数据集在 1 小时内错分的比例，每个坐标值上从左到右五条竖线分别代表数据集 A 到数据集 E 的错误率。从图中可以清晰的看到，数据集 E 的错误率是五个数据集中最小的，这表明前文中提出的

非平衡思想成功的得以实现。图 4.9(b)是展示了数据集 E 的子集所花费的预测时间。用来进行实验的机器配置为 Celeron 3.0G 的 CPU 和 512M 内存。其中最长的预测时间为 0.07 秒，最短为 0.01 秒，这应该算一个很短的时间。值得注意的是，数据集 E 的流量是这五个数据集中最多的。因此其他数据集所需要的预测时间必定更短。

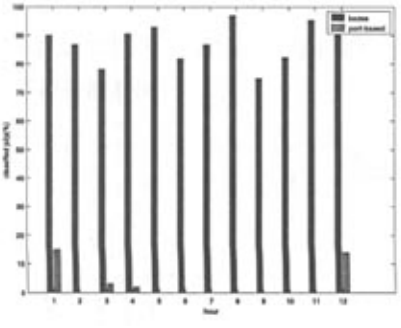
接下来将对多类分类的精度进行评价。本实验选择了基于端口的方法和基于 SVM 的方法进行比较。之所以选取基于端口的方法是因为该方法是应用级流量检测中最常用的方法之一。



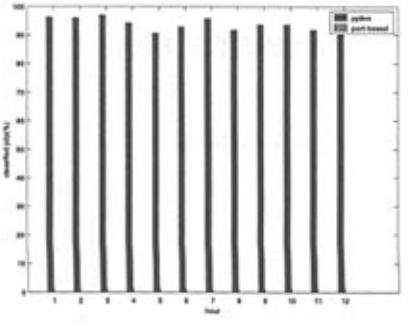
(a) BitTorrent



(b) eDonkey



(c) Kazaa



(d) pplive

图 4.10 多类 SVM 预测结果比较

识别的结果比较如图 4.10 所示。从图 4.10(a)和图 4.10(c)中可以看出，基于端口的方法对 BT 与 Kazaa 流量的识别率几乎为 0。而使用基于多类 SVM 的方法对这四种流量的识别率最低也有 75%，最多甚至可以达到 99%。

以上的几组实验数据可以表明，融合了多类 SVM 的流量检测模型不仅能够精确的识别 P2P 与非 P2P 流量，还能很好的区分不同的 P2P 流量，成功的完成了 P2P 流量的应用级分类。这个事实说明了与机器学习方法相结合的流量检测系统明显比基于传统方法的检测

系统有着更高的分类精度。

值得我们进一步思考的是，在机器学习与模式识别中还有许多其他优秀的学习和分类方法，既然基于 SVM 的方法能够在该应用问题上获得成功，那么这些其他的优秀方法也应当能够获得不错的效果。因此我们可以考虑在今后的工作中将一些合适的方法也应用到该系统当中来，并与基于 SVM 的方法进行对比，这样的结果将更有说服力。

## 4.4 本章小结

本章将 SVM 算法应用到了 P2P 流量检测的实际问题当中，创造性地提出了两种基于 SVM 的 P2P 流量检测新方法。

一开始作者先介绍了应用的背景——P2P 流量及其检测的相关内容，以及本章的出发点——为什么利用 SVM 做流量检测。接下来的一节作者提出了基于两类 SVM 的 P2P 流量检测方法并构造了两个实验框架。第一个实验框架包含了一个反馈系统，使得精度在通过为数不多的几次反馈之后有了很大的提高；第二个实验框架对数据进行了平滑处理，在弱化了噪声影响的同时增强了类可分性。实验的结果表明该方法具有比传统检测方法更高的精度与效率，并达到了构思时提出的四个标准。

在第一种方法成功应用之后，本章又提出了第二个思路：利用多类 SVM 算法来完成对 P2P 流量的应用级分类。该方法利用非平衡两类 SVM 及加权多类 SVM 对非 P2P 流量和多种不同的 P2P 流量进行应用级分类，实验结果再一次证明了 SVM 与流量检测问题的融合可以取得很好的效果。

