

文章编号: 1007-4619(2005)06-0708-09

基于分类回归树分析的遥感影像 土地利用 覆被分类研究

赵 萍^{1,2}, 傅云飞¹, 郑刘根¹, 冯学智³, B. Satyanarayana⁴

(1. 中国科学技术大学 地球与空间科学学院, 安徽 合肥 230026 2. 安徽师范大学 国土资源与旅游学院, 安徽 芜湖 241000
3. 南京大学 城市与资源学系, 江苏 南京 210093 4. Andhra 大学生物系, 印度 维沙卡 530003)

摘 要: 以专家知识和经验为基础, 综合影像光谱信息和其他辅助信息进行分类的基于知识的遥感影像解译方法, 是提高遥感影像分类精度, 实现自动解译的有效途径之一。然而, 知识的获取一直是其得以广泛应用的“瓶颈”问题。以江苏省江宁试验区土地利用 覆被分类为例, 利用分类回归树分析 (CART) 从训练样本数据集中发现分类规则, 集成遥感影像的光谱特征、纹理特征和空间分布特征进行分类实验, 并与传统的监督分类和逻辑通道分类方法进行比较。结果表明, 基于 CART 的分类方法的精度基本在 80% 以上, 与另两种方法相比, 有了较大的提高, 而且该算法复杂性低, 效率高。由此说明, 利用 CART 算法构建决策树获取的分类规则是合理的。它可以快速、有效地获取大量分类规则, 是促进基于知识的遥感影像分类方法在土地利用 覆被分类中广泛应用的一项有效手段。

关键词: 分类回归树分析; 遥感影像; 土地利用 覆被分类; 知识

中图分类号: TP751/TP79 **文献标识码:** A

1 引 言

土地利用 覆被变化 (LUCC) 研究是全球变化研究的核心内容^[1], 而土地利用 覆被数据的获取是其主要的基础工作。近年来, 飞速发展的遥感技术已成为土地利用 覆被调查的一项有效的技术手段。然而, 遥感影像分类技术的提高却远远落后于遥感传感器技术的发展, 不能满足实际应用的需要。

遥感影像分类一直是遥感技术领域研究的一项重要内容。随着计算机技术的发展, 经历了人工目视解译到计算机自动分类的过程。传统的计算机自动分类的方法主要是基于概率统计的监督分类和非监督分类方法。然而, 由于遥感影像本身的复杂性, “同物异谱, 异物同谱”现象的存在, 以及这种概率统计方法对光谱特征正态分布的先验要求, 这种纯粹地依赖影像光谱特征的分类方法在地表状况复杂的地区往往难以得到较高的精度。多年来, 国内外学者一直都在探求能够自动、高效地实现遥感影像

解译的方法, 研究思路大体分为两种: 一是研究新的分类算法, 如人工神经网络方法、模糊数学方法等; 二是利用多源数据, 将专家目视解译时用到的知识加入到计算机自动解译过程中进行综合分类。然而众多研究表明, 单纯地改善分类算法难以达到生产实用的要求, 将多源辅助数据与卫星数据结合, 发展多维信息复合的方法可以大大提高分类的精度, 是提高遥感应用的有效途径之一^[2]。

光谱数据与其他辅助数据综合分类的方法很多, 根据辅助数据与遥感数据集成的时间可分为三大类: (1) 分类前处理, 即辅助数据用于选择训练区或对数据预分层。(2) 分类中处理, 有两种方式: 一是逻辑通道法, 即在传统的统计分类方法中将辅助数据作为输入波段的方法, 二是利用辅助数据确定先验概率的方法。但是, 这种分类中处理的方法在辅助数据不具备某种统计特性时, 会带来不可预测的分类结果。(3) 分类后处理, 主要是利用辅助数据对分类后的误判像元进行修正^[3]。近年来, 随着人工智能技术的发展, 模仿特定领域内人类专家解

收稿日期: 2004-06-28 修订日期: 2004-10-18

基金项目: 中德合作“江宁土地利用与可持续发展研究”(SILUP)项目, 中国科学技术大学“面向 21 世纪教育振兴行动计划”。

作者简介: 赵萍 (1975—), 女, 安徽师范大学, 副教授, 中国科学技术大学博士后, 博士毕业于南京大学地图学与地理信息系统专业, 主要研究方向为遥感图像处理、土地利用 覆被变化研究, 已发表论文 8 篇。

