

doi:10.6046/gtzyyg.2012.02.21

基于决策树方法的海岛土地利用分类研究

杨曦光^{1,2}, 黄海军¹, 严立文¹, 都本绪³

(1. 中国科学院海洋研究所, 青岛 266071; 2. 中国科学院研究生院, 北京 100049;

3. 大连市林业局, 大连 116023)

摘要:以山东省荣成市镆鵁岛为例,利用 SPOT 5 卫星数据,探讨了基于决策树方法的海岛土地利用类型的遥感分类。结果表明,利用决策树分类方法进行海岛土地利用类型分类,可以得到较好的分类结果(分类的平均精度达到 86.46%, Kappa 系数为 0.841 4);与其他分类方法比较,决策树分类法的分类精度有明显的提高,在海岛土地利用类型调查中具有较好的应用潜力。

关键词:分类; 土地利用; 决策树

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 1001-070X(2012)02-0116-05

0 引言

我国海岛众多,面积在 500 km² 以上的岛屿约有 6 500 多个,分布在相当于我国陆域面积 1/3 (约 300 多万 km²) 的海域中^[1]。遥感技术具有高光谱分辨率、高空间分辨率、实时观测、重访周期短等优势,已广泛应用于陆地、海岛等土地利用类型调查中^[2-4]。张华国等^[5]用 IKONOS 数据对福州市琅岐岛的土地覆盖类型进行监督分类,分类总体精度为 76.31%。近年来,基于知识的光谱特征遥感信息模型得到了快速发展,将知识作为分类的辅助信息参与分类过程计算,可有效地改善和提高分类精度^[6]。李春华等^[7]利用 TM 与 ASTER 的融合图像对福州市琅岐岛进行土地利用类型的分类研究,结果表明,基于知识的 Bayes 分类方法比运用单一分类方法的精度明显提高。决策树是一种利用逐层逻辑判别方式使人的知识及判别思维能力与图像处理有机结合起来的知识分析处理方法,具有较高的分类精度,已在陆地土地类型分类研究中得到了成功的应用^[8-12],但在海岛土地利用分类研究中尚不多见。

本文以山东省荣成市镆鵁岛为研究靶区,以 SPOT 5 卫星图像为数据源,采用决策树方法进行海岛土地利用类型的分类研究,通过与传统单一分类方法比较,发掘决策树分类法在海岛土地利用调查中的相对优势、局限性及其应用潜力。

1 研究区与数据源

研究区为山东省荣成市镆鵁岛,位于 E122°31'00", N36°54'48", 属于人工陆连岛。岛屿形状不规则,西南—东北走向,南北长 5.3 km,东西宽 0.84 km,岛陆面积 4.62 km²。岛内陆域土地利用类型包括耕地、林地、草地、住宅用地等多种地类。

研究采用的遥感数据源为 2004 年 8 月获取的 10 m 空间分辨率的 SPOT 5 多光谱数据。

2 研究方法

2.1 训练样本选择

训练样本的选取是建立分类决策树模型的关键步骤,直接关系到规则设定的质量以及分类结果的精度。本研究根据土地利用现状分类标准(GB/T 21010—2007)、海岛调查技术规程^[13]和遥感图像可判读的特点,建立了镆鵁岛土地利用类型的分类系统,包括林地、耕地、草地、居民地(含交通用地)、水域和未利用地共 6 类;并对典型地物进行采样,保证每类地物采样点不少于 150 个。

2.2 主要地物波谱统计特征

图像中主要地物的光谱特性能揭示和反映遥感数据内部及各波段间内在的规律性^[14]。对训练区内主要地物类型进行光谱特征的统计分析,包括各波段的最大值、最小值、均值、方差等(表 1)。