# 第五章 结论与展望

本章对整篇论文的主要工作与所得结论进行了总结性的回顾，并对以后的工作以及研究方向进行进一步展望。

## 5.1 全文工作总结与回顾

支持向量机是上世纪 90 年代才出现的一种新型学习机器，它的理论基础是基于小样本的统计学习理论。本文主要研究了支持向量机的基本性质及其在流量检测与分类问题中的应用。论文的主要贡献一共有四点，理论与实际应用部分各包含两点：

1．理论部分：在详细回顾了支持向量机的基本性质及相关内容之后，本文提出了两种多类 SVM 的改进：基于 1-vs-all 方法的改进和基于 1-vs-1 方法的改进。在基于 1-vs-all 方法的改进中，主要的思想是通过 “排除法”，减少分类平面的个数以及每次参与训练的样本数，从而达到提高分类速度与精度的目的。第二种基于 1-vs-1 方法的改进实际上是把一个多类问题分解成多个三类问题，然后再进行求解。实验证明该方法能够在保证合理精度的前提下大大提高预测速度。

2．应用部分：通过研究 P2P 流量检测问题的背景，我们发现目前针对此问题的解决方法总存在着这样或那样的缺陷，而该问题从本质上来说可以算是分类问题的一种，因此本文考虑使用目前最好的分类器之一：支持向量机来解决这一系列的问题。本论文提出了两种基于SVM的流量检测新方法。第一种是用两类SVM解决P2P流量中的两类识别问题。该方法还融合了反馈与数据平滑处理的思想来提高精度与效率，并得到了实验数据的有力证明。第二种方法是利用多类 SVM 解决 P2P 流量检测中的多类识别问题。里面还用到了加权 SVM 的思想。实验数据再一次证明了新方法在各个方面都要优于传统检测方法。

以上是该论文主要工作的一个简单回顾，下面的一节将对本论文尚未完成的工作和值得进一步研究的内容进行展望。

## 5.2 进一步研究展望

作为机器学习最重要的内容之一，尽管 SVM 已经在模式分类问题中取得了巨大的成功，但仍然有许多地方值得进一步挖掘与研究。作者下一步的工作将主要集中在以下几个方面：

1．多类 SVM 算法精度的提高。

在本论文针对多类 SVM 进行的两种改进中，预测速度固然是得到了极大提高，但是精度方面却没有特别明显的进步。如果能够针对各类样本不同的重要性对选取样本类别以及构造分类平面的顺序进行专门的研究，甚至提出一般化的定理，那么精度方面必将获得巨大进步。

2．已有改进算法与流量检测问题的进一步结合。

作者在将 SVM 算法应用到 P2P 流量检测问题当中的时候，使用的都是较为常规的 SVM 算法。只是在应用过程中根据检测问题本身的需要对某些常规的 SVM 算法或该应用框架进行了一定改进，如加入反馈系统，使用加权 SVM 等。如果在今后的工作中能将本文提到的两种多类 SVM 改进很好的融合到流量检测问题当中去，同时对如何构造非平衡与加权 SVM 提出更为精细的准则，那么必定能使该应用系统的性能更上一层楼。

3．核函数的选取与构造。

目前 SVM 方法中最重要的问题之一就是核函数的选取与构造。虽然已经有许多研究人员提出了各种各样的选取与构造的方法，但总的来说这方面的研究还处于探索的阶段，还有许多值得深入研究的问题。例如在流量检测问题当中，如果能根据数据本身的特性，构造出一个最适合该问题的核函数，那么效果肯定比根据一般经验选取的常用核函数要更加理想。

4．SVM 与微分流形或拓扑学思想的融合。

SVM 与核方法是通过将低维空间中不可分的数据映射到高维空间使其线性可分，有时候映射或者选取核函数的过程会遇到一定的困难。我们可以考虑将本应该在高维空间中完成的任务放到低维弯曲空间来进行，很可能获得更好的效果。这就需要与微分流形或拓扑学进行有机的结合。相信这将是一个很有意义的研究方向。

# 致　谢

本文的工作从研究方向的确定，论文的选题到定稿都是在导师曾迎生副教授的悉心指导下完成的。在研究生学习的两年半时间里，曾老师不但对我严格要求，为我提供良好的学习和研究环境，更培养了我独立进行研究的能力，使我能够顺利的完成硕士阶段的学业和科研任务。在此论文完成之际，谨向我的导师曾迎生副教授表示深深的感谢。