决复杂问题的专家系统技术,以及在模拟人类神经系统的结构和功能的基础上建立起来的人工神经网络技术,为遥感从目视解译到自动化、高精度量化找到了突破口。人工神经网络方法主要是通过增加输入神经元的个数来集成各类辅助数据,类似于逻辑通道法,但是它对数据类型及分布函数没有限制,在集成各类数据进行分类方面具有更大的灵活性。然而人工神经网络方法,无法集成离散的类别数据以及描述性的地学知识和专家经验,而且合理有效的人工神经网络结构的建立相对困难,网络的训练时间会随着输入数据维数的增加而倍增,所形成的网络结构也难以理解,仍是一种数据驱动的分类模型。而利用专家系统技术进行分类,不仅可以综合影像的光谱信息和空间结构信息,而且可以综合关于地物分布的地学知识和专家经验,在处理启发性知识方面具有更大的灵活性,更加符合人类专家解译过程中的特点,可以在很大程度上提高分类的精度。由于它是一种知识驱动的分类方式,很多学者也称之为基于知识的分类方法^[4-12]。然而,传统的利用人类专家的经验分析获取知识的方法通常是一个漫长而反复的过程,在很大程度上阻碍了遥感影像分类的自动化、智能化和实用化。因此,寻找一种有效的方法自动地获取知识,是当前基于知识的分类方法所需要解决的问题。为了解决知识的自动获取问题,在人工智能、知识工程和机器学习领域中,很多学者坚持不懈地展开了研究。20世纪80年代,逐步发展起来的数据挖掘与知识发现技术(Data Mining and Knowledge Discovery DMKD)为数据的理解提供了一种新的智能化手段。数据挖掘与知识发现,简单地说就是从大量数据中挖掘或抽取出隐含的、有价值的模式或规律等知识的复杂过程^[13]。对空间数据实施数据挖掘被称为空间数据挖掘与知识发现 SDMKD。它可以发现普遍的几何知识、空间分布规律、空间关联规则、空间分类规则、空间特征规则、空间区分规则、空间演变规则等,这就为遥感影像解译知识的获取开辟了一条新的途径^[14]。空间数据挖掘的方法很多,其中决策树归纳法是数据挖掘中获取分类规则的主要方法之一,比较适合遥感影像分类,已有一些学者作了初步的探索,取得了较好的分类效果^[15-23]。比较成熟的决策树构建方法有著名的 Quinlan 提出的 D3 C4.5 C5.0 系列,分类回归树分析(Classification and Regression Tree CART), SLIQ, SPRINT 和 CHAID 等^[24-28]。各种算法的构建过程都是分为决策树的生长和修剪两步进

行的,差异主要是决策树生长过程中分枝准则的确立方法和修剪技术^[24]。当前在遥感领域中使用较多的是 Quinlan 系列,该系列算法采用基于信息熵的方法构建分枝规则,然而采用的是预修剪技术,需要使用者对数据分布有较清楚的把握,并且常常需要反复地调整设置参数。本文将讨论 CART 算法在遥感影像土地利用 覆被分类中的应用。

2 分类回归树分析

CART 分析是 Breiman 于 1984 年提出的一种决策树构建算法,并不断进行了改进。其基本原理是通过由测试变量和目标变量构成的训练数据集的循环二分形成二叉树形式的决策树结构。该算法既可以用于分类,也可以用于连续变量的预测。当目标变量为离散的分类类别值时称为分类树;当目标变量为连续值时称为回归树。在土地利用 覆被分类中,目标变量是土地利用 覆被的类型值,测试变量为所利用的分类特征。该算法具有以下优点:结构清晰,易于理解;实现简单,运行速度快,准确性高;可以有效地处理大量数据和高维数据;可以处理非线性关系;对输入数据没有任何统计分布要求;输入数据可以是连续变量也可以是离散值;包容数据的缺失和错误;可以给出测试变量的重要性^[29-30]。

CART 分析在决策树生长过程中,采用经济学领域中的基尼(Gini)系数作为选择最佳测试变量和分割阈值的准则,基本过程如图 1。基尼系数的数学定义如下:

$$GiniIndex = 1 - \sum_{j=1}^J p^2(j|h)$$

其中, $P(j|h) = \frac{n_j(h)}{n(h)}$, $\sum_{j=1}^J P(j|h) = 1$ 。

式中, $p(j|h)$ 是从训练样本集中随机抽取一个样本,当某一测试变量值为 h 时属于第 j 类的概率, $n_j(h)$ 为训练样本中该测试变量值为 h 时属于第 j 类的样本个数, $n(h)$ 为训练样本中该测试变量值为 h 的样本个数, j 为类别个数。

按照上述过程生成的完整决策树往往会出现“过度拟和”的现象,这是因为完整的决策树结构对训练样本特征的描述“过于精确”,包含了噪声信息,失去了一般代表性而无法对新数据进行准确分类,因此有必要对树的结构进行修剪。CART 算法采用交叉验证的方法进行修剪,将样本数据分为训练数据和检测数据两部分,通常分为十等份,每次以

