

文章编号: 1007-4619(2005)04-0363-11

# 一种基于支撑向量机的遥感影像 不完全监督分类新方法

刘志刚<sup>1,2</sup>, 史文中<sup>3</sup>, 李德仁<sup>2</sup>, 秦前清<sup>2</sup>

(1. 北京师范大学遥感科学国家重点实验室, 地理学与遥感科学学院, 环境遥感与数字城市北京市重点实验室, 北京 100875;

2. 武汉大学 测绘遥感信息工程国家重点实验室, 湖北 武汉 430079;

3. 香港理工大学 土地测量与地理咨询系, 地球信息研究中心)

**摘 要:** 不完全监督分类是研究在只有目标类训练样本的情况下如何准确地将目标类从数据集中提取出来。在许多遥感应用问题中, 往往只需要从遥感影像中提取某一类地物。如果分类过程中只要选取目标类训练样本, 将节省在训练样本选取过程中的大量人力物力。因此不完全监督分类是一个值得研究的遥感分类问题。提出了一种基于加权无标识样本支撑向量机(WUS-SVM), 并在其基础发展出一种不完全监督分类方法。该方法分3个步骤: (1)在影像中随机选取一定量的无标识样本, 将它们作为具有不同权重的非目标类训练样本; (2)用目标类的训练样本和加权无标识训练样本一起训练WUS-SVM, 得到初步的分类器; (3)利用初步的分类器确定无标识样本的类别, 并与原目标类训练样本一起再次训练SVM得到最终的分类器。通过对模拟数据和遥感影像的分类试验初步证明了该分类方法的有效性。

**关键词:** 不完全监督分类; 基于加权无标识样本的支撑向量机; 支撑向量机; 遥感

**中图分类号:** TP751.1/TP79 **文献标识码:** A

## 1 引 言

在许多遥感应用问题中, 往往只需要从遥感影像中提取出某一类地物。由于给出影像中每类地物的训练样本非常费时费力, 因此如何设计一种分类器, 使其在只给出目标类的训练样本的情况下就能够将该目标类从遥感影像中准确地提取出来, 是一个值得研究的问题。这种分类问题被称为: 不完全监督分类 (Partially Supervised Classification-PSC)<sup>[1]</sup>。

## 2 不完全监督分类

遥感影像分类方法可以大致地分为两大类: 绝对分类法和相对分类法<sup>[2]</sup>。绝对分类法 (如平行六面体法) 首先定义某种距离 (概率) 来衡量一个未知

样本与某类训练样本的相似程度, 然后给出一个阈值, 如果一个未知样本到某个类别的训练样本的距离 (概率) 小于 (大于) 阈值则将该样本分为这类。相对分类法 (如最大似然法) 也要定义一种相似性度量, 但是必须计算出未知样本与每一个类别的训练样本的相似程度, 然后将该未知样本分类为具有最大相似度的那类。显然如果只识别某种类别, 绝对分类法就只需要目标类的训练样本, 因此这类方法可以直接运用于PSC。而相对分类法却必须所有类别的训练样本, 所以不能直接运用于PSC。绝对分类法虽然简单易行, 但其分类的精度很大程度上取决于阈值的选取, 如何自动地选择适当的阈值恰恰是一个难点。

现有提高PSC精度的主要策略是采用某种方法获取非目标类的信息, 然后用相对分类器来解决PSC问题。Jeon和Landgrebe提出了两种基于概率密度估计的方法<sup>[1,2]</sup>。尽管这两种方法都有助于改

收稿日期: 2003-11-03; 修订日期: 2004-08-09

基金项目: 香港 Government RGC, CRC 计划 (3-ZB40)、香港理工大学研究基金 (1.34.9709)、利用空间数据挖掘技术进行新型影像目标识别和自动分类 (4002300D)、国家自然科学基金项目 (40176032) Electronic Publishing House. All rights reserved. <http://www.cnki.net>

作者简介: 刘志刚 (1976—), 男, 博士后, 2004年获武汉大学摄影测量与遥感专业博士学位。现主要从事遥感影像处理与模式识别等方面工作。已发表论文多篇。

进 PSC 的精度,但是它们都假定目标类的概率密度函数已知或者可以通过训练样本准确地估计出来。但在实际运用中,目标类的概率密度函数往往未知,且由于样本有限,难以准确估算。

支撑向量机 (Support Vector Machine SVM) 是一种不通过概率密度估计直接构造分类面的全新的模式识别方法,具有突出的分类性能<sup>[3]</sup>。但 SVM 也是一种相对分类方法,需要同时具有目标类和非目标类的训练样本,所以不能直接运用于 PSC。文献 [4] 提出了一种利用无标识样本 (即类别未知,既可能是目标类,也可能是非目标类) 的映射收敛算法将 SVM 用于 PSC, 但该算法不能够保证找到正确的分类函数<sup>[5]</sup>。

近年来还发展出了一类用于单类分类的支撑向量机算法,如单类支撑向量机<sup>[6]</sup>和支撑向量数据描述<sup>[7]</sup>。这两种方法的出发点完全不同,但如果都采用 RBF 核函数,两者却是等价的<sup>[8]</sup>。由于本文只采用了 RBF 核函数,因此不妨统称它们为 1-SVM。这类算法有两大优点:其一,只需要目标类训练样本,可以直接用于 PSC;其二,同样采用了“核技巧”<sup>[3]</sup>,所以也能够构造出复杂的分类面。但 1-SVM 同时存在两大缺点:其一,其分类精度很大程度上取决于核函数的参数选取,而至今尚无有效的参数选择方法<sup>[8]</sup>;其二,完全没有考虑非目标类的信息,在一定程度上制约其分类精度。

本文提出了一种基于加权无标识样本支撑向量机,并在其基础提出了一种 PSC 方法。该方法分为 3 个步骤: (1) 在影像中随机选取一定量无标识样本,将它们作为具有不同权重的非目标类训练样本; (2) 用目标类的训练样本和加权无标识训练样本一起训练 WUS-SVM, 得到初步的分类器; (3) 用初步分类器确定无标识样本的类别, 并与原目标类训练样本一起再次训练 SVM 得到最终的分类器。

### 3 加权无标识支撑向量机

给定一个模式向量集 (X):

$$x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n}, x_{m+n+1}, \dots, x_{m+n+t} \in R^B \quad (1)$$