收稿日期: 2011-07-15; 修订日期: 2011-09-20

基金项目: 海洋公益性行业科研专项经费项目(编号: 200905004-7)、中国科学院海洋地质与环境重点实验室开放基金资助项目(编号: MGE2009KG04)和海洋沉积与环境地质国家海洋局重点实验室开放基金资助项目(编号: MASEG200807)共同资助。

表1 SPOT 图像中不同土地利用类型统计特征

Tab. 1 Statistical characteristics of different Land - use types in SPOT image

土地利 用类型	B1 ^①				B2				B3			
	最小值	最大值	均值	方差	最小值	最大值	均值	方差	最小值	最大值	均值	方差
林地	30	37	32.14	1.25	33	37	34.28	0.89	39	55	45.28	2.00
耕地	36	45	40.90	1.62	37	45	40.19	1.28	57	79	70.13	3.76
草地	32	35	33.53	0.67	35	37	36.28	0.51	39	43	41.19	0.80
居民地	37	59	45.41	3.79	38	75	49.35	5.28	42	71	52.97	4.04
水域	30	36	32.16	1.48	32	37	33.86	0.78	25	41	30.15	2.13
未利用地	53	68	63.69	2.67	62	79	73.81	3.15	65	78	73.71	2.37

①B1 - B3 分别表示 SPOT 1 - 3 波段图像。

分析表1可以看出,比较各波段图像上的均值, B1: 林地(32.14) < 水域(32.16) < 草地(33.53) < 耕地(40.90) < 居民地(45.41) < 未利用地(63.69); B2: 水域(33.86) < 林地(34.28) < 草地(36.28) < 耕地(40.19) < 居民地(49.35) < 未利用地(73.81); B3: 水域(30.15) < 草地(41.19) < 林地(45.28) < 居民地(52.97) < 耕地(70.13) < 未利用地(73.71)。6种地物在3个波段的采样点均值曲线如图1所示。

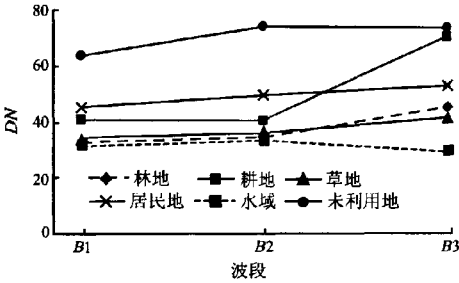


图1 不同地物类型3个波段采样点均值曲线

Fig.1 Mean curves of different object samplings of three bands

分析各波段光谱特征可发现一个明显的特征, 即水域和未利用地的 $B1 < B2$ 且 $B2 > B3$, 而其他地物均表现为 $B1 < B2$ 且 $B2 < B3$; 在3个波段中未利用地的 DN 均值在6种地物中都是最大的, 水域的 DN 均值在6种地物中都是最小的。因此, 利用这些光谱特征基本上可将未利用地及水域区分出来。林地和草地的光谱特征很相似, 但是在 $B1$ 和 $B2$ 波段, 林地的 DN 值小于草地的 DN 值, 在 $B3$ 波段, 林地的 DN 值大于草地的 DN 值。如表1所示, 6种土地利用类型样本 DN 值的最大值和最小值之间存在着交叠, 单纯利用上述3个波段并不能建立很好的分割原则将这些交叠在一起的地物类型区分开来。为增大不同地物类型之间的反差, 减小地物类型之间的交叠, 本文使用 SPOT 图像波段运算的方法提取了4个较为有效的分类特征, 分别是归一化差值植被指数 ($NDVI$)、 $B2 \cdot B3$ 、 $NDVI \cdot B2 \cdot B3$ 和 $\exp(NDVI)$, $NDVI$ 计算公式为

$$NDVI = (B3 - B2) / (B3 + B2)。$$
 (1)

4个分类特征的统计信息见表2。

表2 各土地利用类型 SPOT 图像波段运算统计特征

Tab. 2 Statistical characteristics of different land - use types from SPOT band operations

