

C5.0 决策树 Hyperion 影像森林类型精细分类方法

王怀警¹, 谭炳香¹, 房秀凤¹, 李世明¹, 李太兴²

(1. 中国林业科学研究院 资源信息研究所遥感室, 北京 100091; 2. 吉林省白河林业局, 吉林 延边 133613)

摘要: 以吉林省白河林业局为中心研究区, 利用星载高光谱 Hyperion 数据并结合其他辅助数据, 综合利用影像光谱特征、纹理特征、地形特征、典型地类和主要森林类型外业调查样本数据, 探究针对 C5.0 决策树算法的高光谱影像土地覆盖类型多层次信息提取与森林类型识别的有效方法。在分析典型地物光谱特征的基础上, 优选 8 种纹理特征, 引入主成分分量及与主要森林类型空间分布相关的敏感地形因子, 采用分层分类的策略, 根据光谱特征将地类划分层次, 在层次间建立基于 C5.0 决策树算法的决策树模型, 对研究区的地类进行细分。为便于对比, 以相同的策略采用支持向量机(SVM)分类器进行分类。最后, 结合野外采集样本并参考高分辨率影像, 采用分层随机抽样的独立检验样本对森林类型精细识别结果进行精度验证。结果表明: C5.0 决策树算法可综合利用高光谱影像的光谱、纹理及其他辅助数据, 自动寻找出区分各类别的最佳特征变量及分割阈值, 运算速度快, 占用内存较小且无需人为参与, 其分类精度达到优势树种级别, 总体分类精度达 81.9%, Kappa 系数 0.709 8。图 7 表 6 参 25

关键词: 森林经理学; Hyperion; C5.0 决策树; 分层分类; 森林类型分类; 高光谱

中图分类号: S757.4

文献标志码: A

文章编号: 2095-0756(2018)04-0724-11

Precise classification of forest types use Hyperion image based on the C5.0 decision tree algorithm

WANG Huaijing¹, TAN Bingxiang¹, FANG Xiufeng¹, LI Shiming¹, LI Taixing²

(1. Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China; 2. Baihe Forest Enterprise of Jilin Province, Yanbian 133613, Jilin, China)

Abstract: To explore the potential for using the C5.0 decision tree algorithm with hyperspectral data in precise classification of forest types, Baihe Forestry Bureau in Jilin Province was used as the center of the study area. Hyperion hyperspectral data, other auxiliary data, comprehensive utilization of an image's spectral feature, texture and terrain features, typical land types, and field investigation data of main forest types were used to support the study. Using the hierarchical classification strategy and according to characteristics of the spectrum, forest types were divided into different levels based on analysis of the typical spectral characteristics, eight optimum texture features, a principal component analysis, and sensitive terrain factors related to spatial distribution of the main forest types. Then the decision tree model based on the C5.0 decision tree algorithm was established for different levels. The robust Support Vector Machines (SVM) classifier was selected for comparison using the same strategy. Finally, verification samples were generated from the classification results using stratified random sampling method, combined with high-resolution remote sensing images, to evaluate the classification accuracy. Results showed that the C5.0 decision tree algorithm could comprehensively utilize the spectrum, texture, and other auxiliary data of hyperspectral images to automatically determine the best feature variables and segmentation thresholds for distinguishing each category. The algorithm also had the advantage of

收稿日期: 2017-08-25; 修回日期: 2017-11-01

基金项目: 国防科工委高分辨率对地观测系统重大专项(30-Y20A37-9003-15/17-3); 浙江省省院合作林业科技项目(2017SY04); 国家自然科学基金资助项目(31370635)

作者简介: 王怀警, 从事高光谱遥感林业应用研究。E-mail: wanghuaijing271413@163.com。通信作者: 谭炳香, 研究员, 博士, 从事遥感技术在林业中的应用研究。E-mail: tan@ifrit.ac.cn

fast computing speed, minimum occupied memory, and no human involvement. Classification accuracy reached the level of the dominant tree species, along with an overall classification accuracy of 81.9% and Kappa coefficient of 0.709 8. This method can be used for the precise mapping of forest types of hyperspectral remote sensing images, and can be used as reference for forest type classification of GF-5 hyperspectral images. [Ch, 7 fig. 6 tab. 25 ref.]

Key words: forest management; Hyperion; C5.0 decision tree; hierarchical classification; classification of forest types; hyperspectral

遥感技术具有覆盖范围大、重访周期短、应用成本低等优势,能及时准确地掌握森林类型、分布、面积、结构、质量、现状及动态变化情况,在森林区划、森林资源调查、森林类型精细识别、植被制图^[1]、动态变化监测等方面具有巨大的应用潜力。利用遥感影像开展土地利用类型分类和森林类型识别已有较多研究,主要集中在非参数化智能化分类、多源遥感数据与辅助信息综合分类^[2-3]、知识挖掘和专家系统、多时相复合分类^[4]、面向对象^[5]、新方法引入与分类策略^[6]、多分类器组合^[7-9]等方面,并且在植被分类、林地信息提取、森林类型精细识别^[10-12]、树种(组)分类、动态变化监测等方面获得广泛的应用。当前,针对高光谱影像的分类方法,如光谱角匹配方法、最大似然法、人工神经网络和支持向量机(SVM)等,多为基于像元光谱信息来赋予像元不同的地物类型,比较成熟和稳定。但此类方法往往忽视空间信息,或者对空间信息利用不充分,在没有辅助数据或者辅助数据较少的情况下,如何利用有限的已知样本点所提供的空间位置信息或其他来源的辅助信息提高分类精度值得深究。C5.0 决策树算法是一种以信息熵为核心思想的数据挖掘算法,可以快速寻找现象之间未知的关系和关联。前人对该算法的应用多集中在土地利用信息自动提取^[13]、土地覆盖分类^[14-15]、植被分类^[16]、森林变化检测^[17]等方面;另外,还探究不同特征对分类的影响及不同方法的比较。然而,鲜有应用 C5.0 决策树算法探究高光谱数据森林类型分类。本研究以吉林省白河林业局为中心研究区,充分利用影像多元特征和辅助信息,开展基于 C5.0 决策树算法的森林类型识别方法研究,探究针对主要森林类型或树种(组)C5.0 决策树分类方法的适用性和可行性。