其中  $m, n, B \in N$ 。不失一般性,假定前  $m$  个向量是目标类训练样本,记作  $X_L = \{x_1, \dots, x_m\}$ , 接着的  $n$  个向量为未知样本集  $X_U = \{x_{m+1}, \dots, x_{m+n}\}$ , 从中随机选取的无标识样本,记作  $X_0 = \{x_{m+1}, \dots, x_{m+n}\}$ 。PSC 的目标是从未知样本集中识别出属于

目标类的样本。

#### 3.1 最优分类超平面

首先考虑目标类和非目标类在输入空间线性可分的情况,即存在一个超平面将目标类和非目标类样本分开。为了构造该分类超平面,首先将所有的无标识样本视为非目标类的训练样本,并且赋予它们一定的权重以度量其作为非目标类训练样本的可信度。本文将权重  $S_i$  定义为每个无标识样本  $x_i \in X_U$  到目标类训练样本集  $X_L$  的最小距离:

$$S_i = \min_j D(x_{m+i}, x_j), \quad j=1, \dots, m \quad i=1, \dots, n \quad (2)$$

接着寻找一个分类超平面 ( $\langle w, x \rangle + b = 0$ ) 将  $X_L$  以一定的间隔 ( $2/\|w\|$ ) 尽可能地与  $X_U$  分开 (图 1)。两类的训练误差分别定义为:

$$E_L = \sum_{i=1}^m \xi_i = \sum_{i=1}^m |1 - (\langle w, x_i \rangle + b)|_+ \quad (3)$$

$$E_U = \sum_{i=m+1}^{m+n} S_i \xi_i = \sum_{i=m+1}^{m+n} S_i |(\langle w, x_i \rangle + b) + 1|_+ \quad (4)$$

如果  $x > 0$ , 则  $|x|_+ = x$  否则  $|x|_+ = 0$ 。上述定义体现了如下思想:如果无标识样本离目标类训练样本越远,则其作为非目标类训练样本的可信度越高,被误分为目标类所导致的训练误差越大;相反,如果无标识样本越接近目标类训练样本则其可信度越低,被误分所造成的训练误差越小。

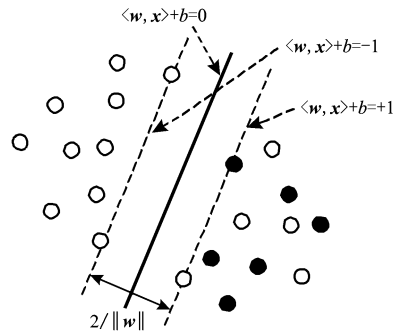


图 1 分类面

(其中黑点为目标类训练样本,白点为测试样本)

Fig 1 Separating plane

假定目标类的训练样本都是正确的,那么首先必须要求分类超平面保持  $X_L$  尽可能完整。即  $E_L$  尽可能小。同时假设目标类样本选取是充分的,即基本上覆盖了目标类在特征空间中的分布区域,此时分类超平面必须靠近目标类训练样本。但如果分类面过分靠近在目标类训练样本,部分目标类样本将

被分类为非目标类,导致漏选,  $E_U$  过小。反之,当分类面过分远离目标类训练样本,部分非目标类样本将被分类为目标类,形成误判,  $E_U$  过大。只有当分类面仅将  $X_U$  中目标类的样本“错误地”分在目标类训练样本一侧,  $E_U$  的大小才是适当的。为了控制  $E_U$  的大小,本文在使  $E_U$  尽可能小的同时,引入一个约束条件——使分类超平面的间隔宽度尽可能大。因为当目标类和非目标类基本可分时,两类样本在特征空间中将存在一定间隙,当分类超平面落在这个间隙中时,其分隔宽度将比较大,同时  $E_U$  较小。显然最大化分隔宽度和最小化训练误差是相互矛盾的两个目标,本文通过引入惩罚参数来协调两者之间的关系。

可以看出最优分类超平面应同时满足 3 个目标:最小化  $E_L$  和  $E_U$ ,以及最大化分类面间隔,即求解如下最优化问题:

$$\begin{aligned} \min_{w \in H, \xi \in R^{m+n}} \tau(w, \xi) &= \frac{1}{2} \|w\|^2 + P_{int} \sum_{i=1}^m \xi_i + P_{others} \sum_{i=m+1}^{m+n} S_i \xi_i \\ \text{s.t.} \quad \forall_{i=1}^m: \langle w, x_i \rangle + b &\geq 1 - \xi_i \\ \forall_{i=m+1}^{m+n}: \langle w, x_i \rangle + b &\geq -1 + \xi_i \\ \forall_{i=1}^{m+n}: \xi_i &\geq 0 \end{aligned} \tag{5}$$

其中  $P_{int}$  和  $P_{others}$  是惩罚参数,均为大于零的常数,它们起到协调 3 个目标的作用。

### 3.2 求解最优化问题

公式 (5) 其实是 SVM 最优化问题的一种推广,因此可以采用与 SVM 类似的方法求解。首先,引入拉格朗日算式

$$\begin{aligned} L(w, b, \xi, \alpha, \beta) &= \frac{1}{2} \|w\|^2 + P_{int} \sum_{i=1}^m \xi_i + P_{others} \sum_{i=m+1}^{m+n} S_i \xi_i - \\ &\sum_{i=1}^m \alpha_i \{ \langle w, x_i \rangle + b - 1 + \xi_i \} - \\ &\sum_{i=m+1}^{m+n} \alpha_i \{ \xi_i - 1 - \langle w, x_i \rangle - b \} - \\ &\sum_{i=1}^{m+n} \mu_i \xi_i \end{aligned} \tag{6}$$

其中  $\alpha (\alpha_1, \dots, \alpha_{m+n} \geq 0)$  和  $\beta (\mu_1, \dots, \mu_{m+n} \geq 0)$  为拉格朗日乘子。由于公式 (6) 中的鞍点即公式 (5) 的解,且在鞍点处公式 (6) 关于  $w, b$  和  $\xi_i$  的偏导都必须为零,因此可以得到如下等式:

$$w = \sum_{i=1}^{m+n} \alpha_i y_i x_i \tag{7}$$

$$\sum_{i=1}^{m+n} y_i \alpha_i = 0 \tag{8}$$

$$\alpha_i = [0.5(y_i + 1)P_{int} - 0.5(y_i - 1)P_{others}S_i] - \mu_i \tag{9}$$

$$\text{其中 } y_i = \begin{cases} +1, & \forall_{i=1}^m \\ -1, & \forall_{i=m+1}^{m+n} \end{cases}$$