土地利 用类型	NDVI				B2 · B3			
	最小值	最大值	均值	方差	最小值	最大值	均值	方差
林地	0.07	0.20	0.14	0.02	1 287	2 035	1 552.30	83.83
耕地	0.18	0.32	0.27	0.02	2 257	3 318	2 821.06	208.88
草地	0.04	0.08	0.06	0.01	1 365	1 591	1 494.71	43.53
居民地	-0.03	0.11	0.04	0.03	1 596	5 325	2 633.11	478.15
水域	-0.15	0.05	-0.06	0.03	850	1 517	1 021.20	82.72
未利用地	-0.03	0.03	-0.000 49	0.01	4 092	6 084	5 447.14	386.66

土地利 用类型	NDVI · B2 · B3				exp(NDVI)			
	最小值	最大值	均值	方差	最小值	最大值	均值	方差
林地	90.82	398.15	214.95	43.72	1.08	1.22	1.15	0.03
耕地	399.59	1 014.60	766.31	103.19	1.20	1.34	1.31	0.03
草地	56.16	119.33	94.85	13.93	1.04	1.07	1.06	0.01
居民地	-145.89	242.96	89.46	57.01	0.97	1.12	1.04	0.03
水域	-129.66	77.80	-57.94	32.36	0.86	1.05	0.94	0.03
未利用地	-151.90	127.88	-4.84	57.69	0.97	1.03	1.00	0.01

分析表2可以看出,比较 $NDVI$ 图像上各地物类型的均值: 水域(-0.06) < 未利用地(-0.000 49) < 居民地(0.04) < 草地(0.06) < 林地(0.14) < 耕地(0.27); 比较 $B2 \cdot B3$ 图像上各地物类型的均

值: 水域(1 021.20) < 草地(1 494.71) < 林地(1 552.30) < 居民地(2 633.11) < 耕地(2 821.06) < 未利用地(5 447.14); 比较 $NDVI \cdot B2 \cdot B3$ 图像上各地物类型的均值: 水域(-57.94) < 未利用地(-4.84) < 居民地(89.46) < 草地(94.85) < 林地(214.95) < 耕地(766.31); 比较 $\exp(NDVI)$ 图像上的地物类型的均值: 水域(0.94) < 未利用地(1.00) < 居民地(1.04) < 草地(1.06) < 林地(1.15) < 耕地(1.31)。

2.3 决策树建立

决策树是利用树结构,按一定的分割原则把数据分为特征更为均质的子集,其基本思想是利用一组或多组自变量来预测每个样本最可能对应的类型。根据上面的统计分析结果制定分类原则,建立

分类决策树。水域在 SPOT 图像的 B3 波段最大 DN 值是 41,基本上可以将水域与其他地物进行区分;然后根据 $NDVI \cdot B3 \cdot B2 \leq 77.80$, $B1 < B2$ 且 $B2 > B3$ 区分出水域;再利用 $\exp(NDVI) \leq 1.07$ 将混淆在水域中的林地、草地区分开来; B3 波段中未利用地 DN 最小值为 65,利用 B3 波段又可以对未利用地与其他地物加以区分;未利用地中混杂的耕地和居民地可以根据 $NDVI \cdot B3 \cdot B2 \geq 399.59$ 来区分;而未利用地和居民地则根据 $B1 \geq 53$,未利用地的 DN 值有 $B1 < B2$ 且 $B2 > B3$ 的特点加以区分;剩余未分的土地利用类型中,居民地可根据 $B2 \geq 38$ 并且 $NDVI \cdot B3 \cdot B2 \leq 242.96$ 加以区分;耕地可根据 $B2 \cdot B3 \geq 2 257$ 加以区分;最后利用 $\exp(NDVI) \leq 1.07$ 区分林地和草地。分类决策树结构如图 2 所示。