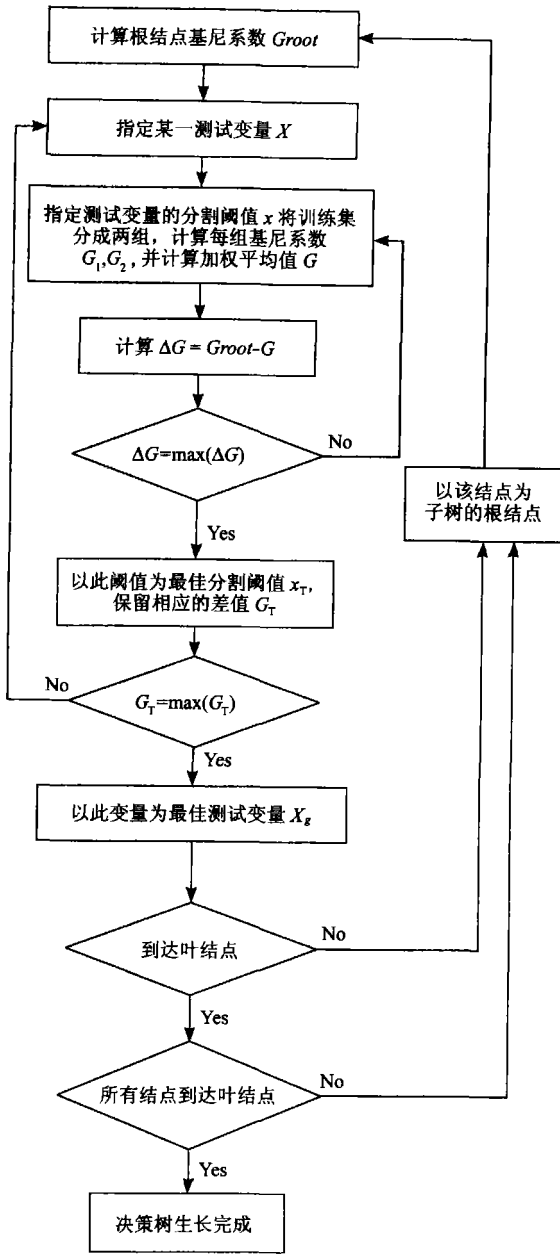


图 1 CART 算法决策树生长的基本过程

Fig 1 Growth of decision tree in CART

其中的九份作为训练数据, 一份作为检验数据, 如此循环交替进行验证。验证过程中引入一个“可调错误率”的概念, 即对某个树枝的所有叶节点增加一个惩罚因子, 如果该树枝仍然能够保持低错误率, 则说明它是强者, 予以保留; 否则它是弱者, 给予剪除。最终的分析结果是一棵兼顾复杂度和错误率的最优二叉树, 一系列二分点定义的每条途径都对应了一个最可能归属类别的判断条件。因此, 这棵树可以看作一系列可以用来对未知值进行分类的规则^[24 29 30]。

3 试验区概况和试验数据

江宁位于长江沿江开发带, 距离南京市 7km, 地理位置在 $31^{\circ}30' - 32^{\circ}00'N$ 与 $118^{\circ}30' - 119^{\circ}25'E$ 之间, 土地总面积 1567km^2 , 全区现辖 23 个乡镇。试验区位于江宁中部, 海拔 10—206m, 东部秦淮河谷水源充足, 以大面积的灌溉水田为主; 西部以黄土丘陵为主, 主要种植旱作; 缓岗与河谷平原的交界处, 水田和旱作交错分布; 边缘的丘陵山地缓坡处为人工林、经济林和一些低矮的灌丛, 陡坡处为针阔混交林所覆盖。整个试验区内各种地物类型交错分布, 地块较为细碎, 难以解译。

所采用的数据为 1999 01-21 的 SPOT XI 影像 (波长分别为 $0.50 - 0.59$ $0.61 - 0.68$ $0.79 - 0.89$ $1.58 - 1.75\mu\text{m}$, 空间分辨率为 20m), 1998 08 10 的 SPOT Pan 影像 (空间分辨率为 10m), 1999 年 1:10 万的土地利用图, 1974 年 1:5 万的地形图。

4 方 法

4.1 数据的预处理

(1)以 1:5 万地形图为基准, 分别选取道路、河流的交叉点等作为控制点, 采用二次多项式变换和最邻近像素重采样的方法对原始影像进行几何纠正和精配准, 误差控制在 0.5 个像素以内。纠正后 SPOT XI 像素大小为 22m, 从中选取 512×512 像素的试验区, SPOT Pan 为 10m。(2)数字化 1:10 万的土地利用图以及 1:5 万地形图上的等高线, 等高距取 10m, 利用其生成试验区 DEM (像素大小与 SPOT XI 影像一致), 并由此派生坡度、坡向图。试验区 SPOT XI (RGB431) 合成影像与 DEM 的叠加如图 2。

4.2 选取训练样本

训练样本的选取是学习的关键步骤, 直接关系到所获取规则的质量。根据该影像的解译能力并结合实际调查, 将试验区的土地利用 覆被类型分为: 水体、居民地、道路、荒草地、水田和林地 (包括园地) 六大类。为了使训练样本能够反映出各种土地利用 覆被类型在空间上的分布特征, 本文采用按照空间坐标随机分层采样的方法。在 10m 的 SPOT 卫星全色影像和 1:5 万地形图参考下, 对试验区进行采样, 阴影也作为一类, 共选取了 3340 个样本。

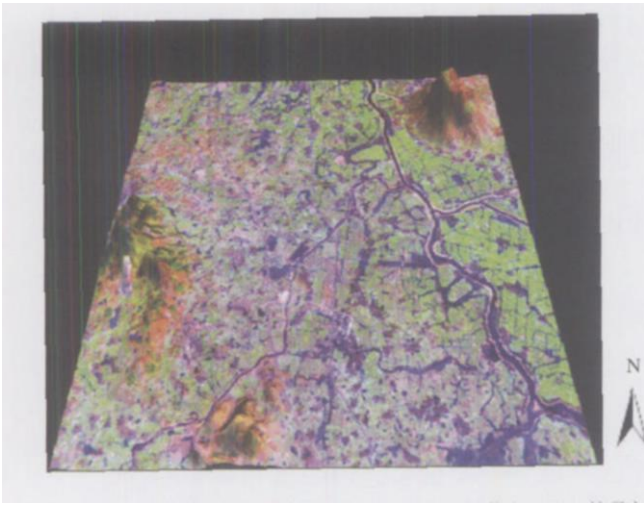


图 2 江宁试验区 SPOT X1 (RGB 7 4 3) 合成影像与 DEM 的叠加
Fig 2 SPOT X1 (RGB 4 3 1) overlaid by DEM in the study area of Jiangning