将上述 3 个条件等式代入公式 (6) 中,可以得到公式 (5) 的对偶问题

$$\begin{aligned} \max_{\alpha \in R^{m+n}} W(\alpha) &= \sum_{i=1}^{m+n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m+n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{s.t.} \quad \forall_{i=1}^m: 0 &\leq \alpha_i \leq P_{int} \\ \forall_{i=m+1}^{m+n}: 0 &\leq \alpha_i \leq P_{others}S_i \\ \sum_{i=1}^{m+n} \alpha_i y_i &= 0 \end{aligned} \tag{10}$$

将该对偶问题的最优解  $\alpha^*$  代入 (7) 中可以求出原最优化问题的最优解  $w^*$ 。最优分类超平面的偏移量  $b^*$  可以由任意一个位于分类间隔面上的向量求得。最后区分目标类和非目标类样本的决策函数即为:

$$f_{w,b}(x) = \text{sgn} \left( \sum_{i=1}^{m+n} \alpha_i^* y_i \langle x_i, x \rangle + b^* \right) \tag{11}$$

前面考虑的是线性可分的情况,但目标类样本在输入空间中常常被非目标类样本所包围,因此很难通过分类超平面将这两类样本分开。注意到公式 (5) 和公式 (10) 中都只涉及到向量的内积,因此可以采用“核技巧”将上述方法推广到非线性可分的情况。具体而言,就是将上述过程中的内积替换为核函数:

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle, \Phi: R^B \rightarrow H \tag{12}$$

其中  $\Phi$  是将数据映射到高维特征空间  $H$  的某种映射。在高维特征空间中构造的分类超平面对应于输入空间的分类超曲面。通过上述方法得到分类器,本文称之为基于加权无标识样本支撑向量机 (Weighted Unlabeled Sample Support Vector Machine WUS-SVM)。

### 3.3 惩罚参数的选取

由于在 PSC 中目标类训练样本通常绝大多数是正确的,这样  $E_L$  越小越好。可以通过给定一个足够大的目标类惩罚参数  $P_{int}$ ,使任何的  $E_L$  都将受到较大的惩罚,这样可以保证大多数的目标类训练样本被正确地分类。此时最优分类面取决于公式 (5) 中的第一项 (分类超平面间隔宽度) 和第三项 ( $P_{others}$  与非目标类训练样本的训练加权误差—— $E_U$ )。当  $P_{others}$  过大,则  $E_U$  的影响将远远超过分类超平面间隔宽度,从而使分类面过分靠近目标类训练样本,导

致大量漏选;反之,如果  $P_{\text{others}}$  过小,  $E_U$  的作用将被忽视,分类面可能过分地远离目标类训练样本,造成大量误选。因此选择大小合适的  $P_{\text{others}}$  是保证识别精度的关键。

本文提出了一种叠代互检验的方法来选取合适的  $P_{\text{others}}$ 。首先,将目标类训练样本  $X_L$  随机分成  $k$  等份。然后对某个  $P_{\text{others}}$  值做  $k$  次检验。每次检验都用其中  $(k-1)$  份目标类训练样本进行训练,用剩余的样本检测精度。 $k$  次检验的精度平均值作为  $P_{\text{others}}$  评价标准。当精度平均值达到一定标准时,认为当前  $P_{\text{others}}$  值是合适的。在叠代互检验的开始,给  $P_{\text{others}}$  赋一个足够大的初值(例如令  $P_{\text{others}} = P_{\text{int}}$ ),此时分类面将非常靠紧目标类训练样本,导致较低的检验精度;然后逐渐减小  $P_{\text{others}}$  的值,分类面相应地渐渐放松,并逐渐靠近目标类和非目标类的分界位置;当平均检验精度达到要求则停止叠代。

#### 4 基于加权无标识支撑向量机的不完全监督分类

与支撑向量机类似的分类算法其速度均取决于有多少训练样本在决策函数中对应的  $\alpha_i \neq 0$ 。具有非零  $\alpha_i$  值的样本被称为支撑向量。根据二次最优化问题的 KKT 可知,所有没被分类超平面以大于  $2/\|w\|$  的间隔分开的训练样本  $x_i$  都将成为 WUS-SVM 中的支撑向量。这就意味着,当分类面将目标类和非目标类训练样本正确分开时,无标识训练样本中那些属于目标类的样本都将成为支撑向量。如果无标识训练样本中目标类样本数较多时, WUS-SVM 中支撑向量将相应增多,降低分类速度。因此 WUS-SVM 在一定程度上失去了 SVM 原有的稀疏性。

其实在得到 WUS-SVM 之后,无标识训练样本的类别已经可以确定,这样就有了目标类和非目标类的两类训练样本,可以用 SVM 算法再训练,得到最终的分类器。它不但具有 WUS-SVM 的精度,而且支撑向量明显减少,分类速度显著提高。因此本文提出四步走的分类方法(图 2),首先利用无标识训练样本训练得到 WUS-SVM;然后用 WUS-SVM 确定无标识训练样本的类别;接着用 SVM 算法再训练得到最终的分类器,本文称之为再训练支撑向量机(Retained Support Vector Machine RT-SVM);最后用 RT-SVM 对所有未知样本进行分类。

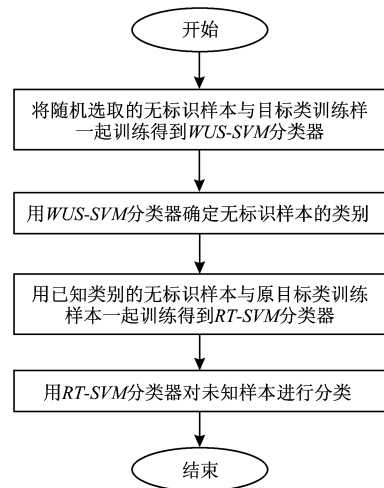


图 2 本文不完全监督分类流程图

Fig 2 Flowchart of proposed partially supervised classification method

### 5 实验与讨论

为了验证本文提出的分类方法的性能,本文分别对二维模拟数据和遥感影像进行了分类实验。对分类结果的评估采用了以下指标:

漏选率  $(E_o) = P\{x \text{ 被分类为非目标类} | x \text{ 为目标类}\}$

误选率  $(E_c) = P\{x \text{ 被分类为目标类} | x \text{ 为非目标类}\}$

平均误差  $(E_a) = E_o * P\{x \text{ 为目标类}\} +$

$E_c * P\{x \text{ 为非目标类}\}$

其中  $P\{*\}$  表示事件“\*”的出现概率。

#### 5.1 基于模拟数据的实验

实验中目标类由 2 个二元正态分布构成,非目标类的由 3 个二元正态分布构成,分别生成了 250 个目标类样本和 750 个非目标类样本(图 3),从中各随机选出 20% 的样本作为训练样本。另外从整个样本集中随机选取 20% 的样本作为无标识训练样本。