1 研究区概况与数据获取

1.1 研究区概况

研究区中心区域位于吉林省延边朝鲜族自治州白河林业局境内,地理坐标为 41°41′49″~42°51′18″N, 127°42′55″~128°16′48″E。研究区覆盖面积约为 4.05 万 hm^2 ,属温带大陆性山地气候,南临长白山保护区,森林覆盖率约 85%。森林植被组成和树种成分较为复杂,属阔叶混交、针叶混交和针阔混交林带,基本特点为次生林,萌生起源为主,兼有实生林及实生树木。研究区主要树种及森林类型为美人松 *Pinus sylvestris* 林,落叶松 *Larix gmelinii* 林,樟子松 *Pinus sylvestris* var. *mongolica* 林,杨树 *Populus bonatii* 林,白桦 *Betula platyphylla* 林,柞树 *Quercus mongolica* 林以及阔叶混交林、针叶混交林、针阔混交林、灌木林。

1.2 数据获取

研究获取 2015 年 7 月 9 日 Hyperion 影像一景,覆盖范围为 41.93°~42.98°N, 127.97°~128.23°E,影像幅宽 7.7 km × 185 km,光谱范围 355~2 500 nm,共有 242 个波段,光谱分辨率为 10 nm,空间分辨率为 30 m。此外,辅助数据包括研究区行政界线矢量数据、先进星载热发射和反射辐射仪全球数字高程模型(advanced spaceborne thermal emission and reflection radiometer global digital elevation model, ASTER GDEM)30 m 分辨率数字高程模型(digital elevation model, DEM)数据、研究区部分二类调查数据(图1),矢量数据用于研究区影像数据裁剪;DEM 数据作为特征因子参与分类;二类调查数据作为精度验证的辅助数据。

Hyperion 影像的预处理包括坏波段剔除、坏线修复、Smile 效应校正、辐射校正、大气校正、几何校正、研究区裁剪,利用 DEM 数据和野外采集的全球定位系统(GPS)坐标点对影像进行正射校正。此外,为消除噪声带来的影响,还对数据进行了光谱平滑处理,为方便起见,反射率数值放大 1.0 万倍。

剔除质量较差的波段，最终保留 148 个波段用于研究，表 1 列出了保留的波段及对应的波长区间(表 1 和图 2)。

表 1 保留的 Hyperion 影像波段
Table 1 Reserved bands of Hyperion image

编号	保留波段/nm	编号	保留波段/nm
1	8~57	3	135~164
	(426.82~925.41)		(1 497.63~1 790.19)
2	79~119	4	192~218
	(932.64~1 336.15)		(2 072.65~2 335.01)

说明：括号内数值为对应的波长区间

1.3 外业数据

2016 年 9 月 3-9 日，以 Hyperion 影像覆盖范围为主要调查区域进行外业调查。样地为直径 45 m 的圆形样地，在样地中心采用 GPS 手持机(Trimble Geo Explorer 6000)记录样地中心位置坐标，同时对样地进行多角度拍摄，记录森林类型、树种组成、优势树种(组)、郁闭度、林龄、林下灌木种类，海拔、坡度、坡向、经营活动等主要调查因子，详实反映样地及其周围林分植被生长状况，使样地的纹理、属性和位置信息一一对应。野外调查共采集 Hyperion 影像条带覆盖区域实测样地 81 个。样地在 Hyperion 影像条带上的空间分布如图 2 所示，实测样地类型及样地数见表 2。由于样地可达性限制，部分不能抵达观测的样地在图纸和记录表中分别记录其相对位置和相关属性，可用于辅助精度验证。

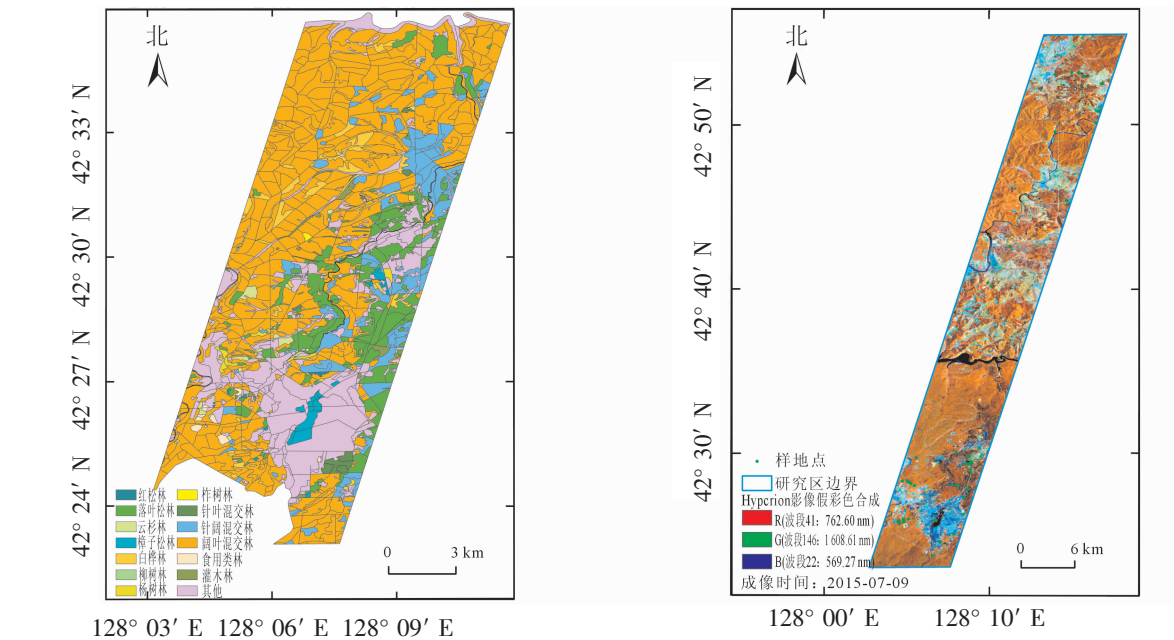


图 1 研究区部分森林类型分布图
Figure 1 Distribution of some forest types in the study area

图 2 研究区 Hyperion 影像及样地分布图
Figure 2 Hyperion image and plot distribution map of the study area