4.3 确定测试变量

光谱响应特征是多光谱遥感影像地物识别最直接也是最重要的解译元素。地表的各类地物由于物质组成和结构不同而具有独特的波谱反射和辐射特性,在图像上反映为各类地物在各波段上灰度值的差异。

纹理也是遥感影像的重要信息,它不仅反映了影像的灰度统计信息,而且反映了地物本身的结构特征和地物空间排列的关系,是进行目视判读的重要标志之一^[31]。许多研究表明,原始影像光谱信息加上纹理信息对提高影像的解译精度起着较好的作用^[32-36]。在遥感影像中,纹理大多为随机型纹理,服从统计分布,常用基于灰度共生矩阵的方法加以描述。Maceau 通过大量试验表明,基于灰度共生矩阵的纹理描述方法可以在很大程度上改进土地利用 覆被分类的精度^[37]。所谓灰度共生矩阵是由影像灰度级之间二阶联合条件概率密度所构成的矩阵,反映了影像中任意两点间灰度的空间相关性。在此基础上定义了 8 个纹理特征统计值:均值、方差、逆差矩、对比度、非相似度、熵、角二阶矩、相关性,它们从多个侧面描述了影像的纹理特征^[38]。利用灰度共生矩阵进行纹理分析涉及到 3 个重要参数:移动窗口的大小、移动步长和移动方向。合适的窗口大小对于纹理分析尤为重要,较小的窗口虽然能够保证窗口内像素的同质性,却无法准确有效地描述各类别像素排列的分布规律性;而较大的窗口虽然能够保证同类地物的纹理统计具有代表性,但又可能包含多余的异类地物像素的信息,使介于不

同种地物间的像素模糊。一般说来,窗口的大小取决于原始影像纹理结构本身,较小的窗口表征细微的纹理特征,较大的窗口表征粗糙的纹理特征。移动步长的选择也取决于影像中纹理的粗细度,短步长比较适于细纹理,长步长比较适于粗纹理, Treitz 等研究指出步长取 1 对于不同的纹理来说都比较有效。移动方向通常取 0° , 45° , 90° , 135° 4 个主要的方向或这 4 个方向的平均值^[35]。然而, Franklin 和 Pebble 研究认为单一方向的纹理特征比 4 个方向的平均效果要更好^[39]。由于高分辨率卫星影像具有更为丰富的纹理信息,因此本文选用时间相近的空间分辨率为 10m 的 SPOT 卫星的全色影像进行基于灰度共生矩阵的纹理特征分析。通过对不同窗口大小和步长的分析和比较,选取了 5×5 的窗口,大小为 1 的步长。同时发现,对试验区地物来说,在相关性纹理统计值上都没有可分性。

此外,地物在空间上的分布受地域自然条件的控制和人为因素的干预,往往存在某种地域分异规律。本试验区海拔较低,没有明显的垂直地带性,但是由于受地形的影响,地物分布存在有一定的规律。如水田主要分布在中部的平原、沟谷地区;林地和园地主要分布在丘陵山地;阴影一般在山的阴面;水体和居民地一般不会分布在坡度较陡的地方等。因此,本文将地理坐标和地形因子也作为预测变量。

综上所述,本研究共采用以下 16 个测试变量:地理坐标 (X, Y), SPOT X 影像的 4 个波段灰度值 (B1 B2 B3 B4), 7 个纹理特征值 (Mean Variance Homogeneity Contrast Dissimilarity Entropy Second Moment (SM)) 和 3 个地形因子 (DEM, Slope Aspect)。其中测试变量坡向 Aspect 为离散变量 (取值为 0 1, ..., 8), 其余测试变量均为连续变量。目标变量为:水体、居民地、道路、荒草地、水田、林地和阴影。

4.4 基于 CART 分析的分类

利用 CART 分析对由上述测试变量和目标变量构成的 3340 个样本进行学习,得到的决策树结构共 48 个叶结点,学习精度为 93.8% (决策树结构图略),验证精度为 91%,训练时间为 13.55s。该决策树结构可以很方便地表示成 If ~ Then 形式的产生式规则,从树的根结点到每一个叶结点的测试路径对应一条规则,共 48 条规则。例如:

If (B1 > 98.5 && B4 <= 80.5 && Mean <= 129.5) Then (class = 水体) CF = 0.962

If (Mean > 115.5 && Dissimilarity > 37.5 && B1 >

98.5 && B1 <= 114.5 && B4 > 80.5 && B4 <= 136.5 && Homogeneity > 5.5) Then(class = 居民地)CF=0.943

If(B4 > 80.5 && Mean > 115.5 && Dissimilarity <= 37.5 && B1 > 98.5 && B1 <= 110.5 && DEM <= 17 && B3 <= 91.5 && Variance > 13) Then(class = 道路)CF=0.654

If(B1 > 98.5 && Mean <= 115.5 && B4 > 80.5 && B4 <= 123.5 && B3 > 102.5 && Slope <= 3.5) Then(class = 水田)CF=0.966

.....