实验中还同时采用了 SVM 和 1-SVM 两种分类器作为对比。在 SVM 的训练中利用了两类的训练样本,算法采用标准的 C-SVM 算法,其中 C 固定为常用值 100。在 1-SVM 的训练中只需用目标类训练样本,其实现采用 Libsvm<sup>[9]</sup> 中的标准程序。1-SVM 的参数  $\nu$  表示在目标类训练样本集中被分类为非目标类的样本所占的最大比例<sup>[6]</sup>。由于本文实验中目标类训练样本基本上是正确的,因此  $\nu$  被设为 0.01。在 WUS-SVM 的训练中同时采用了目标训练

样本和无标识训练样本,其中  $P_{in}$  被固定为 100,  $P_{others}$  则根据前面所述的互检验算法自动获取。在 RT-SVM 的训练中,采用标准的 C-SVM 算法,其中 C 固定为常用值 100。这 3 种分类器均采用了 RBF 核函数:  $K(x_i, x_j) = \exp\{-\gamma\|x_i - x_j\|^2\}$ 。为了检验它们对参数选取的敏感程度,核参数  $\gamma$  采用了 3 个值: 1, 10 和 100。

图 3 给出了 3 种算法采用的不同  $\gamma$  值所构造的分类面。可以看出各分类器的分类面都随着  $\gamma$  的增

大而逐渐收紧。尽管如此, SVM, WUS-SVM 和 RT-SVM 的平均分类精度始终维持在 85% 以上,且三者的分类面非常接近。而 1-SVM 的分类精度对参数  $\gamma$  的取值非常敏感。当  $\gamma$  取 1 时,1-SVM 的分类面接近于圆形,误选率达 20% 左右;随着参数值增大,误选率降低,但漏选率上升;当  $\gamma$  为 100 时,1-SVM 的分类面紧紧地收缩在目标类训练样本周围,漏选率升至 35% 左右。

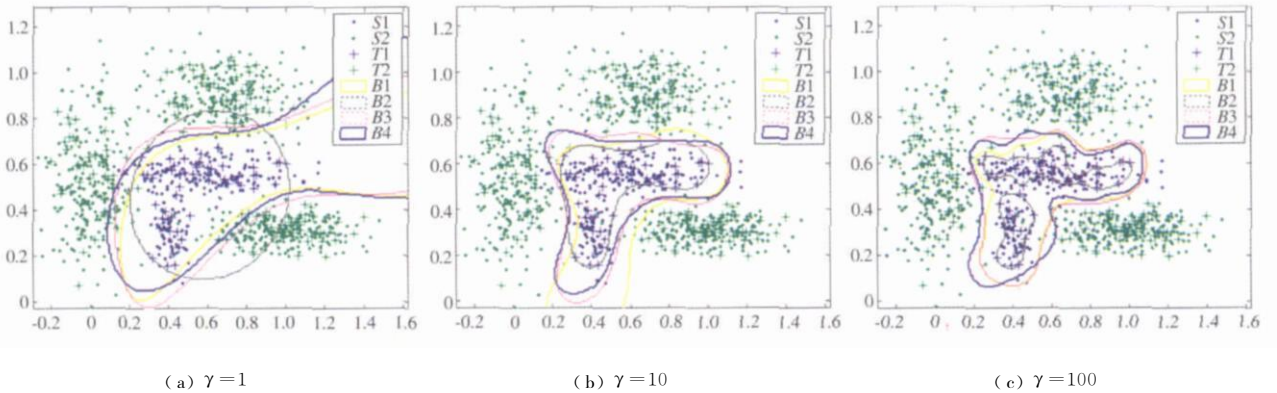


图 3 模拟数据实验中多种算法的分类结果

Fig 3 Classification results of different algorithms in the experiment of simulated data

(S1: 目标类样本; S2: 非目标类样本; T1: 目标类训练样本; T2: 非目标类训练样本; B1: SVM 分类面; B2: 1-SVM 分类面; B3: (采用 20% 无标识样本) WUS-SVM 分类面; B4: (采用 20% 无标识样本) RT-SVM 分类面。)

为了分析无标识样本的数量对 RT-SVM 的影响,实验中还采用随机选取的 10% 的样本以及所有样本作为无标识样本用于训练 WUS-SVM,得到的 RT-SVM 分类面如图 4。从采用不同  $\gamma$  值的 3 组实验结果均可以看出,当无标识样本由 20% 减少到 10% 时,RT-SVM 的分类面有所松弛,漏选率略有上

升;当无标识样本由 20% 增加到 100% 时,RT-SVM 的分类面几乎没有发生变化。然而随着无标识样本数的增加,WUS-SVM 的支撑向量数却不断增加。RT-SVM 的支撑向量数量虽然明显少于 WUS-SVM,但也随无标识样本数的增加而不断上升(表 1)。

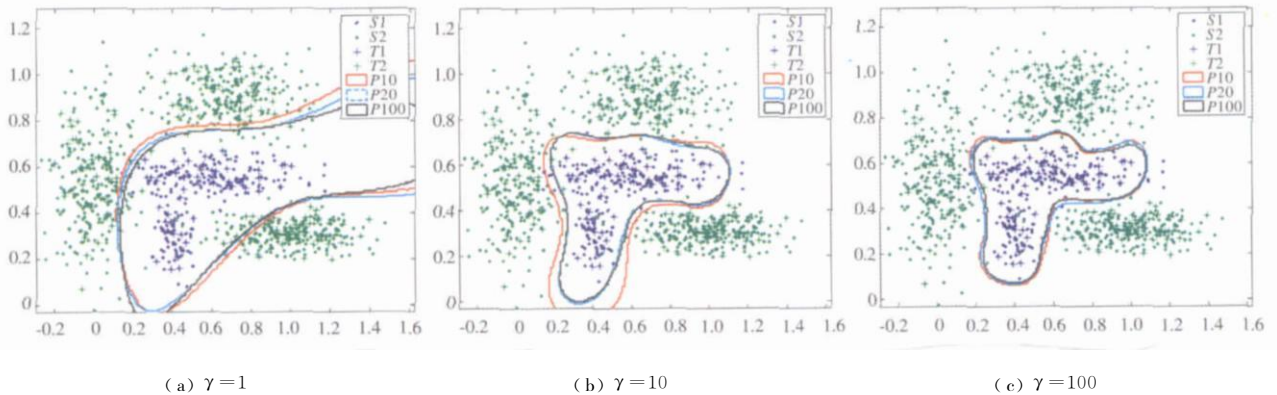


图 4 模拟数据实验中采用不同数量的无标识样本时 RT-SVM 的分类结果

Fig 4 Classification results of RT-SVM with different amount of unlabeled sample in the experiment of simulated data

