

基于决策树的龙口市土地利用/ 覆盖分类研究*

王大鹏,王周龙,崔青春,李德一
(鲁东大学 地理与资源管理学院,山东 烟台 264025)

摘 要 :LUCC 是全球变化研究的重要内容,土地利用/ 覆盖信息的获取是其前提和基础。决策树分类法是每次利用一定的规则,对遥感影像进行逐级细化的分类模型,具有直观、高效等特点。在综合分析每一种土地利用/ 覆盖类型各种特征的基础上,探讨将不同土地利用/ 覆盖类型的光谱特征、空间特征、波段比值特征等信息融入决策树分类模型,对龙口市土地利用/ 覆盖信息进行自动分类的方法,取得了较高的分类精度。
关键词 :遥感;决策树;土地利用/ 覆盖;自动分类
中图分类号 :F301. 24;TP79 文献标识码 :A 文章编号 :1005-3409(2008)01-0114-03

Land Use/ Cover Classification Based on Decision Tree in Longkou

WANG Da-peng, WANG Zhou-long, CUI Qing-chun, LI De-yi
(College of Geography & Resource Management, Ludong University, Yantai, Shandong 264025, China)

Abstract :Land use/ cover information is the precondition and foundation of LUCC (land use/ cover change) which is indispensable to globe change research. Decision tree classifier is a kind of classification model which uses certain rules to gradually thinning the image. It has intuitionistic and high-effect features. Classify Longkou land use/ cover on the basis of synthetically analyzing various characters of each kind of land use/ cover ,melting spectral characteristic ,spectral characteristic ,band ratio etc. into the decision tree classifier ,which has a good result.
Key words :remote sensing ;decision tree ;land use/ cover ;automatical classification

1 引 言

土地利用/ 覆盖变化(LUCC)是全球变化研究的重要内容,特别是进入 20 世纪 90 年代以后,随着国内外对其研究不断深入,逐渐成为全球变化研究的前沿和热点。同时其研究的范围亦由全球性研究不断地转向有针对性的生态环境脆弱带或人文因素高度复合带^[1]。土地利用/ 覆盖信息的获取是土地利用/ 覆盖变化研究的前提和基础,同时,及时有效地获取土地利用/ 覆盖信息对指导当地的经济、社会、环境发展和土地资源的合理利用及科学管理有着重要意义。利用遥感和地理信息系统技术获取土地利用/ 覆盖信息具有获取信息速度快、信息更新周期短、信息现势性强等特点。传统的监督分类和非监督分类单纯地利用像元的亮度特征对影像进行分割,分类精度低,往往不能满足实际应用的需要。随着人们对遥感认识的不断深入,新的分类方法不断提出,如纹理分析法、波段组合法、神经网络法、模糊分类法、植被指数法、决策树分类法等^[2-5]。
决策树分类法(Decision Tree Classifier)是以各像元的特征值为设定的基准位,按照一定的知识规则,分层逐次进行比较,从影像中分离目标地物的分类方法。一个决策树由一系列的节点和分支组成,每一分支有一个父节点,父节点经由一次属性判别生成 2 个或多个子节点,父节点和子节点之间形成分支,每一个子节点又可以作为父节点生成下一级子节点。最初一级父节点称为根节点,最末一级子节点称为叶节点,叶节点包含决策结果。在遥感影像分类中,根节点

即为分类影像,叶节点为分类目标地物。由于用决策树方法对遥感影像进行分类,可以方便融合影像以外的各种知识,且不需要任何先验统计假设条件,因而在遥感信息提取和影像分类中有着广泛的应用^[5-9]。以龙口市为例研究将各种地物的光谱知识、空间知识以及其他相关知识融入决策树,对土地利用/ 覆盖类型进行计算机自动分类的方法。