其中, CF 的值域为 [0, 1], 表明该规则的置信水平。

当该值为 0 时, 则排除当前像元为结论所给出类别的可能性; 而当取值为 1 时, 则已有可信度值不变。

利用上述规则采用以下简单的匹配策略进行演绎推理, 得到的结果如图 3(a):

(1) 当只满足其中的某一条规则时, 则取规则的输出类别为分类类别;

(2) 当同时满足多条规则, 则取置信度最大的规则的输出类别为分类类别;

(3) 当所有规则都不满足时, 则为未分类别。

为了与一般的分类方法相比, 分别进行了逻辑通道法(辅助数据以附加波段的形式加入进行监督分类)和监督分类试验, 如图 3(b), (c)。

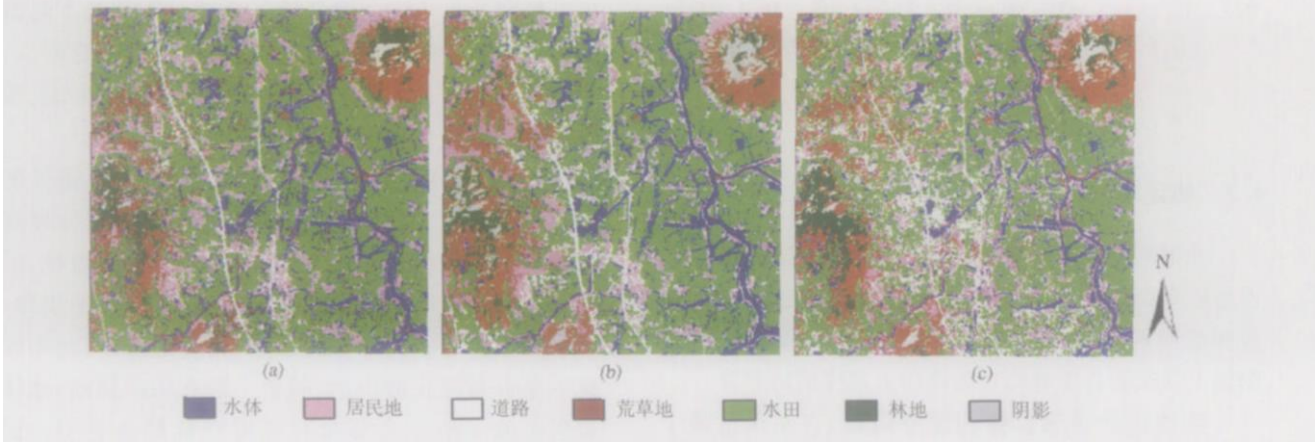


图 3 江宁试验区土地利用 覆被分类结果

(a) 基于 CART 的分类; (b) 逻辑通道法; (c) 监督分类

Fig 3 Results of land use /cover classification in the study area of Jiangning
(a) CART based classification; (b) logic channel classification; (c) supervised classification

4.5 精度评价

精度评价是遥感数据分类过程中一项不可缺少的工作。通过精度分析, 分类者能确定分类模式的有效性, 改进分类模式, 提高分类精度; 使用者能根据分类结果的精度, 正确、有效地获取分类结果中的信息^[40]。基于混淆矩阵的方法是目下普遍推荐的遥感影像分类精度的评价方法。本文参考 10m 的 SPOT 全色影像、1:50000 地形图和 1:100000 土地利用图, 并结合实地调查, 随机地选取了 1024 个检测样本, 通过目视判读建立混淆矩阵, 计算相关指标进行评价, 基于 CART 的分类精度评价如表 1, 不同方法的精度比较如表 2 和图 4。

5 结论与讨论

(1) 从表 1 可知, 除道路和荒草地以外, 所有地

物的使用精度和生产精度都在 80% 以上。使用精度与生产精度分别从两个方面衡量每一类别的分类精度, 前者对应的是错分误差, 后者对应的是漏分误差。由表 1 和图 3(a) 可以看出, 居民地和荒草地容易被错分为道路, 荒草地容易被错分为林地、道路和居民地。从图 4 可以看出, 除林地外基于 CART 的分类方法所获取的各类地物的使用精度都比其他方法要高, 除荒草地以外生产精度也都比其他方法要高。

(2) 从表 2 可知, 基于 CART 的分类方法总精度和 Kappa 系数均最高, 总精度为 87.79%, 比监督分类高 8.78%, 比逻辑通道法分类高 6.83%; Kappa 系数为 0.8489 比监督分类高 0.1073 比逻辑通道法高 0.0853。逻辑通道法对分类精度的改善并不大, 这是因为地形辅助因子往往不能满足一定的统计分布要求, 使用监督分类方法进行分类时, 通常会造造成不可预测的分类结果。

表 1 江宁试验区基于 CART 的分类精度评价结果

Table 1 Accuracy assessment on CART based classification in the study area of Jiangning

类型	水体	居民地	道路	荒草地	水田	林地	阴影	总和	使用精度 %
未分类	0	0	0	0	1	0	0	1	—
水体	111	0	0	6	3	0	0	120	92.5
居民地	1	141	3	8	11	0	0	164	85.98
道路	1	16	59	16	4	0	0	96	61.46
荒草地	1	3	0	184	13	3	0	204	90.20
水田	2	0	1	8	288	2	0	301	95.68
林地	2	0	0	11	4	74	1	92	80.43
阴影	0	0	0	4	0	0	42	46	91.3
总和	118	160	63	237	324	79	43	1024	—
生产精度 %	94.07	88.13	93.65	77.64	88.89	93.67	97.67	—	—