(S1: 目标类样本; S2: 非目标类样本; T1: 目标类训练样本; T2: 非目标类训练样本; P10: 采用 10% 无标识样本的 RT-SVM 分类面; P20: 采用 20% 无标识样本的 RT-SVM 分类面; P100: 采用 100% 无标识样本的 RT-SVM 分类面。)

表 1 在模拟数据实验中各分类器的支撑向量数量

Table 1 The support vector number of different classifiers in the experiment of simulated

$\gamma$	WUS-SVM	RT-SVM	WUS-SVM	RT-SVM	WUS-SVM	RT-SVM	1-SVM	SVM
	10% 无标识样本		20% 无标识样本		100% 无标识样本			
1	57	22	101	31	402	86	3	30
10	39	14	137	19	256	25	7	15
100	95	73	160	103	482	155	23	86

## 5.2 遥感影像实验

该实验数据是 HYDICE 高光谱影像, 共有 220 个波段, 波谱范围为:  $0.4-2.40\mu\text{m}$ 。实验中去除了水蒸气吸收波段, 剩余 191 个波段用于分类。影像大小为 307 列 280 行, 覆盖区为华盛顿部分街区

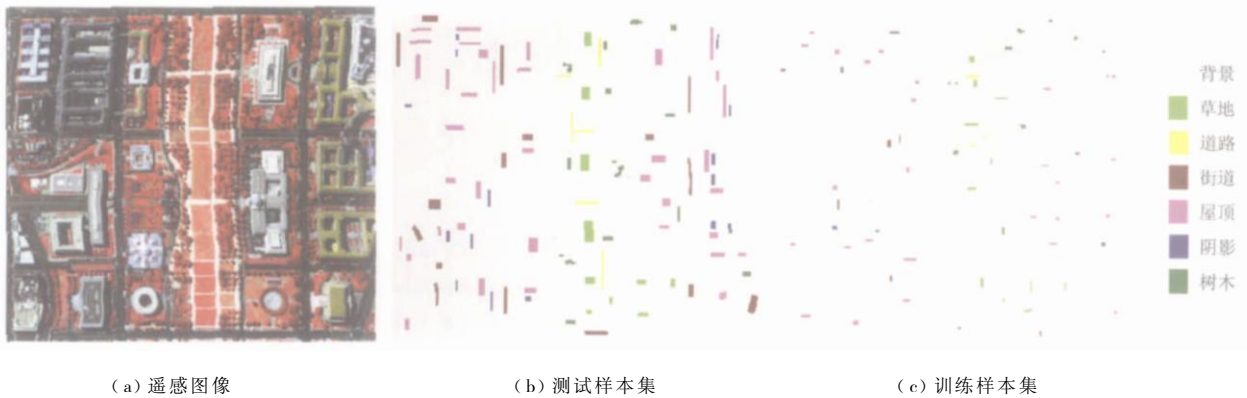


图 5 遥感影像以及测试样本和训练样本

Fig 5 Original HYDICE image with training and testing samples

表 2 训练样本和测试样本的选取

Table 2 Training and testing samples

类号	类名	训练样本数	测试样本数
1	草地	131	564
2	砾石路	36	288
3	街道	93	992
4	屋顶	292	1846
5	阴影	41	334
6	树木	116	402

本组实验采用了 RT-SVM 和 1-SVM, 两者的核参数  $\gamma$  采用了 3 个值: 0.5, 2.6 和 5.2。此外还采用了光谱角制图法 (SAM)<sup>[12]</sup>, 这是一种常用的高光谱影像分类算法。由于只有目标类训练样本, SAM 需要给定一个弧度阈值  $r$  当测试样本与训练样本间的波谱角小于  $r$  则认为是目标类, 否则为非目标类。因此这里 SAM 是一种绝对分类法。由于文章篇幅的限制, 本文选出两种代表性的类别: “树木”和“屋

顶”作为目标类。在该实验影像中“树木”的波谱特性比较单一, 在影像中所占比例较小, 而“屋顶”的波谱特性复杂, 且在影像中的分布较广。图 6, 图 7 和表 3 给出了 3 种算法采用不同的参数分别以“树木”和“屋顶”为目标类的分类结果。此时 RT-SVM 在整幅影像中随机选取 1% 的样本作为无标识样本 (图表中表示为  $u=1\%$ )。

从实验结果可以看出, SAM 的分类精度明显受阈值大小的影响。此外当  $r=0.2$  时, 在屋顶的分类实验中尽管出现大量误选, 但是其漏选率也依然高达 20.7%, 这同时说明 SAM 的分类能力是有限的。1-SVM 的分类精度取决于核参数  $\gamma$  的取值, 随着  $\gamma$  的增大其漏选率上升, 误选率下降。这在屋顶的分类结果中体现得非常明显。RT-SVM 对核参数  $\gamma$  取值敏感程度低于 1-SVM, 而且分类精度稳定且明显高于 1-SVM 和 SAM。但是 1-SVM 和 SAM 的训练速度和分类速度均快于 RT-SVM。

从实验结果可以看出, SAM 的分类精度明显受阈值大小的影响。此外当  $r=0.2$  时, 在屋顶的分类实验中尽管出现大量误选, 但是其漏选率也依然高达 20.7%, 这同时说明 SAM 的分类能力是有限的。1-SVM 的分类精度取决于核参数  $\gamma$  的取值, 随着  $\gamma$  的增大其漏选率上升, 误选率下降。这在屋顶的分类结果中体现得非常明显。RT-SVM 对核参数  $\gamma$  取值敏感程度低于 1-SVM, 而且分类精度稳定且明显高于 1-SVM 和 SAM。但是 1-SVM 和 SAM 的训练速度和分类速度均快于 RT-SVM。