表 2 实测样地类型及样地数
Table 2 Measured sample types and number

地物类型	样地数/个	地物类型	样地数/个	地物类型	样地数/个
美人松林	10	阔叶混交林	4	耕地	4
落叶松林	12	针叶混交林	8	园地	4
樟子松林	3	针阔混交林	5	水体	2
白桦林	6	灌木林地	2	未利用地	3
杨树林	3	无立木林地	2	建设用地	2
柞树林	5	湿地	6		

2 研究方法

2.1 分类系统

表 3 分类系统	
Table 3 Classification system	
1 级类型	2 级类型
林地	有林地
	乔木林
	纯林
	美人松林
	落叶松林
	樟子松林
	白桦林
	杨树林
	柞树林
	混交林
	阔叶混交林
	针叶混交林
	针阔混交林
	灌木林地
	无立木林地
湿地	
非林地	耕地
	园地
	水域
	未利用地
	建设用地

2.2 样本选择

由于样地可达性及工作量等因素限制,野外调查的样地数据比较有限,不足以进行分类,因此需对样本量进行适量扩充。具体做法如下:根据地物在空间分布具有连续性的特点,综合已测样地位置、纹理、影像中光谱曲线及记录表中相关信息,充分利用高光谱影像的精细光谱特征,重点针对森林类型,围绕样地周边像元进行拓展,当拓展样本与已测样地的光谱、纹理及记录表中相关信息相匹配时,即可判定该拓展样本的类别,样本选择结果见表4。

2.3 分类策略

根据影像中地物光谱反射差异,选择特征波段,先区分出易于区分的地类,将已区分的类别进行掩膜处理,可大大减少其对后期分类工作的干扰,利于提高分类精度。利用不同地物的特征波段提取某一

表 4 地类列表及样本数量

Table 4 List of land types and sample number

地类编号	地物类型	样本数(像元)/个	地类编号	地物类型	样本数(像元)/个
1	美人松林	475	10	灌木林地	49
2	落叶松林	66	11	无立木林地	69
3	樟子松林	63	12	湿地	103
4	白桦林	255	13	耕地	2 453
5	杨树林	63	14	园地	195
6	柞树林	70	15	水体	1 828
7	阔叶混交林	1 033	16	未利用地	121
8	针叶混交林	700	17	建设用地	1 102
9	针阔混交林	829			

地类或某几类地物, 即对地物进行分层, 随后在层次间对地类进一步细分, 针对层次间光谱特征较为相似的地类, 综合利用纹理信息和地形因子, 构建基于 C5.0 算法的决策树模型并进行分类。最后对分类结果进行决策级融合, 得到最终分类结果。

2.4 特征提取

高光谱数据波段多, 波段间相关性较高, 在进行分类研究时, 通常需进行特征提取或特征选择, 以达到降维或波段优选的目的。常通过微分变换的导数谱、对数变换、微分对数变换相结合、主成分分析 (PCA)^[20-21] 等进行高维数据的压缩。经试验对比, 选用 PCA 法进行降维, 选取 PCA 的前 4 个分量 (信息量达 99% 以上) 参与分类。

对于光谱特征相似的地物, 往往通过其纹理差异加以区分^[22-23]。根据研究区地类特征和影像纹理特征差异, 经反复测试, 分别计算 PCA 前 3 个分量各自对应的 8 个纹理特征^[24]参与分类, 选取的纹理特征如表 5 所示。

表 5 灰度共生矩阵计算的纹理因子

Table 5 The texture factor of gray level co-occurrence matrix

编号	纹理因子	编号	纹理因子	编号	纹理因子
1	均值	4	对比	7	二阶矩
2	方差	5	相异性	8	相关性
3	一致性	6	信息熵		

研究区主要森林类型及优势树种 (组) 的生态学特性 (如喜光、喜阴), 树种群落特征和生态适应范围 (如柞树多生长在向阳的山坡上), 垂直向分布特征等较为明显, 因此将数字高程模型 (DEM), 坡度 (aspect), 坡向 (slope) 作为树种分类的有效特征和辅助信息, 以期进一步提高森林类型识别精度。

2.5 C5.0 决策树算法

C5.0 决策树算法最早的原型是由 QUINLAN 于 1979 年提出的 ID3 算法, 后经不断改进形成 C4.5 算法, C5.0 算法是基于 C4.5 的进一步改进。C5.0 增加了 Boosting 算法以提高分类精度。该算法以信息增益率为标准确定最佳分组变量和最佳分割点, 其核心概念是信息熵。信息熵又称为先验熵, 是信息发送前信息量的数学期望值^[25]。C5.0 以信息熵的下降速度作为选取最佳分支变量和分割阈值的依据, 信息熵的下降意味着不确定性下降。

信息 $u_i (i=1, 2, \dots, r)$ 的发生概率 $P(u_i)$ 组成信源数学模型 $\sum_{i=1}^r P(u_i)=1$, 信息量和信息熵的计算公式如下:

$$I(u_i)=-\log_2 P(u_i);$$

$$H(U)=-\sum_{i=1}^r P(u_i)\log_2 P(u_i)。$$

信息熵 $H(U)$ 为 0 时表示只存在唯一的可能性, 不存在不确定性; 如果信源的 k 个信号有相同的发出概率, 即所有的 u_i 有 $P(u_i)=1/k$, $H(U)$ 达到最大, 不确定性最大, $P(u_i)$ 差别越小, $H(U)$ 就越大。设 S 是 1 个样本集合, 目标变量 C 有 K 个分类, $f_{\text{req}}(C_i, S)$ 表示属于 C_i 类的样本数, $|S|$ 表示样本集合 S 的样本数, 则集合 S 的信息熵定义为:

$$I_{\text{info}}(S)=-\sum_{i=1}^k [f_{\text{req}}(C_i, S)/|S| \log_2 [f_{\text{req}}(C_i, S)/|S|]。$$