总精度 = 87.79% Kappa 系数 = 0.8489

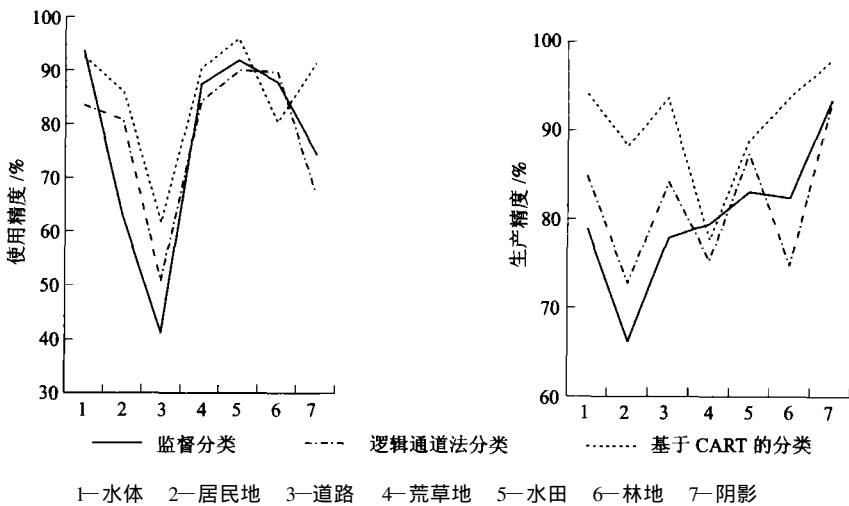


图 4 3种方法对江宁试验区各类地物分类的精度比较

Fig 4 Comparison of classification accuracy for all types of objects in the study area of Jiangning

表 2 江宁试验区 3种方法分类精度的比较

Table 2 Comparison of classification accuracy for three methods in the study area of Jiangning

分类方法	样本数	总精度 %	Kappa 系数
监督分类	1024	79.01	0.7416
逻辑通道法	1024	80.96	0.7636
基于 CART 的分类	1024	87.79	0.8489

(3)从图 3和图 4可以看出, 逻辑通道法和基于 CART 的分类方法都较明显地改善了道路和居民地的提取效果, 这是因为加入了高分辨率的全色卫

星影像的纹理特征, 从而融合了高分辨率影像的几何信息, 辅助多光谱影像的分类, 很大程度上改善了了几何结构信息丰富的地物类型的提取精度。

(4)尽管由于该地区土地利用 覆被类型复杂, 而且缺乏同一时期的实际验证资料, 受目视判读本身误差的局限, 分类精度的分析可能存在一定的偏差。但总体说来, 基于 CART 的方法的分类精度都比传统的监督分类和逻辑通道法有了较大的提高。而且该方法使用方便, 复杂度低, 计算简单, 运行速度快, 本例中学习时间仅 13.55s。由此说明, 利用分类回归树算法构建决策树获取分类规则是合理的。它可以快速、有效地获取大量的分类规则, 甚至

发现人类专家所无法分析出的知识,是促进基于知识的遥感影像分类方法在土地利用 覆被分类中广泛应用的一项有效手段。但是在应用该算法时应注意以下问题:在测试变量的选择中只考虑了单个变量,然而实际过程中类别的划分常与多个属性相关,可通过事先对测试变量加以组合的方法改进;当测试变量为类型值且取值很多、分布极为分散时,决策树结构会过于茂盛,改进的方法是将样本量较少的类合并;分类规则的建立仍很大程度上取决于样本的质量,样本的选取不仅要具有代表性,而且要尽量均匀,否则样本数多的类别属性将占优。

(5)本文仅探讨了分类回归树分析在遥感影像分类中的应用,进一步还可以研究其他空间数据挖掘与知识发现方法。如研究基于地物目标的空间数据挖掘方法,充分利用 GIS 辅助数据,发掘出更多的知识辅助分类,如地物的几何知识、地物间邻近关系、空间演变规则等。

参 考 文 献 (References)