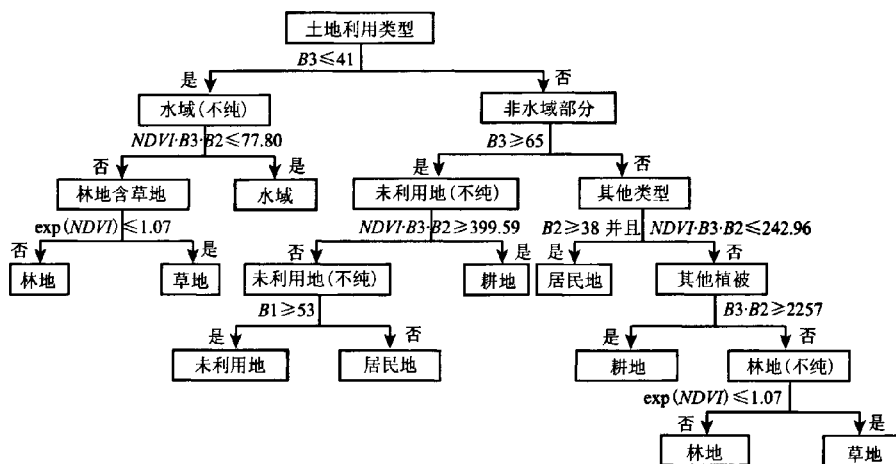


图 2 分类决策树结构

Fig. 2 Structure of decision - tree for classification

3 结果与分析

利用上文建立的决策树对研究区 SPOT 5 图像进

行分类,分类后利用计算机按土地利用类型随机生成 384 个检验点。通过目视解译的方法对 384 个检查点进行判读,采用误差矩阵和 Kappa 系数对决策树分类结果进行精度评价(计算出的误差矩阵见表 3)。

表 3 误差矩阵计算结果

Tab. 3 Result of error matrix

类别	未分类	林地	耕地	草地	居民地	水域	未利用地	行和	制图精度/%	用户精度/%	Kappa 系数
未分类	53	0	0	0	1	0	0	54	—	—	—
林地	0	55	3	2	1	6	1	68	87.30	80.88	0.771 3
耕地	0	0	63	0	7	0	1	71	91.30	88.73	0.862 6
草地	0	5	0	44	0	12	0	61	95.65	72.13	0.683 4
居民地	0	0	2	0	30	0	2	34	68.18	88.24	0.867 1
水域	0	3	0	0	0	41	0	44	69.49	93.18	0.919 4
未利用地	0	0	1	0	5	0	46	52	92.00	88.46	0.867 3
列和	53	63	69	46	44	59	50	384	总分类精度 = 86.46%		总 Kappa = 0.841 4

制图精度中草地最高(95.65%),居民地最低(68.18%);用户精度中水域最高(93.18%),草地最

低(72.13%),总体分类精度为 86.46%,总 Kappa 系数为 0.841 4。从整体上来看,耕地、未利用地和林

地分类精度相对较高,而草地、水域和居民地分类精度略低。

根据误差矩阵,并对比 SPOT 原始图像和实地调查数据,对 6 种地物的分类精度及其原因分析如下:

1)草地主要错分为林地和水域。其原因为草地在图像上的光谱特征与林地十分相似,容易产生错分。对比 SPOT 原始图像可以看到,草地与水域错分的情况主要集中在水陆交界处。水陆交界处的像元既包含水体信息,又包含部分植被信息,这种混合光谱现象导致了草地与水体的错分,也导致了水域与林地的错分。

2)居民地分类精度较低且与耕地和未利用地之间存在错分。主要是因为该岛屿耕地多以小斑块形式存在,斑块之间又有纵横交错的田间小路,而且这种道路在图像上多表现得细碎零散,在分类过程中易与相邻地物发生混淆。未利用地多分布在居民地附近或其中,二者的光谱特征相似,也是导致居民地与未利用地错分的原因。