同时还要感谢其他关心和帮助过我的所有老师。首先感谢贺汉根教授，是他给予了我继续学习与深造的宝贵机会。贺老师无微不至的关怀与严格的督促使我的数学与英语水平得到了质的飞跃。同时他为我们创造了许多宝贵的学术交流机会和一个有张有弛的学习氛围。特别感谢徐昕副研究员一直以来无私的帮助。在整个硕士阶段的学习以及论文的写作过程中，徐老师给了我许多重要乃至关键性的建议。他多次在研究方向上给予我中肯的意见，并利用自己的宝贵时间为我修改学术论文，使我的学习和研究能力得到了进一步的提高。感谢吴涛老师在许多前沿理论和方法上给予我的重要提示及论文上对我的悉心指导，让我的思维得到很大的启发。还要感谢314教研室的董国华老师，他在数学方面给我的巨大帮助令我难忘。

感谢两年多来一起学习与生活的同学：李健、代凯乾、王剑波、王涛等。感谢他们对我学习与生活上的关心和支持。特别是李健同学和代凯乾同学。通过与李健同学多次交流学习与研究经验，我感觉自己拓宽了知识面，并对自己的研究方向有了更深入的理解。代凯乾同学则是在编程方面给予了我很大的帮助。同时还要感谢同一个实验室的同学们：郑兴林、吴立珍和吴唯一。通过和他们的讨论与交流同样使我颇有收获。

另外，要特别感谢计算机学院的硕士研究生王锐，在与他合作撰写三篇英文论文的过程中，我感觉受益非浅。他扎实的专业基础、深厚的编程功底以及对支持向量机的浓厚兴趣是这三篇论文得以成功发表的保障。

在我漫长的学习与生活过程中，父母的支持和关爱一直是我在学术上持续探索和攀登的动力来源，我的成长和每一点进步都凝聚着亲人的心血。谨以此文献给含辛茹苦养育我、关心我的父母，并希望他们永远幸福快乐。

# 参考文献

[1] V. Vapnik, The Nature of Statistical Learning Theory, New York: Springer-Verlag, 1995.

[2] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[3] B. E. Boser, I. M. Guyon, and V. Vapnik. A Training Algorithm for Optimal Margin Classifiers. In D. Haussler, editor, Proceedings of the Annual Conference on Computational Learning Theory, pages 144-152, Pittsburgh, PA, July 1992. ACM Press.

[4] C. Cortes and V. Vapnik, Support Vector Networks, Machine Learning, 20:273-297, 1995.

[5] V. Vapnik, S. Golowich, and A. Smola, Support Vector Method for Function Approximation, Regression Estimation and Signal Processing, in Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 1997, vol. 9.

[6] A. Smola and B. Scholkopf, On A Kernel-based Method for Pattern Recognition, Regression, Approximation and Operator Inversion, Algorithmica, 22:211-231, 1998.

[7] C. Bahlmann, B. Haasdonk and H. Burkhardt, On-line Handwriting Recognition with Support Vector Machines-A Kernel Approach, In Proceeding of the 8th Int. Workshop on Frontiers in Handwriting Recognition (IWFHR), pp: 49–54, 2002.

[8] M. Pontil and A. Verri, Support Vector Machines for 3D Object Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 637-646, June 1998.

[9] K. S. Goh, E. Y. Chang, and B. Li, Using One-Class and Two-Class SVMs for Multiclass Image Annotation, IEEE Transactions on Knowledge and Data Engineering, 1333-1346, October 2005.

[10] A. Graf and C. Wallraven, Multi-class SVMs for Image Classification using Feature Tracking, Technical Report No. 099, Max Planck Institute for Biological Cybernetics, August 2002.

[11] G. Guo and S. Z. Li, Content-Based Audio Classification and Retrieval by Support Vector Machines, IEEE Transactions on Neural Networks, 209-215, January 2003.

[12] A. Ganapathiraju, J. E. Hamaker and J. Picone, Applications of Support Vector Machines to Speech Recognition, IEEE Transactions on Signal Processing, 2348-2355, August 2004.

[13] P. J. Phillips, Support Vector Machines Applied to Face Recognition. Advanced in Neural Information Processing System, MIT Press, 803-809, 1998.

[14] E. Osuna, R. Freund and F. Girosi, Training Support Vector Machines: an Application to Face Detection, Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 97), 1997.