如果某属性变量 T , 有 m 个分类, 则属性变量 T 引入后的条件熵定义为:

$$I_{\text{info}}(T)=-\sum_{i=1}^n (|T_i|/|T|) I_{\text{info}}(T_i)。$$

属性变量 T 带来的信息增益为:

$$G_{\text{ain}}(T)=I_{\text{info}}(S)-I_{\text{info}}(T)。$$

C5.0 算法使用十折交叉验证的方法, 分类过程中及时反馈训练样本的质量, 及时修改模型, 避免出

现“过度拟合”现象, 保证较高的分类精度。C5.0 采用后剪枝 (post-pruning) 策略自叶节点向上逐层剪枝, 使用统计置信区间的误差估计方式, 直接在训练数据中估计误差, 若待剪子树中叶节点误差大于父节点的误差, 则予以剪去。C5.0 在选择最佳分组变量时, 通常会将带有缺失值的样本当作临时样本剔除, 并进行权数调整处理, 使得算法对样本具有一定的容错能力。

2.6 影像分类

针对不易区分的几种地类, C5.0 决策树算法首先计算输入样本集的熵和不同特征波段加入后的信息增益, 根据信息增益最大的字段拆分样本, 第 1 次拆分确定的属性作为树的根节点, 随后根据其他属性再次拆分, 后建立的决策树重点考虑之前被错分和漏分的数据, 直到样本子集不能再被拆分; 而属性阈值的分割则是以信息熵下降最快为准。如此建立一株完整的决策树。最后根据样本集对生成的决策树进行剪枝, 剪枝的标准是叶子节点的错误率小于父节点。为提升决策树的性能, 采用交叉验证技术, 对选出的属性进行投票。根据构建的决策树模型进行细分, 直至所有类别均被区分。对分类结果进行决策级融合并进行重编码, 得到所有地类的分类结果。对分类结果进行分类后处理, 合并分类结果中碎小的细部。为了体现本研究方法的优势, 在相同的分类策略下, 选用性能稳健优异的 SVM 分类器进行分类, 采用网格参数寻优法寻找不同层次分类的最优参数, 确保分类结果较好。

3 结果与分析

3.1 光谱特征分析与地物分层方案

根据样本对影像进行分类统计, 求取各波段反射率均值, 得到各类别的光谱曲线。图 3 为获取的各类别反射率曲线, 可用于分析各类别的光谱差异及确定地物分层方案。

由光谱曲线可见, 建设用地在蓝光到红光区间反射率呈上升趋势, 不同于其他地类; 水体在整个光谱区间内反射率较低; 植被在 447 nm 处有明显的吸收谷, 绿光区间有绿峰, 红光区间有吸收谷, 近红外区间有明显的反射峰; 柞木林和灌木林在近红外到短波红外区间反射率较其他植被高很多; 园地在蓝光到绿光区间内, 反射率先快速上升而后趋于平稳, 在 447 nm 处的吸收谷消失, 近红外区间的反射峰值低于其他植被; 耕地在中红外区间表现出较强的反射特性, 在 1 659 nm 处达到峰值, 在大于 2 000 nm 的波长区间, 反射率较其他地类高许多。樟子松林、美人松林、湿地、无立木林地等 4 类的光谱曲线较为相近, 由于部分水体与植被相邻, 为综合多种属性提高分类精度, 将此 4 类分为同一层; 白桦林、杨树林、落叶松林、阔叶混交林、针叶混交林、针阔混交林、建设用地等 7 种地类光谱曲线差异不明显, 不易区分, 需辅以其他信息方能加以区分。

根据地物光谱曲线差异选择特征波段, 对研究区地类进行分层, 特征波段选择和分层策略见图 4, 分别第 4 层和第 7 层中应用 C5.0 决策树算法构建决策树模型, 对同一层中较难区分的地类进行细分。

3.2 分类结果与精度评价

为验证方法的适用性和森林类型精细识别的精度, 采用分层随机抽样产生独立验证样本。检验样本数与训练样本数比例约为 1:3, 参考野外采集样本、高分辨率影像(Google Earth)和研究区部分二类调查数据, 对森林类型进行精度验证。为验证森林类型分类精度, 将非森林类型合并为其他类, 选取总体精度、Kappa 系数、用户精度和生产精度作为评价指标。C5.0 决策树算法森林类型总体分类精度为 81.9%, Kappa 系数为 0.709 8, SVM 分类器森林类型总体分类精度为 84.2%, Kappa 系数为 0.717 8, 2