3)林地除被错分为水域和草地外,还被错分为居民地、耕地和未利用地。主要是因为居民地内包含植被信息,混合像元存在带来的“异物同谱”现象造成了错分。

由于样本的选择是随机进行的,因此分析结果能在整体上反映分类的总体精度。但是误差矩阵法只是分类精度分析的一种方法,其分析结果与测试样本的选择密切相关。因此,这里的分析只是从一个方面或是局部地反映了决策树分类的精度情况。

为了进一步了解决策树分类法在海岛土地利用

类型分类调查中的相对优势、局限性及应用潜力,本文还利用相同的统计样本,分别使用平行六面体分类法、支持向量机分类法、最小距离分类法、马氏距离分类法和最大似然分类法分别对研究区 SPOT 5 图像(图 3(左))进行了监督分类,并利用上文的 384 个检验点进行了分类精度评价,分别从总分类精度和总 Kappa 系数两个方面对上述 6 种分类方法进行了比较(表 4)。

表 4 各种分类方法分类精度比较
Tab.4 Comparison of precisions among different methods of classification

分类方法	总分类精度/%	总 Kappa 系数
平行六面体	80.47	0.772 2
支持向量机	66.41	0.607 5
最小距离	63.02	0.569 3
马氏距离	60.16	0.533 6
最大似然	67.97	0.628 1
决策树	86.46	0.841 4

由表 4 可知,无论从总分类精度还是总 Kappa 系数的结果来看,决策树方法明显优于其他分类方法。可见,这种基于知识的决策树分类方法在海岛土地利用类型调查中具有更好的精度和利用空间。

研究中利用决策树分类法获取到镭铈岛土地利用类型专题图(图 3(右))。从分类结果中统计出镭铈岛居民地面积为 1.431 1 km²,耕地面积为 1.903 1 km²,林地面积为 0.827 5 km²,草地面积为 0.076 6 km²,水体面积为 0.064 8 km²,未利用地面积为 0.365 6 km²。

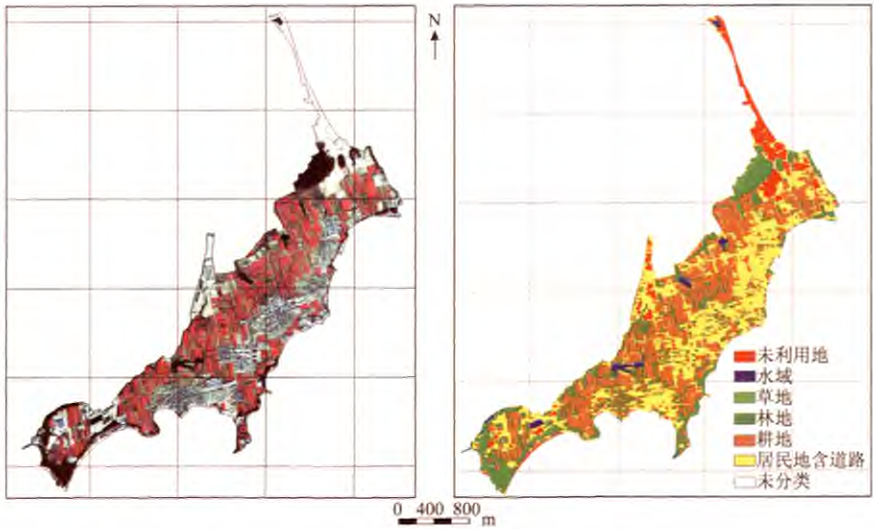


图 3 研究区 SPOT 图像(左)与决策树分类结果(右)
Fig.3 SPOT image(left) and the result of decision - tree classification (right) of study area

4 结 论

1) 基于高空间分辨率的 SPOT 5 多波段数据, 采用决策树分类方法进行馍鄒岛土地利用类型分类, 总分类精度达到 86.46%, 总 Kappa 系数为 0.841 4。