[15] B. Heisele, P. Ho and T. Poggio, Face Recognition with Support Vector Machines: Global versus Component-based Approach, Proceedings of IEEE International Conference on Computer Vision (ICCV), 688-694, 2001.

[16] L. M. Manevitz and M. Yousef, One-Class SVMs for Document Classification, Journal of Machine Learning Research, 139-154, 2001.

[17] H. Drucker, D. Wu and V. Vapnik, Support Vector Machines for Spam Categorization, IEEE Transactions on Neural Networks, 1048-1054, September 1999.

[18] S. Avidan, Support Vector Tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1064-1072, August 2004.

[19] P. Paysan, Stereovision Based Vehicle Classification Using Support Vector Machines, Master Thesis, MIT, February 2004.

[20] R. K. Begg, M. Palaniswami and B. Owen, Support Vector Machines for Automated Gait Classification, IEEE Transactions on Biomedical Engineering, 828-838, May 2005.

[21] B. Sun, D. Huang and H. Fang, Lidar Signal Denoising Using Least-Squares Support Vector Machine, IEEE Signal Processing Letters, 101-104, February 2005.

[22] I. E. Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos and R. M. Nishikawa, A Support Vector Machine Approach for Detection of Microcalcifications, IEEE Transactions on Medical Imaging, 1552-1563, December 2002.

[23] K. I. Kim, K. Jung, S. H. Park and H. J. Kim, Support Vector Machines for Texture Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1542-1550, November 2002.

[24] Drucker, C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support Vector Regression Machines, In: Neural Information Processing Systems, Vol. 9. MIT Press, Cambridge, MA, 1997.

[25] B. Scholkopf, Support Vector Learning, Ph.D. Thesis, Berlin University, 1997.

[26] C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Knowledge Discovery and Data Mining, vol. 2, pp. 121–167, June 1998

[27] N. Cristianini and J. S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, Cambridge, UK, 2000.

[28] A. Smola, Regression Estimation with Support Vector Learning Machines, M.S Dissertation, Technische Universit t München, 1996.

[29] A. Smola, Learning with Kernels, Ph.D. Thesis, Technische Universität Berlin, 1998.

[30] S. Amari and S. Wu, Improving Support Vector Machines by Modifying Kernel Functions, Technical report, RIKEN, 1999.

[31] C. Burges and B. Scholkopf, Improving the Accuracy and Speed of Support Vector Learning Machines. In: Advances in Neural Information Processing Systems 9, pages 375–381, Cambridge, MA, MIT Press, 1997.

[32] G. Cauwenberghs and T. Poggio, Incremental and Decremental Support Vector Machine Learning, In Advances in Neural Processing Systems, 2001.

[33] B. Scholkopf, A. Smola, R. Williamson, and P. L. Bartlett, New Support Vector Algorithms, Neural Computation, 12:1207–1245, 2000.

[34] C. Chang and C. J. Lin, Training $v$-Support Vector Classifiers: Theory and Algorithms, Neural Computation, 13(9): 2119–2147, 2001.

[35] I. Steinwart, On the Optimal Parameter Choice for $v$-Support Vector Machines. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003.

[36] A. Chalimourda, B. Scholkopf, and A. Smola, Choosing $v$ in Support Vector Regression with Different Noise Models — Theory and Experiments, In Proceedings of the International Joint Conference on Neural Networks, Como, Italy, 2000.

[37] N. Cristianini, C. Campbell, and J. S. Taylor, Dynamically Adapting Kernels in Support Vector Machines, In Advances in Neural Information Processing Systems, volume 11, MIT Press, Cambridge, MA, 1999.

[38] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, Choosing Kernel Parameters for Support Vector Machines, Machine Learning, 2002.

[39] T. Joachims, Making Large-scale Support Vector Machine Learning Practical, In: Advances in Kernel Methods-Support Vector Learning, Massachusetts: The MIT Press, 1999.

[40] J. Platt, Fast Training of Support Vector Machines Using Sequential Minimal Optimization. Advances in Kernel Methods-Support Vector Learning, Cambridge, MA: MIT Press, 1999.

[41] B. Scholkopf, A. Smola and Bartlett, Advances in Large Margin Classifiers, MIT Press, 2000.

