

# 一种多光谱遥感图象的自适应最小距离分类方法

朱建华 刘政凯 俞能海

(中国科学技术大学电子工程与信息科学系, 合肥 230027)

**摘要** 该文提出了一种适用于多光谱遥感图象分类的自适应最小距离算法. 这一算法通过对样本集合的自适应精细划分来调整最小距离分类器的参数. 对 TM 遥感图象分类的实验结果表明, 该方法对 16 类数据进行有监督分类, 精度可达 92.9%, 适用于多类别遥感图象分类.

**关键词** 计算机图象处理 遥感 最小距离分类

中图分类号: TP751 文献标识码: A 文章编号: 1006-8961(2000)01-0021-04

## Remote Sensing Image Classification Using an Adaptive Min-Distance Algorithm

ZHU Jian-hua, LIU Zheng-kai, YU Neng-hai

(Department of Electronic Engineering and Information Science,  
University of Science and Technology of China, Hefei 230027)

**Abstract** This article presents an adaptive min-distance algorithm to classify multi-spectral remote-sensing images. This method approximates the distribution of the classes by dividing the sample sets, and adjusts the parameters of the min-distance classifier adaptively. Experiments with TM remote-sensing images demonstrate that this approach achieves an accuracy of 92.9% in the supervised classification of 16 classes. The experimental results verify the applicability of this approach in classifying of multicategory remote-sensing images.

**Keywords** Computer image processing, Remote-sensing, min-distance classification

## 0 引言

遥感技术是人类研究地球资源环境的重要技术手段. 随着资源环境信息的爆炸性增长, 地理信息系统(GIS)在土地资源管理、环境监测等方面发挥着越来越重要的作用. 当前, GIS正在与遥感(RS)、全球定位系统(GPS)紧密结合, 向3S一体化方面发展. 遥感分类是遥感地理信息系统中的关键技术之一, 快速、高精度的遥感图象分类算法是实现资源环境的动态监测、评价、预报的关键. 传统的统计模式分类方法有许多难以克服的困难: 运算量太大, 无法获得各类的先验概率和概率分布密度函数, 难以形成复杂的判别函数和分割界面等. 人工神经网络分

类方法(如BP网络<sup>[3]</sup>)具有良好的容错能力和自适应性, 对模式先验概率分布的要求较小, 因而具有比较好的性能. 但是神经网络方法也有一些缺点, 如局部极值点问题、训练收敛速度太慢、分类性能对各类差别较大等等.

最小距离分类<sup>[2]</sup>是一种最简单的有监督分类方法, 其思想为将待分类点归入到与其距离最小的一类. 受模式散布的影响, 这类方法一般精度不高. 针对简单最小距离分类的缺点, 我们提出了一种自适应的最小距离分类方法, 在训练过程中, 将各类样本集合自适应地分解为子集树, 定义待分类点到子集树的距离作为分类依据. 实验结果表明, 这一算法具有较高的分类正确率(对6波段16类数据分类精度达92.9%)和较快的分类速度和训练速度, 是一种简单高效的遥感分类算法.

# 1 自适应最小距离分类原理

最小距离分类通过定义待分类点到各类的距离, 将其归入距离最小的一类, 按照距离的不同定义, 可以有多种具体方式. 最简单最小距离分类器表达为: 设数据为  $M$  个波段,  $N$  个类别分别用标准样本  $W_1, W_2, \dots, W_N$  表示, 则根据最小分类原理, 待分类点  $P$  到一类的距离可定义为:

$$d_i(P) = \sqrt{\sum_{j=1}^M (p_j - w_{ij})^2}, i = 1, 2, \dots, N$$

设第  $i$  类训练样本集合  $\{x_{ik}, k = 1, 2, \dots, C_i\}$ , 标准样本一般选择为一类训练样本的中心:

$$w_{ij} = \frac{1}{C_i} \sum_{k=1}^{C_i} x_{ik}, i = 1, 2, \dots, N, j = 1, 2, \dots, M$$

分类准则为

$$d_i(P) < d_j(P), \forall j \neq i \Rightarrow P \in \tilde{\omega}$$

一般来说, 最小距离分类只能用于低维、类别少、样本数小、样本散布小的情况, 很难在遥感分类中应用, 通常作为研究分类方法时的一种参考方法.