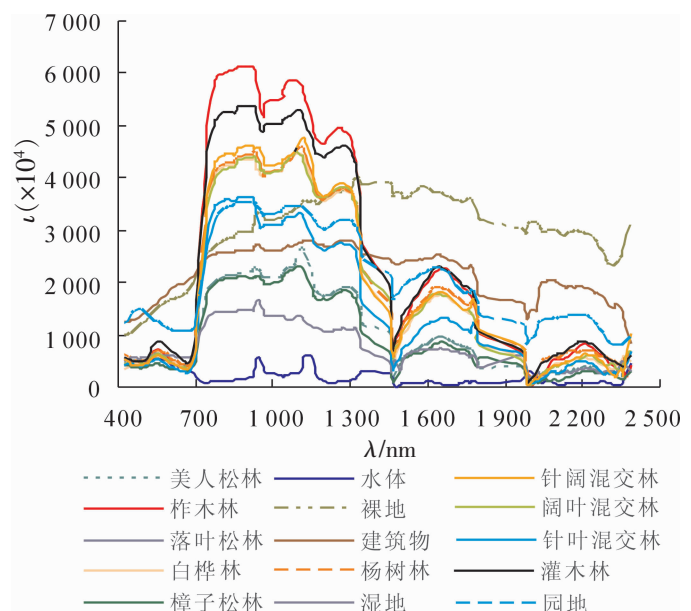


图 3 典型地物光谱曲线

Figure 3 Spectral curves of typical objects

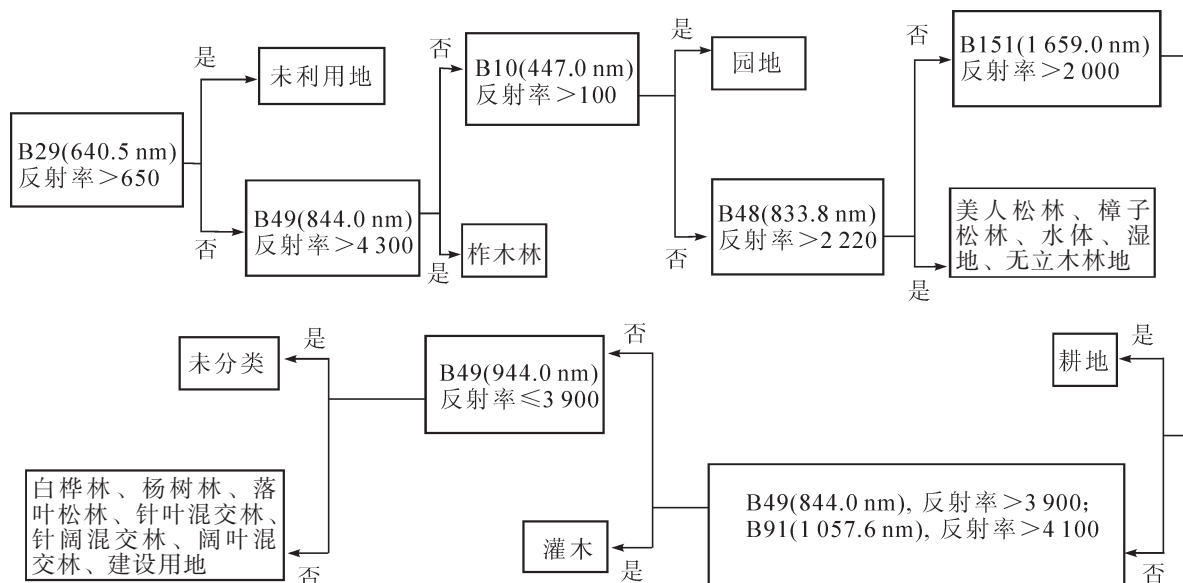


图 4 地物分层分级方案及阈值

Figure 4 Hierarchical classification scheme and threshold

种方法的分类结果和分类精度评价分别见图 5 和表 6。

C5.0 算法决策树模型构建过程中特征变量使用的频率可作为衡量各变量对分类贡献大小的依据, 该算法区分第 4 和第 7 层地物使用的特征变量不同。经优化后的第 4 层和第 7 层地物分类决策树模型见图 6。

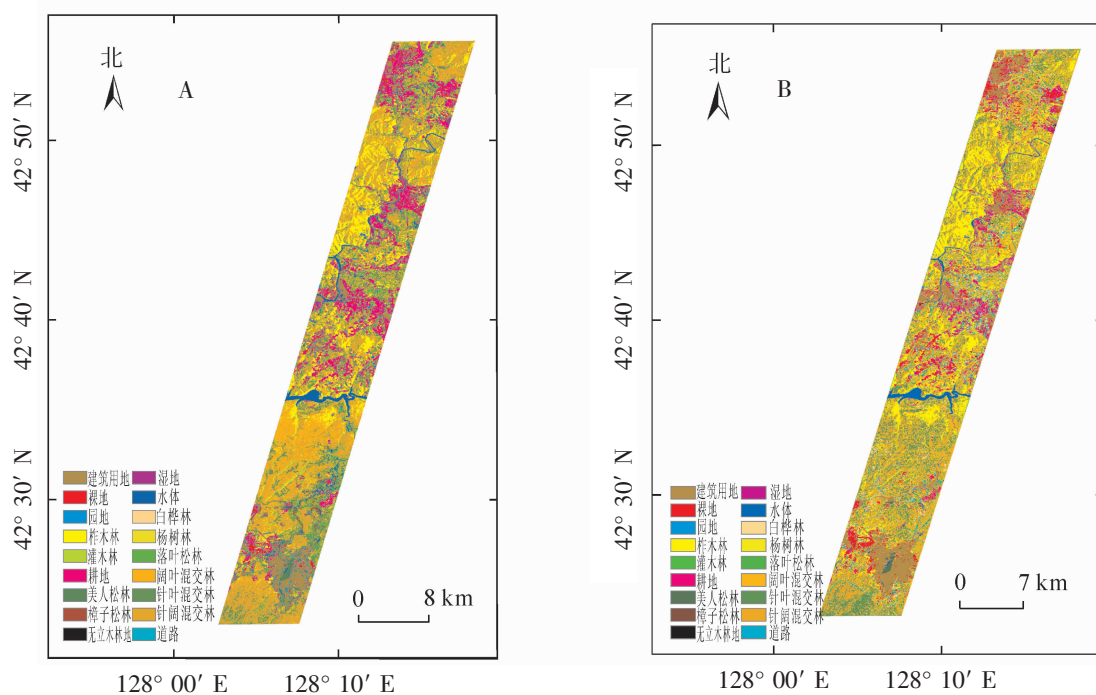


图 5 C5.0 决策树分类法(A)与 SVM 法(B)森林类型分类图

Figure 5 Classification Result of C5.0 Decision tree Classifier(A) and SVM Classifier(B)

为了进一步印证 C5.0 选择特征波段的有效性, 对各类别相应的特征波段进行统计分析, 选取各类型特征波段均值绘制如图 7 所示的离散直方图, 从图 7 可以清晰的看出各类别不同特征间存在明显差异, 相比光谱特征具有更好的区分性, 从侧面证实了 C5.0 决策树算法构建的决策树模型的正确性。

由表 6 可见: 2 种分类方法结果较一致, 其中落叶松林、樟子松林、白桦林、杨树林、阔叶混交林分类精度相对较低; 阔叶混交林、针叶混交林和针阔混交林间混分现象较明显; 美人松林、桉树林、针