[42] E. Osuna and F. Girosi, Reducing Run-time Complexity of SVMs, Proc. 14th International Conference on Pattern Recognition, Brisbane, Australia, 1998.

[43] R. Collobert and S. Bengio, SVMTorch: Support Vector Machines for Large-scale Regression Problems, Journal of Machine Learning Research, 1:143–160, 2001.

[44] H. Yu, J. Yang and J. Han, Classifying Large Data Sets Using SVM with Hierarchical Clusters, SIGKDD'03 Washington, DC, USA, 2003.

[45] B. Scholkopf and A. Smola, Learning with Kernels, MIT Press, 2002.

[46] 邓乃扬，田英杰，数据挖掘中的新方法——支持向量机，北京：科学出版社，2004.

[47] V. Vapnik, Estimation of Dependences Based on Empirical Data, Nauka: 1979.

[48] V. Vapnik and A. Chervonenkis, Theory of Pattern Recognition, Nauka: 1974.

[49] 袁亚湘，孙文瑜，最优化理论与方法，北京：科学出版社，1997.

[50] H. W. Kuhn and A. W. Tucker, Nonlinear programming, In: Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics, pp: 481–492, Berkeley, University of California Press, 1951.

[51] K. Crammer and Y. Singer, On the Algorithmic Implementation of Multi-class SVMs, JMLR, 2001.

[52] Krebel, Pairwise classification and support vector machines, In: Advances in Kernel Methods: Support Vector Machine. MIT Press, Cambrige, MA, pp 255-268, 1998.

[53] J. Friedman, Another Approach to Polychotomous Classification, Dept. Statist., Stanford Univ., Stanford, CA, 1996.

[54] J. C. Platt, N. Cristianini and J. S. Taylor, Large Margin DAGs for Multi-class Classification, In: Advances in Neural Information Processing Systems, vol 12, MIT Press, Cambridge, MA, pp: 547-553, 2000.

[55] C. W. Hsu and C. J. Lin, A Comparison of Methods for Multi-class Support Vector Machines, IEEE Transactions on Neural Networks, 13:415-425, 2002.

[56] R. Rifkin and A. Klautau, In Defense of One-vs-all Classification, Journal of Machine Learning Research 5:101-141, 2004.

[57] J. Weston and C. Watkins, Multi-class Support Vector Machines, Technical Report CSD-TR- 98-04, Royal Holloway, University of London, 1998.

[58] T. G. Dietterich and G. Bakiri, Solving Multi-class Learning Problems via Error-correcting Output Codes, Journal of Artificial Intelligence Research, 2: 263-286, 1995.

[59] L. Xu and D. Schuurmans, Unsupervised and Semi-supervised Multi-class Support Vector Machines.

[60] 李蓉, 叶世伟, 史忠植, SVM-KNN 分类器：一种提高SVM分类精度的新方法, 电子学报, pp: 745-748, 2002.5.

[61] 李昆仑, 黄厚宽, 田盛丰, 模糊多类SVM模型, 电子学报, pp: 830-832, 2004.5.

[62] C. F. Lin and S. D. Wang, Fuzzy Support Vector Machines, IEEE Transactions on Neural Networks, Vol. 13, No. 2, March 2002.

[63] Ping Zhong Masao Fukushima, A New Multi-class Support Vector Algorithm, Optimization Methods and Software, 1–18, February 2005.

[64] 黄景涛, 马龙华, 钱积新, 一种用于多分类问题的改进支持向量机, 浙江大学学报(工学版), 第38卷第12期, 2004.12.

[65] L. Khan, and Q. Chen, Effective Intrusion Detection Using Support Vector Machines, In South Central Information Security Symposium, SCISS, Rice University, Houston, Texas, April 2004.

[66] C. Bregler and S. Omohundro. Nonlinear Manifold Learning for Visual Speech Recognition, In Proc. IEEE ICCV, pp: 494-499, 1995.

[67] M. Sotelo, J. Nuevo, D. Fernandez, I. Parra, L. M. Bergasa, M. Ocana, and R. Flores, SVM-based Obstacles Recognition for Road Vehicle Applications" , pp: 1740-1741, IJCAI 2005.

[68] http://asi.insa-rouen.fr/~arakotom/toolbox/

[69] http://download.joachims.org/svm_multiclass/examples/example4.tar.gz

[70] T. Karagiannis, A. Broido, N. Brownlee, K. Claffy and M. Faloutsos, Is P2P Dying or Just Hiding? In Globecom, Dallas, TX, USA, November 2004.