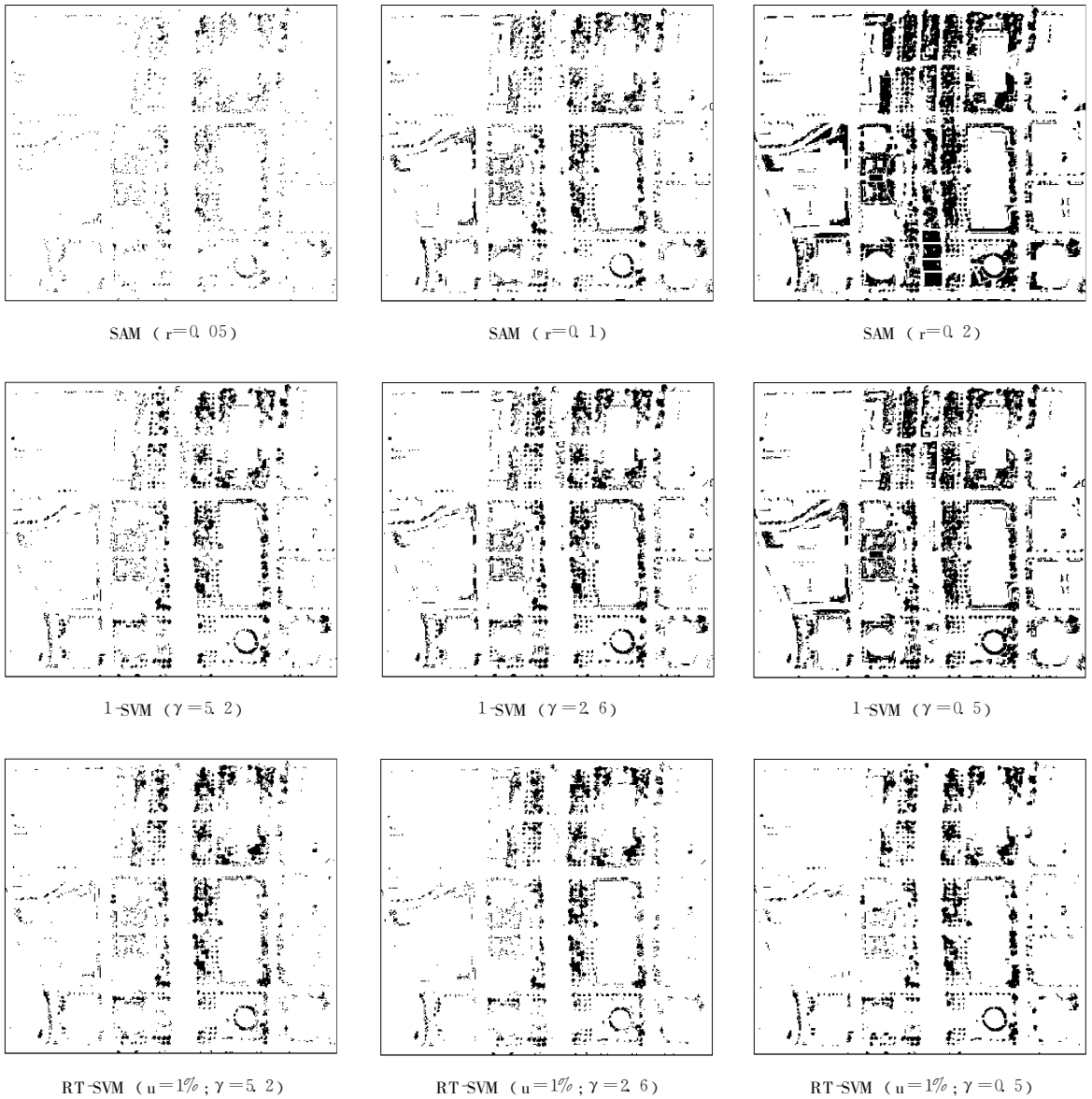


图 6 不同算法的分类结果(树木)

Fig 6 Classification results by different algorithms( tree)

表 3 不同算法的分类精度

Table 3 Classification accuracies of different algorithms

地类	精度	SAM			1-SVM			RT-SVM		
		r=0.05	r=0.1	r=0.2	gamma=5.2	gamma=2.6	gamma=0.5	u=1% gamma=5.2	u=1% gamma=2.6	u=1% gamma=0.5
树木 %	E <sub>o</sub>	62.2	23.6	4.8	19.8	9.7	2.7	10.0	9.2	5.7
	E <sub>c</sub>	0	0	12.2	0.2	0.3	3.3	0.2	0.2	0.1
	E <sub>a</sub>	5.6	2.1	11.5	1.7	1.2	3.2	1.1	1.0	0.6
屋顶 %	E <sub>o</sub>	90.1	63.7	20.7	52.4	34.3	10.3	15.7	17.8	11.4
	E <sub>c</sub>	0.3	1.2	18.2	8.9	16.0	35.4	5.9	5.7	5.4
	E <sub>a</sub>	37.0	26.6	19.2	26.6	23.5	25.1	10.6	10.6	7.8