我们知道,  $N$  维模式空间中的一个连续点集可

以用一组  $N$  维球体的并集来任意逼近. 图 1 是球体逼近在二维空间(平面)上的示意图. 分类实质上就是划分出各类间的边界. 针对样本散布的情况, 我们通过样本集合的分解得到样本分布的一个估计, 根据分解的结果求出待分类点到一类的距离并作为分类的依据. 而这个分解过程可以通过对样本的分类自适应地实现. 由此, 我们提出了一种自适应的最小距离分类方法.

自适应最小距离分类的基本原理表述为: 将每一类模式点集近似为一组球体的并集, 对待分类的一点, 判断其落在哪个球体中, 并赋予相应的类编号. 取得模式集合划分的所有球体中心半径等参数的过程, 称为分类器的训练. 鉴于样本的代表性, 这一过程可以通过样本集合的划分(球体逼近)来实现. 我们设计的自适应最小距离分类器的集合划分是一个由上到下, 逐步细化的过程, 即首先用半径较大的球尝试逼近样本点集, 并用所得到的分类器对样本进行分类实验. 如果分类精度达不到要求, 则将大球进一步细分为半径比较小的球. 根据对样本分类的精度要求自适应地控制样本集合的分解过程, 就可以使不同类的小球尽可能分离, 从而提高分类精度. 如图 2 所示.

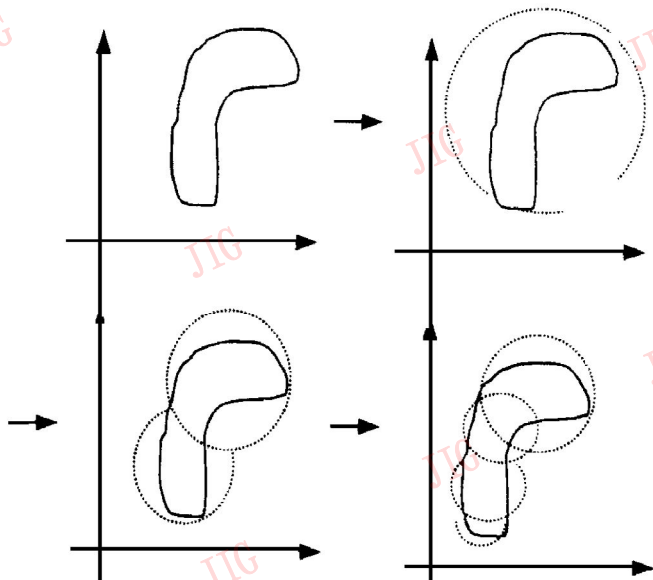


图 1 二维空间球体逼近示意图

我们使用的集合细分方法是 K-Means 算法, 每次分解为两个子集, 样本集合的细分是以二叉树方式进行的. 每一类生成一棵二叉树, 称为该类的子集树. 二叉树的一个节点对应一个球体, 记录每个球体的球心半径, 以及对应的样本子集. 球体的中心定义为该节点上样本子集的中心, 半径是该节点上的样本点到球心欧氏距离的最大值. 由一个节点的子集

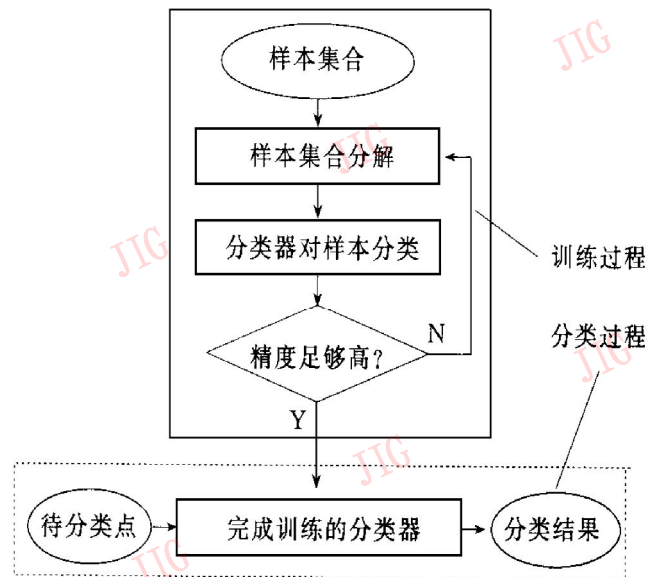


图 2 自适应最小距离分类算法流程图

细分得到的球体则表示为该节点的两个子节点.