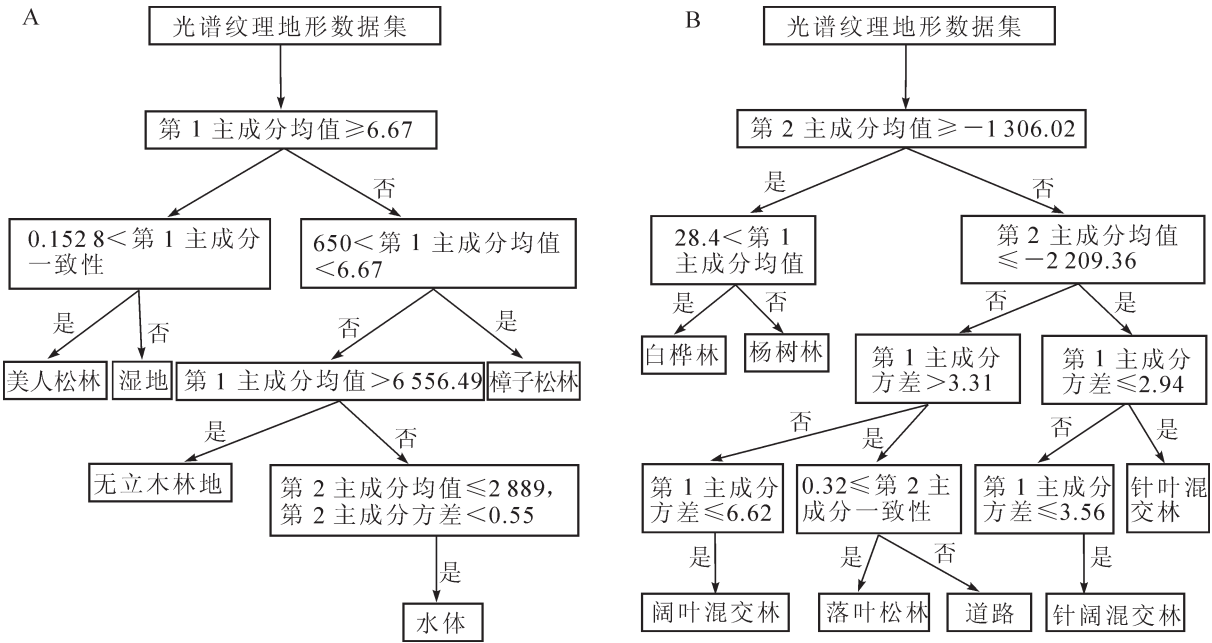


图 6 第 4 层(A)和第 7 层(B)地物分类决策树模型

Figure 6 Classification decision tree model of the fourth layer(A) and the seventh layer(B) terrain

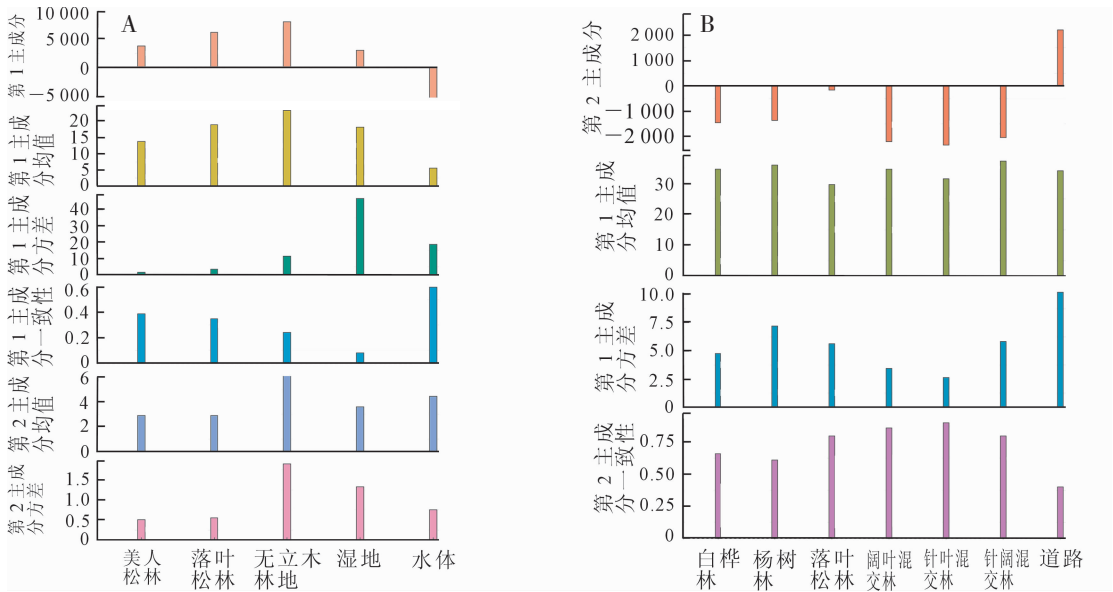


图 7 第 4 层(A)和第 7 层(B)森林类型分类所选特征波段均值的离散直方图

Figure 7 Discrete histograms of selected characteristic bands for fourth (A) and seventh (B) forest types

叶混交林、针阔混交林分类精度相对较高。C5.0 算法对落叶松林、樟子松林、杨树林、针叶混交林和针阔混交林分类精度优于 SVM 法，SVM 分类器对白桦林、柞木林和阔叶混交林区区分较好，其他类别两者分类度相差不大。

对比 C5.0 决策树法和 SVM 法，2 种方法总体分类精度均较高，SVM 法总体分类精度高于 C5.0 决策树算法，但 SVM 法在内存占用、耗时等方面远高于 C5.0 决策树法，SVM 分类器存在最优参数选择的问题，而 C5.0 决策树算法则参数较少。2 种方法对植被的识别能力均弱于其他地物，但相比而言，C5.0 决策树算法对绿色植被识别能力优于 SVM 法；2 种方法对不同地物的识别能力不同，存在优势互补的可能。

3.3 分类结果讨论

训练样本选择准确对分类至关重要。野外调查时间与影像成像时间相隔约 1 a，研究中忽略了此间变化产生的影响。研究区地物类型丰富，不同地类混杂较严重，导致分类结果较为破碎。研究区森林覆

表 6 C5.0 决策树算法与 SVM 法森林类型分类精度评价

Table 6 Forest land type classification accuracy evaluation result of C5.0 decision tree algorithm and SVM