[71] S. Sen, O. Spatscheck, and D. M. Wang, Accurate, Scalable In-Network Identification of P2P Traffic Using Application Signatures, WWW2004, New York, USA, 2004.

[72] M. Roughan, S. Sen, O. Spatscheck and N. Duffield, Class-of-service Mapping for Qos: A Statistical Signature-based Approach to IP Traffic Classification, In IMC 04: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, New York, USA, 2004.

[73] H. Bleul and E. P. Rathgeb, A Simple, Efficient and Flexible Approach to Measure Multi-Protocol Peer-To-Peer Traffic, IEEE International Conference on Networking, 2005.

[74] F. Constantinou and P. Mavrommatis, Identifying Known and Unknown Peer-to-Peer Traffic, 2005.

[75] T. Karagiannis, A. Broido and M. Faloutsos, Transport Layer Identification of P2P Traffic, Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, New York, USA, pp: 121-134, 2004.

[76] M. S. Kim, H. J. Kang and J. W. Hong, Towards Peer-to-Peer Traffic Analysis Using Flows, Lecture Notes in Computer Science, Springer, Heidelberg, Germany, pp: 55-67, 2003.

[77] I. Dedinski, H. Meer, L. Han and L. Mathy, Cross-Layer Peer-to-Peer Traffic Identification and Optimization Based on Active Networking, in Proceedings of the Seventh Annual International Working Conference on Active and Programmable Networks (IWAN'05), CICA, France, November 21-23, 2005.

[78] T. M. Mitchell, Machine Learning, McGRAW-HILL, 1997.

[79] C. W. Hsu, C. C. Chang and C. J. Lin. A Practical Guide to Support Vector Classification.

# 附　录

作者在攻读硕士期间发表的主要论文：

[1] **Yang Liu**, Rui Wang, Yingsheng Zeng, "An Improvement of One-against-all Method for Multi-class Support Vector Machine", accepted by the *4th IEEE International Conference on Science of Electronics, Technology of Information and Telecommunications (SETIT 2007)*, Tunisia, 2007.3.

[2] **Yang Liu**, Rui Wang, Heyun Huang, Yingsheng Zeng, Hangen He, "Applying Support Vector Machine to P2P Traffic Identification with Smooth Processing", In: *Proceedings of 8th IEEE International Conference on Signal Processing (ICSP2006)*, pp: 1802-1805, 2006.11. **(Indexed by EI and ISTP)**

[3] Rui Wang, **Yang Liu**, Yuexiang Yang, Xiaoyong Zhou, "Solving the App-Level Classification Problem of P2P Traffic Via Optimized Support Vector Machines", In: *Proceedings of Sixth IEEE International conference on Intelligent Systems Design and Applications (ISDA2006)*, volume2, pp: 534-539, 2006.10. **(Indexed by EI and ISTP)**

[4] Rui Wang, **Yang Liu**, Yuexiang Yang, Hailong Wang, "A New Method for P2P Traffic Identification Based on Support Vector Machine", In: *Proceedings of ICGST International Conference on Artificial Intelligence and Machine Learning (AIML06)*, 2006.6.

[5] Xinling Zheng, **Yang Liu**, Yingsheng Zeng, "Signal Processing of Automobile Millimeter Wave Radar Base on BP Neural Network", *ICGST International Conference on Artificial Intelligence and Machine Learning (AIML06)*, June 2006.

[6] Jian Li, **Yang Liu**, Xiangjing An, Hangen He, "A Neighborhood Memory System for Parallel Image Computing", In: *Proceedings of IEEE International Conference on Sensing, Computing, and Automation (ICSCA2006)*, 2006.5. **(Indexed by SCI)**

[7] Heyun Huang, **Yang Liu**, Xiang Pan, "Combined wavelet coefficients for least overlapping and its application in underwater signal classification", *7th International Conference on Theoretical and Computational Acoustics (ICTCA)*, Hangzhou, 2005.9.

[8] Heyun Huang, **Yang Liu**, Xiang Pan, "Underwater signal classification based on wavelet and Kolmogorov complexity", *7th International Conference on Theoretical and Computational Acoustics (ICTCA)*, Hangzhou, 2005.9.