2) 在不同地物类型的区分过程中, 仅考虑 SPOT 图像波段本身信息建立的分割原则并不能完全将不同地物类型逐一区分; 而当引入波段运算统计数据后, 各土地利用类型的划分标准更容易选择且可区分性更好。

3) 对比决策树分类结果与平行六面体分类器、支持向量机分类器、最小距离分类器、马氏距离分类器和最大似然分类器的分类结果表明, 决策树分类法明显优于其他分类方法, 证明了决策树分类法在海岛土地利用调查中的潜力和发展空间。

参考文献:

- [1] 张耀光, 胡宜鸣. 关于做好海岛县土地利用总体规划工作的探讨——以辽宁长海县为例[J]. 中国土地科学, 1994, 8(6): 42-46.
- [2] 吕惠萍. SPOT 卫片在土地利用分类中的应用效果初探[J]. 地质地球化学, 1989(5): 53-55.
- [3] 吴均平, 毛志华, 陈建裕, 等. 一种基于分割图斑的海岸带遥感图像分类方法[J]. 海洋学研究, 2006, 24(2): 70-78.
- [4] Xu K M. The Analysis of SPOT 5 Characteristics on Land Cover Classification[J]. Science of Surveying and Mapping, 2004, 29(S1): 108-116.
- [5] 张华国, 黄韦良, 周长宝. 应用 IKONOS 卫星遥感图像监测南麂列岛土地覆盖状况[J]. 遥感技术与应用, 2003, 18(5): 306-312.
- [6] 杨存建, 周成虎. 基于知识的遥感图像分类方法的探讨[J]. 地理学与国土研究, 2000, 17(1): 72-77.
- [7] 李春华, 沙晋明. 基于知识的遥感影像的 BAYES 分类方法研究——以福州市琅岐岛土地覆盖/土地利用类型为例[J]. 水土保持研究, 2006, 13(6): 126-128, 132.
- [8] 刘小平, 彭晓晴, 艾 彬. 像元信息分解和决策树相结合的影像分类方法[J]. 地理与地理信息科学, 2004, 20(6): 35-39.
- [9] 刘勇洪, 牛 铮, 王长耀. 基于 MODIS 数据的决策树分类方法研究与应用[J]. 遥感学报, 2005, 9(4): 405-412.
- [10] 杜明义. 决策树方法在土地荒漠化分类中的应用研究[J]. 测绘科学, 2006, 31(2): 81-82, 86.
- [11] 黄 颖, 周云轩, 吴 稳, 等. 基于决策树模型的上海城市湿地遥感提取与分类[J]. 吉林大学学报: 地球科学版, 2009, 39(6): 1156-1162.
- [12] 朱宏波. 一种基于决策树的 SPOT-5 影像分类方法[J]. 电力勘测设计, 2010(3): 19-21, 29.
- [13] 国家海洋局 908 专项办公室. 海岛调查技术规程[M]. 北京: 海洋出版社, 2005: 59-62.
- [14] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003: 227-228.

Land - use Classification of Islands Based on Decision - tree Method

YANG Xi - guang^{1,2}, HUANG Hai - jun¹, YAN Li - wen¹, DU Ben - xu³

(1. Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China; 3. Dalian Forestry Bureau, Dalian 116023, China)

Abstract: With Moye island of Rongcheng city in Shandong province as the study area, the authors investigated the application of decision - tree method to land - use classification based on SPOT 5 satellite data. The results show that the decision - tree method is suitable for classification, with the mean precision and Kappa index being 86.46% and 0.8414 respectively. A comparison with the other traditional classification methods shows that the precision of decision - tree method is obviously higher, suggesting that the method of decision - tree classification has better applicable potential than the other methods in land - use classification research.

Key words: classification; land - use; decision - tree

第一作者简介: 杨曦光(1983 -), 男, 在读博士研究生, 主要从事海洋遥感与地理信息系统研究。E - mail: yangxiguang21@163.com。

(责任编辑: 李 瑜)