自适应最小距离分类器对待分类点  $P$  的分类步骤为:

(1) 求出该点到各类对应的子集树的距离  $D$ . 待分类点  $P$  到子集树  $T$  的距离  $D(T, P)$  定义为:

① 如果  $P$  到  $T$  的根节点对应的球心欧氏距离  $d$  大于该节点球体的半径的两倍, 则忽略该节点细

分得到的所有小球,并令  $D = d$ ;

② 如果  $T$  的根节点已经是叶节点,则令  $D = d$ ;

③ 若①,②均不满足,则  $D$  递归定义为  $P$  到  $T$  的左右子树  $T_1, T_2$  的距离  $D_1, D_2$  的最小者,即:  $D_1 = D(T_1, P)$ ;  $D_2 = D(T_2, P)$ ;  $D = \min(D_1, D_2)$ .

(2) 根据最小距离原理,将距离  $D$  最小的一子集树的类号赋予待分类点.

比较可见,前面最简单的小距离分类中的距离定义相当于某个子集树只有根节点的情况,只用一个中心计算距离,因此受样本散布影响比较大.显然,我们定义的距离,比起最简单的距离定义更为精细,因而能够实现高精度分类.距离计算中使用了二叉树查找,使得只有一部分节点参与  $D$  的求取,因此即使细分得到的球体总数量很多,实际计算  $d$  的次数并不多.这就保证了在样本数量增大或者细分得到球体增多的情况下计算量不会增加很多.

分类器的训练过程是一个自适应的过程,即:使用当前的分类器依次对各个子集树上的各个叶节点上样本子集进行分类,如果分类精度不够高,说明与其他类的球体发生了交叠,这时使用 K-Means 算法将其分解为两个子节点(半径较小的球体)以避免交叠.自适应最小距离分类器的训练步骤为:

(1) 初始化各个类的子集树为只有一个根节点的情况,即根节点上的样本点集均为训练样本集中属于该类的子集.同时求出各个根节点上样本点集的中心和半径.

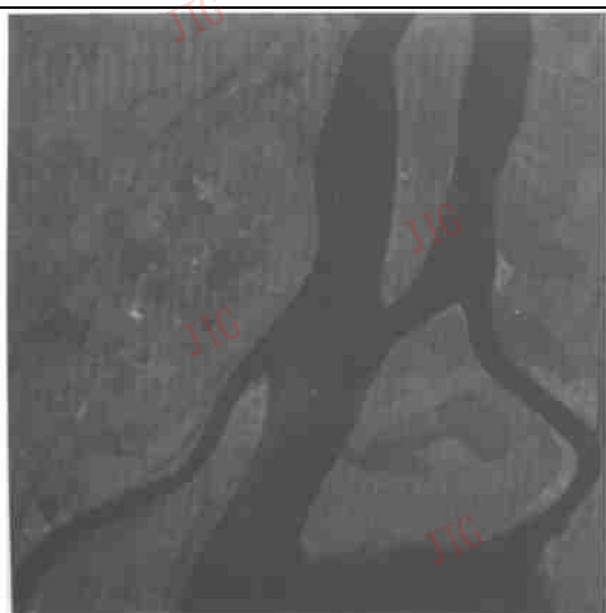
(2) 对各个子集树进行训练 对各个叶节点,如果按照前述分类算法对该节点上的样本点集进行分类的正确率低于给定的门限,则将该节点进行细分,即生成其左右子节点,将样本点集用 K-Means 分类算法分解为两个子集,分别作为两个子节点的样本点集,并求出左右子节点对应球体的中心和半径等参数.

(3) 重复(2)对各个子集树依次进行训练,直到没有一个子集树被更新.训练结束后,各个节点上的样本集合已经没有必要保留.