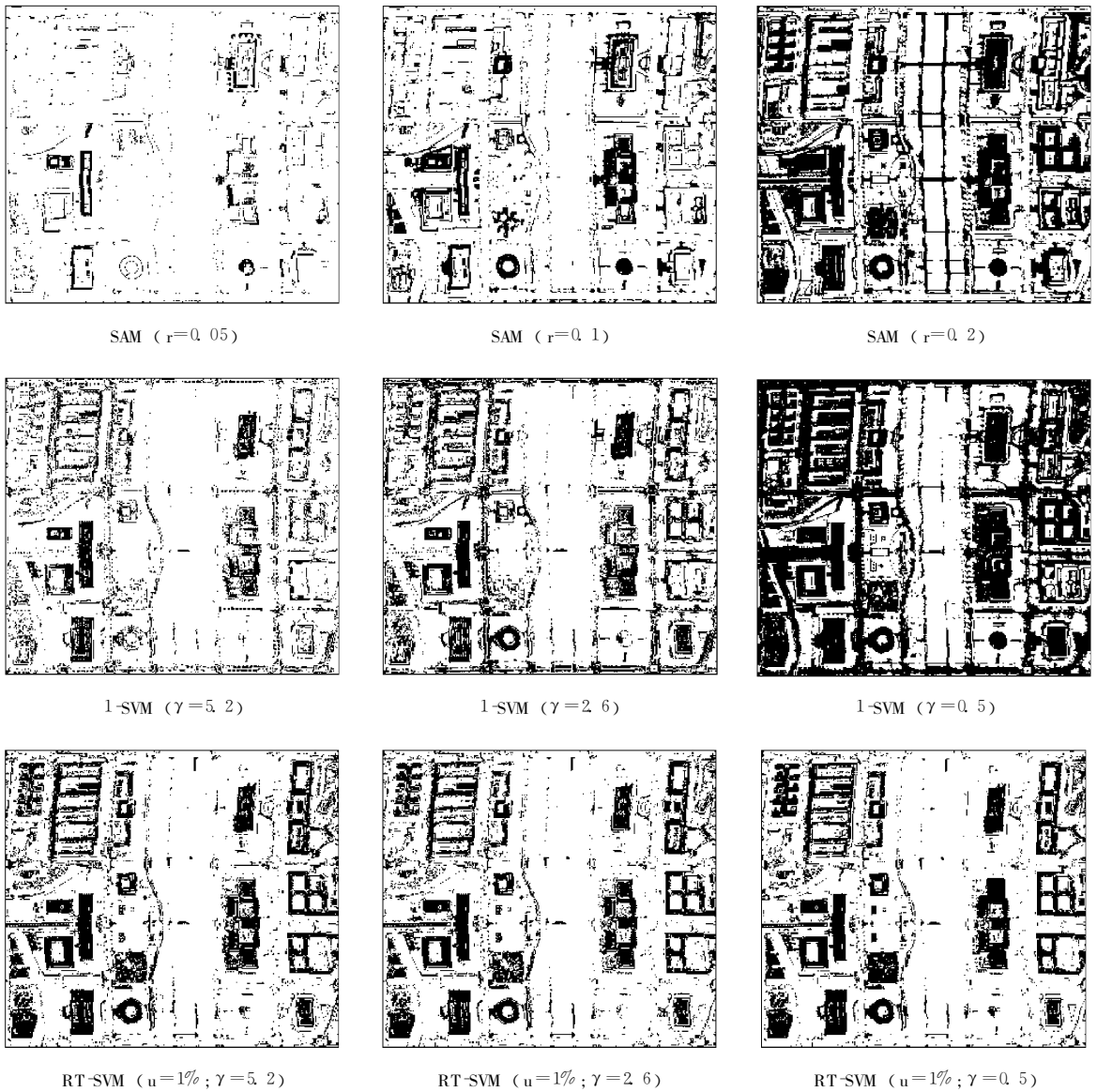


图 7 不同算法的分类结果(屋顶)

Fig 7 Classification results by different algorithms ( roof)

为了进一步分析无标识样本的数量多少对 RT-SVM的影响,实验中分别在影像中随机选取了 0.05%, 0.2%, 0.5%, 1%和 5%的样本作为无标识

样本。对“树木”和“屋顶”的两组试验的分类结果如图 8,图 9和表 4所示。

表 4 基于不同数量无标识样本的 RT-SVM ( $\gamma=0.5$ )的分类误差和支撑向量数  
Table 4 Classification accuracy and support vector number of RT-SVM ( $\gamma=0.5$ ) with different amount of unlabeled samples

无标识样本百分比 %	树 木					屋 顶				
	分类误差 %			支撑向量数		分类误差 %			支撑向量数	
	$E_o$	$E_c$	$E_a$	WUS-SVM	RT-SVM	$E_o$	$E_c$	$E_a$	WUS-SVM	RT-SVM
0.05	2.2	0.9	1.0	48	36	5.1	18.8	13.3	101	78
0.2	2.7	0.7	0.9	74	49	11.0	8.3	9.4	150	84
0.5	5.0	0.3	0.9	104	53	8.1	7.2	7.5	254	89
1	5.7	0.1	0.6	185	78	11.4	5.4	7.8	479	121
5	4.7	0.1	0.6	463	103	12.4	3.9	7.3	1561	350

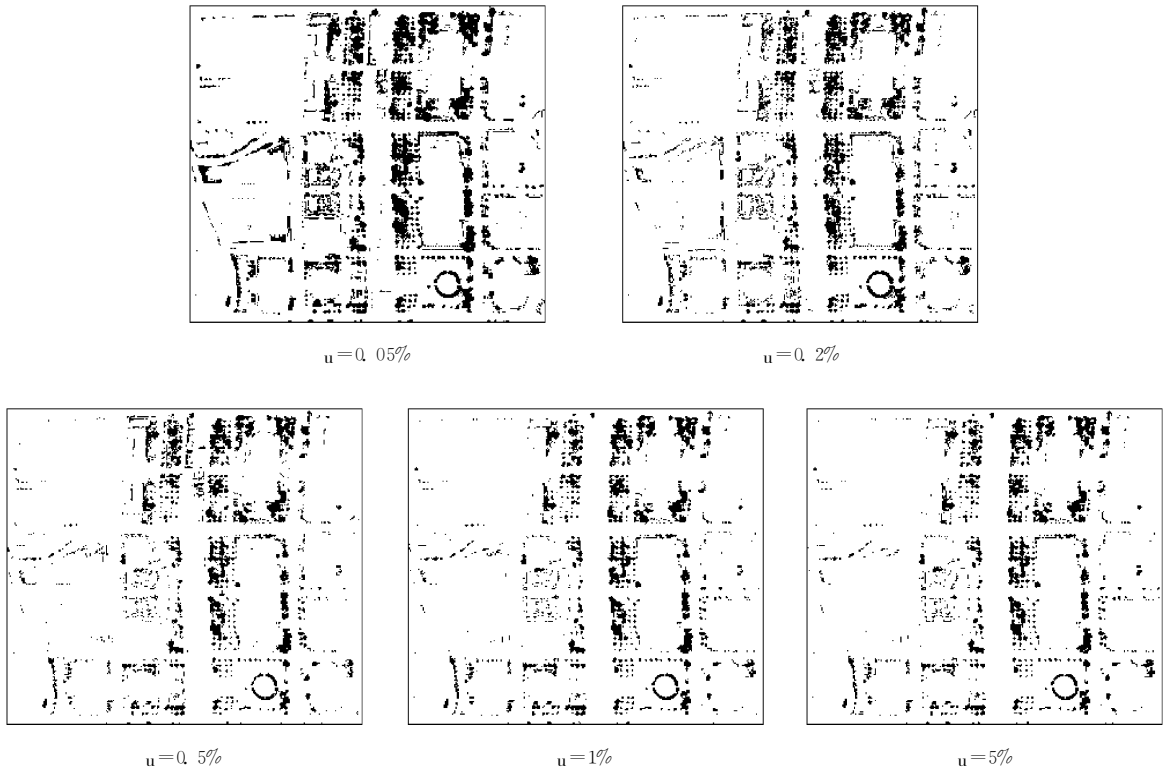


图 8 采用不同数量无标识样本的 RT-SVM ( $\gamma=0.5$ )对树木的提取结果

Fig 8 Classification results(tree) of RT-SVM ( $\gamma=0.5$ ) with different amount of unlabeled samples



图 9 采用不同数量无标识样本的 RT-SVM ( $\gamma=0.5$ )对屋顶的提取结果

Fig 9 Classification results(roof) of RT-SVM ( $\gamma=0.5$ ) with different amount of unlabeled samples