地类	美人松林	落叶松林	樟子松林	白桦林	杨树林	柞木林	针叶混交林	针阔混交林	阔叶混交林	灌木林	其他	用户精度/%
美人松林	139	9	0	0	0	0	6	0	12	0	15	76.8
落叶松林	1	69	0	8	0	0	4	0	7	0	33	56.6
樟子松林	0	5	10	0	0	0	0	0	9	0	3	34.5
白桦林	0	0	0	39	0	0	0	0	0	0	31	55.7
杨树林	0	0	0	0	9	0	0	0	0	0	2	81.8
柞木林	0	1	0	0	0	98	54	32	17	2	8	46.2
针叶混交林	17	27	1	10	0	0	268	14	71	0	10	63.8
针阔混交林	3	3	0	0	4	9	1	256	31	0	13	79.5
阔叶混交林	1	1	0	1	0	0	43	94	181	0	5	55.5
灌木林	0	0	0	1	0	3	5	3	16	9	72	8.2
其他	2	16	5	4	2	3	2	0	10	2	2 392	
生产者精度/%	85.2	52.7	62.5	61.9	60.0	86.7	70.0	64.2	51.1	69.2	92.6	

C5.0 决策树算法总体分类精度为 81.9%; Kappa 系数为 0.709 8

地类	美人松林	落叶松林	樟子松林	白桦林	杨树林	柞木林	针叶混交林	针阔混交林	阔叶混交林	灌木林	其他	用户精度/%
美人松林	137	0	1	0	0	0	0	0	0	0	0	99.3
落叶松林	3	61	1	0	0	0	0	0	5	0	1	66.3
樟子松林	1	8	9	0	0	0	6	2	16	0	3	37.5
白桦林	0	2	0	37	2	2	1	0	1	0	36	45.7
杨树林	0	0	0	1	5	0	1	0	0	0	0	71.4
柞木林	0	1	0	1	0	102	81	36	36	2	28	35.5
针叶混交林	10	9	0	6	0	3	227	37	19	0	5	71.8
针阔混交林	7	1	0	4	3	0	16	117	13	0	1	74.5
阔叶混交林	0	33	2	6	2	0	34	11	190	0	72	53.1
灌木林	0	0	0	1	0	3	0	0	0	8	47	14.3
其他	2	6	2	2	2	0	0	0	3	3	2 525	
生产者精度/%	85.6	50.4	60.0	64.9	35.7	91.9	62.0	57.6	67.1	61.5	92.9	

SVM 法总体分类精度为 84.2%; Kappa 系数为 0.717 8

盖度较高, 针叶混交林、针阔混交林和阔叶混交林混杂严重, 加之三者特征相近, 故较难区分。由于不同地类训练样本数量不同, 训练样本数较少, 可能导致其分类精度较低。可见, 纹理信息有助于提高某些地物分类精度, 但限于影像空间分辨率, 导致纹理信息不足以精细地描述不同森林类型的差异; 地物尺寸较大时纹理信息对分类精度提升明显, 例如流线状的河流和道路。纹理因子对分类的贡献率与影像分辨率间的关系, 有待进一步探究。

结合以上实验分析, 总结如下: ①针对土地覆盖类型复杂的区域, 采用分层分类的策略可提高总体的分类精度, 且分层数不宜过多; ②建立的决策树模型深度越深, 其精度越小, 故决策树的深度不宜过深; ③C5.0 决策树算法对绿色植被的识别能力弱于其他类别; ④分层分类策略结合 C5.0 决策树算法应用于高光谱森林类型分类, 易与其他辅助数据结合, 森林类型识别可达到优势树种(组)级别, 可行性好, 可应用于实际生产; ⑤综合遥感影像的光谱、纹理和地形信息的分层信息提取方法, 森林类型识别更为精细, 能满足复杂地形条件下星载高光谱影像森林类型精细识别的应用需求, 对中国 GF-5 号高光谱遥感数据林业应用具有参考价值。

4 结论

研究表明: ①该算法对绿色植被的识别能力弱于其他类别; ②该算法在训练样本数较少时依然表现较好, 可充分利用影像的光谱、纹理及其他辅助信息, 不会或较少出现局部收敛现象; ③地形较复杂区

域进行分类时,地形因子对分类贡献度较高;④C5.0 决策树算法,计算速度快,占用内存小,自动选择特征变量和分割阈值,且生成的决策树规则易于理解,总体分类精度较高;⑤该算法在利用纹理因子参与分类时,均值、方差、一致性、相异性 4 个变量对分类贡献较高。

研究采用的数据空间分辨率较低,混合像元现象严重,以后可考虑从混合像元分解入手改进分类策略和算法。C5.0 决策树算法存在与其他分类器(如 SVM)优势互补的潜力,今后可考虑多分类器融合或组合策略,以提高森林类型分类精度。