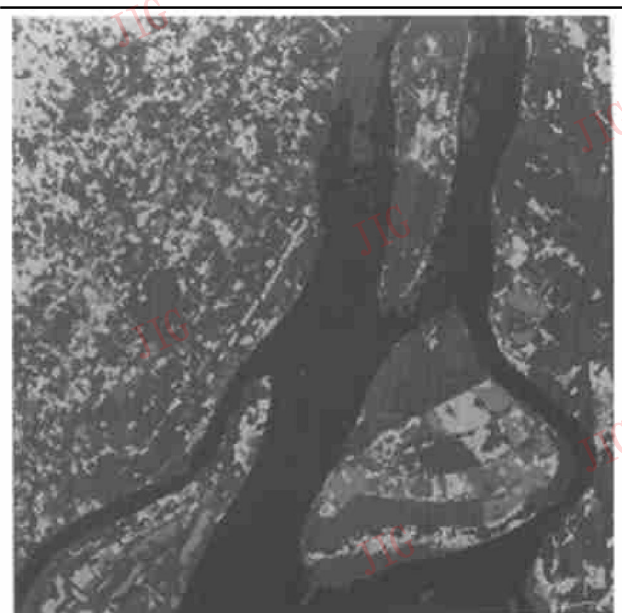
自适应最小距离分类用一组球体(一组训练样本子集)的中心来定义距离,较单个中心更准确.由于样本的集合分解是在对样本分类实验的反馈指导下自适应地进行的,因此这一算法能有效提高分类精度.

## 2 多波段遥感图象分类实验

我们实验所用的原始图象是葡萄牙里斯本地区泰吉河流域遥感图象.原始图象为  $256 \times 256$ (6 个波段),地面采样得到的样本为 16 类,共 8047 点.16 类样本中包含土地、水和当地若干典型植被.我们从 8047 点样本中随机选 3000 点作训练样本集,然后对这 8047 点采样样本进行分类,以便统计分类精度,得到的实验数据如表 1 所示.对全图进行分类的结果图象见图 3(b).



(a) 原始遥感图



(b) 分类影像

图 3 自适应最小距离分类前后图象

表1 几种有监督分类结果比较

分类算法	正确点数	正确率
自适应最小距离	747 8	92.9%
简单最小距离分类	326 0	40.5%
最大似然分类	637 6	79.2%

由表1可以看出,自适应最小距离分类取得了比较理想的精度,而简单的最小距离分类效果最差,最大似然分类是一种基于统计的分类方法,由于实际应用中,遥感图象的模式不一定满足正态分布,这类方法的精度往往有限。

### 3 结论和讨论

实验结果表明,本文提出的基于样本集合划分的自适应最小距离分类算法,是一种简单有效的有监督分类算法.它避免了传统统计分类对先验知识的需求,精度较高,适用于多类别遥感图象分类.由此可见,对训练样本做适当分解处理是提高有监督分类精度的有效方法。



**刘政凯** 1964年毕业于中国科学技术大学无线电遥控遥测专业,现为中国科学技术大学教授,博士生导师.主要从事遥感图象处理、人工神经网络及模式识别方面的研究.已发表论文60余篇,出版专著6本。



**朱建华** 1996年毕业于中国科学技术大学,获学士学位,现为中国科学技术大学硕士研究生.主要研究方向为遥感图象处理与模式识别。



**俞能海** 1992年毕业于清华大学,获工学学士学位,现为中国科学技术大学副教授,在职博士生.主要研究领域为图象处理、地理信息系统、多媒体技术等.已发表学术论文十余篇,获省部级科技进步三等奖两项。

### 参考文献

- 1 刘政凯. 微型计算机数字图象处理技术. 合肥:安徽科学技术出版社,1991.
- 2 沈清,汤霖. 模式识别导论. 长沙:国防科技大学出版社.
- 3 Houqiang Li, Zhengkai Liu. A new approach to pattern recognition of remote sensing Images using artificial neural network. IEEE, IG ARSS93.
- 4 李厚强,王宜主,刘政凯. 一种适用于多类别遥感图象分类的新方法——复合神经网络分类方法. 遥感学报, 1997, 11.
- 5 靳文戟,刘政凯. 多类别遥感图象的复合分类方法. 环境遥感, 1995, 10(4).