就精度而言,随着无标识样本数量的增加,RT-SVM的误选率逐渐下降,漏选率逐渐上升。当无标识样本过少,则误选率偏高;当无标识样本过多,则漏选率偏高。这是由于随着无标识样本采样率的提高,其中目标类样本的数量也不断增加,这迫使WUS-SVM在训练中越来越偏重于降低训练误差,导致分类面逐步收紧。不过在这个过程中无标识样本的惩罚系数  $P_{others}$  逐渐减小,在一定程度上起到了缓和作用,因此平均的分类精度依然比较稳定。

就速度而言,无标识训练样本的增加,将直接增加WUS-SVM中的支撑向量数,导致训练中计算量的增加。而且为了选取适当大小的  $P_{others}$  需要反复地进行WUS-SVM训练,因此增加无标识训练样本将明显降低训练速度。尽管RT-SVM中的支撑向量较WUS-SVM有明显减少,但是依然随着无标识训练样本的增加而增加(表4),因此增加无标识训练样本也将降低分类速度。

### 5.3 实验结论

通过模拟数据和遥感影像数据的实验表明,本文提出的RT-SVM能够以优于1-SVM和SAM的精度实现PSC,而且有效地克服了后两者对参数取值高度敏感的缺点。无标识样本的数量需要控制在一定范围内,过少可能增加误选率,过多则将影响训练和分类速度。值得指出的是,由于在WUS-SVM的训练阶段需要通过循环检测来自动选取参数,因此在分类器的训练阶段耗费的时间较长。但由于整个训练过程是自动化的,且不需要采集非目标类训练样本可以节省大量人力物力,因此该方法不失为一种方便有效的PSC方法。

## 6 结 论

本文提出了一种基于加权无标识支撑向量机(WUS-SVM),并在其基础提出了一种新的不完全监督分类方法。其内在思想在于通过选取无标识样本间接地利用非目标类的信息,将不完全分类问题转化为两类分类问题。在模拟数据和遥感影像的分类试验均达到了理想的分类精度。如果在WUS-SVM

的训练之前对训练样本做适当的滤筛选有可能提高该方法的训练和分类速度。这需要进一步的研究验证。值得指出的是本文提出的方法同样可以运用于网页搜索、基于内容的图像检索以及故障检测等问题。

### 参 考 文 献 (References)

- [ 1 ] Jeon B, Landgrebe D A. Partially Supervised Classification Using Weighted Unsupervised Clustering [ J ]. IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(2): 1073-1079.
- [ 2 ] Jeon B, Landgrebe D A. A New Supervised Absolute Classifier [ A ]. Proc of IEEE International Geoscience and Remote Sensing Symposium ( IGARSS ), 1990.
- [ 3 ] Vapnik V. The Nature of Statistical Learning Theory [ M ]. Springer NY, 1995.
- [ 4 ] Yu H, Han J, Chang K C. PEBL: Positive Example Based Learning for Web Page Classification Using SVM [ A ]. In Proc 8th Int Conf Knowledge Discovery and Data Mining [ C ]. Edmonton Canada, 2002.
- [ 5 ] Lee W S, Liu B. Learning from Positive and Unlabeled Examples Using Weighted Logistic Regression [ A ]. Proceedings of the Twentieth International Conference on Machine Learning [ C ]. Washington DC USA, 2003.
- [ 6 ] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the Support of A High-Dimensional Distribution [ J ]. Neural Computation, 2001, 13(7): 1443-1471.
- [ 7 ] Tax D. One-class Classification [ D ]. USA: Delft University of Technology, 2001.
- [ 8 ] Schölkopf B, Smola A. Learning with Kernels Support Vector Machines Regularization, Optimization and Beyond [ M ]. MIT Press Cambridge MA, 2002.
- [ 9 ] Chang C C, Lin C J. LIBSVM: A Library for Support Vector Machines [ P ]. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, 2004.
- [ 10 ] Zhang Q, Landgrebe D. Design of an Adaptive Classification Procedure for the Analysis of High-Dimensional Data with Limited Training Samples [ D ]. USA: Purdue University, 2001.
- [ 11 ] Dundar M, Landgrebe D. A Model-Based Mixture Supervised Classification Approach in Hyperspectral Data Analysis [ J ]. IEEE Transactions on Geoscience and Remote Sensing, 2002, 40(12): 2692-2699.
- [ 12 ] Knuse F, Lefkoff A, Boardman J, et al. The Spectral Image Processing System ( SIPS )—Interactive Visualization and Analysis of Imaging Spectrometer Data [ J ]. Remote Sensing of Environment, 1993, 44: 145-163.

## Partially Supervised Classification of Remotely Sensed Imagery Using Support Vector Machines

LIU Zhi-gang<sup>1,2</sup>, SHI Wen-zhong<sup>3</sup>, LI De-ren<sup>2</sup>, QIN Qian-qing<sup>2</sup>

(1. The State Key Laboratory of Remote Sensing Science and Beijing Key Laboratory for Environmental Remote Sensing and Digital Cities  
School of Geography and Remote Sensing Science, Beijing Normal University, Beijing 100875, China;

2. National Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Hubei Wuhan 430079, China;

3. Advanced Research Centre for Spatial Information Technology, Department of Land Surveying and Geo-Informatics  
The Hong Kong Polytechnic University, Hong Kong, China)

**Abstract** In many real world of remote sensing applications, one may be interested in identifying if the samples belong to one class from a remote sensing imagery. It is always expensive in terms of time and manpower to collect an exhaustive training sample set. Therefore, it is useful to design a classifier by which a user needs only to collect training samples of the objective class for identifying if a pixel belongs to the class of interest or not. This is the technique of partially supervised classification. In this paper, we present an algorithm called Weighted Unlabeled Sample Support Vector Machine (WUS-SVM), based on which a new partially supervised classification method is proposed. In this method, a certain amount of unlabeled samples is first randomly selected from the test set and labeled as other classes with different weights. Second, a primary classifier is defined by WUS-SVM, in which a hyperplane is constructed to separate training samples of objective class and other classes with low weighted error and large margin width. Third, the primary classifier is used to determine the class of unlabeled samples. Thus all training samples are labeled and again they are used as training samples of support vector machines to obtain a final classifier, which is used to classify other test samples. Experimental results with both simulated and real data sets show that the proposed method is very effective.

**Key words** partially supervised classification (PSC); weighed unlabeled sample support vector machine (WUS-SVM); support vector machine (SVM); remote sensing