- [1] Li X B. A Review of the International Researches on Land Use / Land Cover Change[J]. *Acta Geographica Sinica* 1996 51(6): 553— 557. [李秀彬. 全球环境变化研究的核心领域——土地利用 土地覆被变化的国际研究动向[J]. 地理学报, 1996 51(6): 553— 557]
- [2] Li S H. An Efficient Way to Improve the Classification by Using Remote Sensing Data [J]. *Remote Sensing for Land & Resources* 1995 (4): 1— 4. [李四海. 提高遥感数据分类应用性的有效途径[J]. 国土资源遥感, 1995 (4): 1— 4]
- [3] Hutchinson Charles F. Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement [J]. *Photogrammetric Engineering & Remote Sensing* 1982 48(1): 123— 130.
- [4] Srinivasan A, Richards J A. Knowledge based Techniques for Multi Source Classification[J]. *International Journal of Remote Sensing* 1990 11(3): 505— 525
- [5] Bolstad P V, Lillesand T M. Rule based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data [J]. *Photogrammetric Engineering & Remote Sensing* 1992 58(7): 965— 971
- [6] Richards J A, Jia X P. Remote Sensing Digital Image Analysis [M]. New York: Springer 1993.
- [7] Wilkinson G G. A Review of Current Issues in the Integration of GIS and Remote Sensing [J]. *International Journal of Geographical Information Systems* 1996 10(1): 85— 101
- [8] Dobson M C. Knowledge Based Land-Cover Classification Using ERS 1/JERS 1 SAR Composites[J]. *IEEE Transaction on Geoscience & Remote Sensing* 1996 34(1): 83— 99
- [9] Yang C J. Thematic Information Extraction from the Remote Sensing Images on the Basis of Knowledge Discovered[D]. Beijing: Institute of Geographical Sciences and Natural Resources Research CAS 1998. [杨存建. 基于知识发现的遥感影像专题信息提取[D]. 北京: 中国科学院地理研究所, 1998]
- [10] Qin Q M. Problems and Solutions of Automatic Interpretation of Remote Sensing Images[J]. *Science of Surveying and Mapping* 2000 25(2): 21— 24. [秦其明. 遥感图像自动解译面临的问题与解决的途径[J]. 测绘科学, 2000 25(2): 21— 24.]
- [11] Luo J C. Remote sensing Intelligent Geointerpretation Model and Its Geocognition Issue[J]. *Progress in Geography* 2000 19(4): 289— 296. [骆剑承. 遥感影像智能图解及其地学认知问题探索[J]. 地理科学进展, 2000 19(4): 289— 296]
- [12] Li S Ding S Y, Xu S M. The Comparative Study of Remote Sensing Image Classification[J]. *Journal of Henan University (Natural Science)*, 2002 32(2): 70— 73 [李爽, 丁圣彦, 许叔明. 遥感影像分类方法比较研究[J]. 河南大学学报(自然科学版), 2002 32(2): 70— 73.]
- [13] Zhu M. Data Mining[M]. Hefei: Anhui P. R. China University of Science and Technology of China Press 2002. [朱明. 数据挖掘[M]. 安徽合肥: 中国科学技术大学出版社, 2002.]
- [14] Di K C. Spatial Data Mining and Knowledge Discovery [M]. Wuhan: Wuhan University Press 2001. [邱凯昌. 空间数据挖掘与知识发现[M]. 武汉: 武汉大学出版社, 2001.]
- [15] Huang X Q, Jensen J R. A Machine Learning Approach to Automated Knowledge Based Building for Remote Sensing Image Analysis with GIS Data[J]. *Photogrammetric Engineering & Remote Sensing* 1997 63(10): 1185— 1194
- [16] Friedl M A, Brodley C E. Decision Tree Classification of Land Cover from Remotely Sensed Data [J]. *Remote Sensing Environment* 1997 61: 399— 409
- [17] DeFries R, Hansen M, Townshend J R G, et al. Global Land Cover Classifications at 8 km Spatial Resolution: The Use of Training Data Derived from Landsat Images in Decision Tree Classification[J]. *International Journal of Remote Sensing* 1998 19(16): 3141— 3168.
- [18] Simard M, Sansan S S, DeGrandi G, et al. The Use of Decision Tree and Multiscale Texture for Classification of JERS 1 SAR Data Over Tropical Forests[J]. *IEEE Transaction on Geoscience & Remote Sensing* 2000 38(5): 2310— 2321
- [19] Lawrence R L, Andrea W. Rule based Classification Systems Using Classification and Regression Tree (CART) Analysis[J]. *Photogrammetric Engineering & Remote Sensing* 2001 67(10): 1137— 1142
- [20] You D A, Jiang D H, Yu X C. Bayes Classification Assisted by GIS of Remote Sensing Images[J]. *Journal of Institute of Surveying and Mapping* 2001 18(2): 113— 117 [游代安, 蒋定华, 余旭初. GIS 辅助下的 Bayes 法遥感影像分类[J]. 测绘学院学报, 2001 18(2): 113— 117]
- [21] Meier D K, Friedl M A. Using Prior Probabilities in Decision Tree Classification of Remotely Sensed Data[J]. *Remote Sensing of Environment* 2002 81: 253— 261.
- [22] Li S Ding S Y. The Decision Tree Classification and Its Application Research in Land Cover [J]. *Remote Sensing*