5 参考文献

- [1] 张志明,张征凯,郭银明,等.高原山区遥感植被制图研究综述[J].云南大学学报(自然科学版),2013,35(3): 416 – 427.
ZHANG Zhiming, ZHANG Zhengkai, GUO Yinming, *et al.* Mountain vegetation mapping using remote sensing [J]. *J Yunnan Univ*, 2013, 35(3): 416 – 427.
- [2] VOISIN A, KRYLOV V A, MOSER G, *et al.* Supervised classification of multisensory and multiresolution remote sensing images with a hierarchical copula-based approach[J]. *IEEE Trans Geosci Remote Sens*, 2014, 52(6): 3346 – 3358.
- [3] 任冲,鞠洪波,张怀清,等.多源数据林地类型的精细分类方法[J].林业科学,2016,52(6): 54 – 65.
REN Chong, JU Hongbo, ZHANG Huaiqing, *et al.* Multi-source data for forest land type precise classification [J]. *Sci Silv Sin*, 2016, 52(6): 54 – 65.
- [4] 竞霞,王锦地,王纪华,等.基于分区和多时相遥感数据的山区植被分类研究[J].遥感技术与应用,2008,23(4): 394 – 397.
JING Xia, WANG Jindi, WANG Jihua, *et al.* Classifying forest vegetation using sub-region classification based on multi-temporal remote sensing images [J]. *Remote Sens Technol Appl*, 2008, 23(4): 394 – 397.
- [5] 温一博,范文义.多时相遥感数据森林类型识别技术研究[J].森林工程,2013,29(2): 14 – 20.
WEN Yibo, FAN Wenyi. Remote sensing image recognition for multi-temporal forest classification [J]. *For Eng*, 2013, 29(2): 14 – 20.
- [6] BENZ U C, HOFMANN P, WILLHAUCK G, *et al.* Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information [J]. *ISPRS J Photogramm Remote Sens*, 2004, 58(3/4): 239 – 258.
- [7] FOODY G M, BOYD D S, SANCHEZ-HERNANDEZ C. Mapping a specific class with an ensemble of classifiers [J]. *Int J Remote Sens*, 2007, 28(8): 1733 – 1746.
- [8] KITTILER J. Combining classifiers: a theoretical framework [J]. *Pattern Anal Appl*, 1998, 1(1): 18 – 27.
- [9] FREUND Y, SCHAPIRE R E. Experiments with a new boosting algorithm[C]// *ICML96 Proceesings of the Thirteenth International Conference on International Conference on Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc, 1996: 148 – 156.
- [10] MELVILLE P, MOONEY R J. Creating diversity in ensembles using artificial data [J]. *Inf Fusion*, 2005, 6(1): 99 – 111.
- [11] HELMER E H, RUZYCKI T S, BENNER J, *et al.* Detailed maps of tropical forest types are within reach: forest tree communities for Trinidad and Tobago mapped with multi-season Landsat and multi-season fine-resolution imagery [J]. *For Ecol Manage*, 2012, 279(6): 147 – 166.
- [12] 董心玉,范文义,田甜.基于面向对象的资源 3 号遥感影像森林分类研究[J].浙江农林大学学报,2016,33(5): 816 – 825.
DONG Xinyu, FAN Wenyi, TIAN Tian. Object-based forest type classification with ZY-3 remote sensing data [J]. *J Zhejiang A&F Univ*, 2016, 33(5): 816 – 825.
- [13] 李明诗,彭世揆,周林,等.基于 ASTER 数据的决策树自动构建及分类研究[J].国土资源遥感,2006,18(3): 33 – 36, 42.
LI Mingshi, PENG Shikui, ZHOU Lin, *et al.* A study of automated construction and classification of decision tree classifiers based on ASTER remotely sensed datasets [J]. *Remote Sens Land Resour*, 2006, 18(3): 33 – 36, 42.
- [14] 齐红超,祁元,徐瑱.基于 C5.0 决策树算法的西北干旱区土地覆盖分类研究:以甘肃省武威市为例[J].遥感技术与应用,2009,24(5): 648 – 653.

- QI Hongchao, QI Yuan, XU Zhen, *et al.* The study of the northwest arid zone land-cover classification based on C5.0 decision tree algorithm at Wuwei City, Gansu Province [J]. *J Remote Sens Technol Appl*, 2009, **24**(5): 648 – 653.
- [15] 高玉蓉, 许红卫, 丁晓东. 基于 C5.0 的钱塘江流域地区土地利用/覆被信息提取研究[J]. 生态科学, 2012, **31**(5): 481 – 487.
- GAO Yurong, XU Hongwei, DING Xiaodong. Extraction of land use/cover information based on C5.0 algorithm in Qiantang River drainage area [J]. *Ecol Sci*, 2012, **31**(5): 481 – 487.
- [16] 李梦莹, 胡勇, 王征禹. 基于 C5.0 决策树和时序 HJ-1A/B CCD 数据的神农架林区植被分类[J]. 长江流域资源与环境, 2016, **25**(7): 1070 – 1077.
- LI Menyng, HU Yong, WANG Zhengyu. Study on vegetation classification in Shennongjia forest district based on decision tree and HJ-1 A/B data [J]. *Resour Environ Yangtze Basin*, 2016, **25**(7): 1070 – 1077.
- [17] KEMPENEERS P, SEDANO F, SEEBACH L, *et al.* Data fusion of different spatial resolution remote sensing images applied to forest-type mapping [J]. *IEEE Trans Geosci Remote Sens*, 2012, **49**(12): 4977 – 4986.
- [18] 侯瑞萍, 黄国胜, 李应国, 等. LY/T 2188.1–2013 森林资源数据采集技术规范 第 1 部分: 森林资源连续清查[S]. 北京: 中国标准出版社, 2014.
- [19] 唐小平, 陈雪峰, 翁国庆, 等. GB/T 26424–2010 森林资源规划设计调查技术规程[S]. 北京: 中国标准出版社, 2011.
- [20] 王志慧, 李世明, 张艺伟. 基于 C5.0 算法的森林资源变化检测方法研究: 以山东省徂徕山林区为例[J]. 西北林学院学报, 2011, **26**(5): 185 – 191.
- WANG Zhihui, LI Shiming, ZHANG Yiwei. Methodological study on the detection of the variations of forest resources based on C5.0 algorithm: a case of Culai Forest in Shandong [J]. *J Northwest For Univ*, 2011, **26**(5): 185 – 191.
- [21] 崔宾阁, 马秀丹, 谢小云. 小样本的高光谱图像降噪与分类[J]. 遥感学报, 2017, **21**(5): 728 – 738.
- CUI Binge, MA Xiudan, XIE Xiaoyun. Hyperspectral image de-noising and classification with small training samples [J]. *J Remote Sens*, 2017, **21**(5): 728 – 738.
- [22] 张晓羽, 李凤日, 甄贞, 等. 基于随机森林模型的陆地卫星-8 遥感影像森林植被分类[J]. 东北林业大学学报, 2016, **44**(6): 53 – 57.
- ZHANG Xiaoyu, LI Fengri, ZHEN Zhen, *et al.* Forest vegetation classification of Landsat-8 remote sensing images based on random forests model [J]. *J Northeast For Univ*, 2016, **44**(6): 53 – 57.
- [23] 董连英, 邢立新, 潘军, 等. 高光谱图像植被类型的 CART 决策树分类[J]. 吉林大学学报(信息科学版), 2013, **31**(1): 83 – 89.
- DONG Lianying, XING Lixin, PAN Jun, *et al.* Vegetation classification in hyperspectral image with CART decision tree [J]. *J Jilin Univ Inf Sci Ed*, 2013, **31**(1): 83 – 89.
- [24] 赵英时. 遥感应用分析与方法[M]. 北京: 科学出版社, 2003: 156 – 158.
- [25] HARALICK R M. Statistical and structural approaches to texture [J]. *Proc IEEE*, 1979, **67**(5): 786 – 804.