- Technology and Application* 2002 17(1): 6—11. [李爽, 丁圣彦. 决策树分类法及其在土地覆被分类中的应用 [J]. 遥感技术与应用, 2002 17(1): 6—11.]
- [23] Pal M, Mather P M. An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification [J]. *Remote Sensing of Environment* 2003 86(4): 554—565
- [24] Center of Data Mining at the Statistics Department in Renmin University of China. Decision Tree Techniques and Its Application among the Data Mining [J]. *Statistics & Information Tribune* 2002 17(2): 4—10 [中国人民大学统计学系数据挖掘中心. 数据挖掘中的决策树技术及其应用 [J]. 统计与信息论坛, 2002 17(2): 4—10.]
- [25] Tan X, Wang L Z, Zhuo M. Researches on Discovery of Classification Rules based on Decision Tree [J]. *Journal of Yunnan University(Natural Science)*, 2000 22(6): 415—419. [谭循, 王丽珍, 卓明. 利用决策树发掘分类规则的算法研究 [J]. 云南大学学报(自然科学版), 2000 22(6): 415—419.]
- [26] Liu H Y, Chen J, Chen G Q. Review of Classification Algorithms for Data Mining [J]. *Journal of Tsinghua University (Science & Technology)*, 2002 42(6): 727—730. [刘红岩, 陈剑, 陈国青. 数据挖掘中的数据分类算法综述 [J]. 清华大学学报(自然科学版), 2002 42(6): 727—730.]
- [27] Zhou X Y, Li S Z, Qi X F *et al*. An Overview of Data Mining Techniques [J]. *MiniMicro Systems* 2002 23(3): 342—346. [周晓宇, 李镇之, 戚晓芳等. 数据挖掘技术初探 [J]. 小型微型计算机系统, 2002 23(3): 342—346.]
- [28] Zhang T, Pan H P. The Formal Algorithm of Decision Tree and Its Application in Geomatics [J]. *Bulletin of Surveying and Mapping* 2002 (7): 51—53 [张彤, 潘和平. 决策树的形式算法及其在地理信息学中的应用 [J]. 测绘通报, 2002 (7): 51—53.]
- [29] Breiman L, Friedman JH, Olshen R A *et al*. Classification and Regression Trees [M]. Monterey, California, U. S. A.: Wadsworth International Group, 1984. 1—358
- [30] Yohannes Y, Hoddinott J. Classification and Regression Tree: An Introduction [M]. Washington, D. C., U. S. A.: International Food Policy Research Institute, 1999.
- [31] Haralick R M, Shanmugan K, Dinstein I. Texture Features for Image Classification [J]. *IEEE Transactions on Systems Man and Cybernetics* 1973 3(6): 610—621.
- [32] Peddle D R, Franklin S E. Image Texture Processing and Data Integration for Surface Pattern Discrimination [J]. *Photogrammetric Engineering & Remote Sensing* 1991, 57(4): 413—420
- [33] Ryherd S, Curtis W. Combining Spectral and Texture Data in the Segmentation of Remotely Sensed Images [J]. *Photogrammetric Engineering & Remote Sensing* 1996 62(2): 181—194.
- [34] Zhang Y. Optimisation of Building Detection in Satellite Images by Combining Multispectral Classification and Texture Filtering [J]. *ISPRS Journal of Photogrammetry & Remote Sensing* 1999 54 50—60
- [35] Treitz P, Howarth P. Integrating Spectral, Spatial and Terrain Variables for Forest Ecosystem Classification [J]. *Photogrammetric Engineering & Remote Sensing* 2000 66(3): 305—317
- [36] Narasimha R P V, Sesha S M V R, Sreenivas K *et al*. Textural Analysis of RS 1D Panchromatic Data for Land Cover Classification [J]. *International Journal of Remote Sensing* 2002 23(17): 3327—3345
- [37] Marceau D J, Howarth P J, Dubois J M, *et al*. Evaluation of the Grey Level Co-Occurrence Matrix Method for Land Cover Classification Using SPOT [J]. *IEEE Transactions on Geoscience and Remote Sensing* 1990 28: 513—519
- [38] Xia D S, Fu D S. Technology and Application of the Modern Image Processing [M]. Nanjing: Jiangsu P. R. China Southeast University Press, 1997. [夏德深, 傅德胜. 现代图像处理技术与应用 [M]. 南京: 东南大学出版社, 1997.]
- [39] Franklin S E, Peddle D R. Spectral Texture for Improved Land Discrimination in Complex Terrain [J]. *International Journal of Remote Sensing* 1989 54 1727—1734
- [40] Wu J B, Yang X W. Accuracy Analysis of Classification of Remotely Sensed Data [J]. *Remote Sensing Technology and Application* 1995 10(1): 17—24 [吴健平, 杨星卫. 遥感影像分类结果的精度分析 [J]. 遥感技术与应用, 1995 10(1): 17—24.]

Cart based Land Use /cover Classification of Remote Sensing Images

ZHAO Ping^{1,2}, FU Yun-fei¹, ZHENG Liu-gen¹, FENG Xue-zhi³, B Satyanarayana⁴

(1 School of Earth & Space Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China)

2 College of Territorial Resources and Tourism, Anhui Normal University, Wuhu, Anhui 241009, China

3 Department of Urban & Resources Science in Nanjing University, Nanjing, Jiangsu 210093, China

4 Department of Zoology, Andhra University, Vishakhapatnam 530003, India

Abstract Nowadays investigations on land use / land cover change detections constitute a main objective for the global research. As a part of rapid development in technology, remote sensing has become an important tool to acquire the information of the land use /cover. Therefore, how best the extraction of timely and accurate information from these remotely sensed images is an impending problem. Recently, the knowledge-based interpretation of these images has become an effective and efficient approach to realize the automatic interpretation, which can integrate the spectral and other associated information based on experts' knowledge and experience to improve the accuracy. However, it is a bottleneck problem to obtain the knowledge for its wide application. A case study on the land use /cover classification of Jiangning study area in Jiangsu Province is discussed in the present article. At first, the data are preprocessed, then the relevant sixteen variables including geographical coordinate, grey value of four bands, textural statistics, DEM, slope and aspect are selected and extracted. The defined training sample areas are picked up by stratified random sampling techniques based on geographical coordinates. Thirdly, classification rules are discovered from these samples through Classification and Regression Tree (CART) Analysis, which integrates spectral, textural and the spatial distribution characters. Fourthly, the interpretation was performed by a judgment based on these rules. Finally, the traditional supervised as well as logic channel classifications are also performed to check the classification accuracies. The results have suggested that the accuracy of classification based on the CART is higher than others', which can obtain a lot of reasonable rules most quickly and effectively. So, it was felt that it is a good way to promote the wide application of knowledge-based interpretation of remote sensing images.

Key words Classification and Regression Tree (CART) Analysis; remote sensing; land use /cover classification; knowledge