2 研究区概况

龙口市(东经 120°22′ - 120°75′,北纬 37°46′ - 37°76′)地处胶东半岛西北部,渤海湾南岸,是一座新兴的沿海开放港口城市,东临烟台,南接青岛,西与潍坊毗邻,东北与天津、大连、秦皇岛、北戴河等名城以及朝鲜半岛隔海相望,是中国环渤海经济区中最具发展活力的地区之一。龙口市地势南高北低,西北部为西南向东北倾斜的滨海平原,地势平坦,灌溉条件良好,农田、果园广布。东南部为低山丘陵,地势较高处分布着大面积的森林,山间平地 and 山谷里亦有果园和盖度极低的旱田分布。农田里主要作物为小麦,由于干旱,旱田里的小麦已经收割,使旱田呈裸露状态;而水浇地内的小麦大部分还未收割。

3 数据源及数据处理

采用龙口市 2004 年 6 月 5 日 TM 影像一景,有 TM1 - TM5, TM7 等 6 个波段,影像质量良好,无云,无坏线。同一时期研究区 SPOT - 5 全色波段影像,研究区 1 : 5 万地形

* 收稿日期:2006-10-26
基金项目:山东省自然科学基金(Z2003E01);鲁东大学基金项目
作者简介:王大鹏(1981 -),男,山东台儿庄人,硕士研究生,主要从事资源环境遥感研究。

图,龙口市 1996 年土地利用现状图及其相关图文数据。

校正 1 5 万数字化地形图,使之具有正确的地理坐标。采用二次多项式的方法,通过采集地面控制点实现影像的几何精校正。由于遥感影像的校正精度直接影响到多源数据综合应用的效果,因而应注重地面控制点的数量和精度,并应均匀分布于整个研究区,在变形较大的地区要适当增加控制点的数量和密度。该研究地面控制点从数字地形图和 GPS 实测获得。重采样方法为最邻近像元法。实地考察检验校正精度,校正误差在 0.5 个像元以内,满足精度要求。矢量化地形图等高线,生成 DEM 和坡度图。

4 土地利用/覆盖信息提取

4.1 土地利用/覆盖类型划分

土地利用与土地覆盖有一定的对应关系,但土地利用并不等同于土地覆盖。土地覆盖是遥感影像所直接反映的地表的客观存在,侧重于地物的自然属性;土地利用则通过各种土地覆盖的组合、结构和模式间接获取,反映人地关系,侧重于地物的人文社会属性^[10-11]。目前,国际上还没有土地利用/覆盖分类的统一标准,主要有国际地图和生物圈计划(IGBP)标准、美国国家地质局(USGS)标准、Olson 世界生态系统综合体体系、中国国家土地管理局的城镇土地分类系统等^[12]。为了便于进一步研究土地利用/覆盖变化,结合历史土地利用图,将研究区土地利用/覆盖类型划分为建设用地(包括居民地、工矿用地和道路)、水浇地、旱田、果园、林地、滩涂、水域、干河床/海滩等 8 类。

4.2 土地覆盖类型特征分析

根据历史土地利用现状图和野外实地调查的结果,在影像中分别给各种土地利用/覆盖类型选择一定数量的训练样本,测定各种土地利用/覆盖类型在各个波段的亮度范围,并据此生成各土地利用/覆盖类型光谱响应曲线图(图 1)。

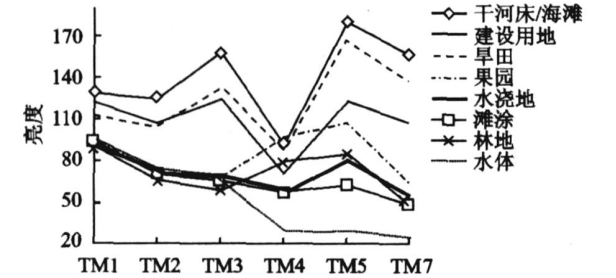


图 1 各土地利用/覆盖类型光谱响应曲线

水体同陆地的差别比较明显,亮度从 TM1 到 TM7 整体呈下降趋势,在 TM4, TM5, TM7 远低于其它地物。植被同非植被的区别也很明显,没有植被覆盖的地物 TM3 远大于 TM4,而有植被覆盖的地物则不具备此特征。干河床/海滩、旱田和建设用地等 3 种没有植被覆盖的土地利用/覆盖类型有着相似的光谱特征,虽然建设用地的亮度平均值在 TM5 略低于旱田和干河床/海滩,但三者整体亮度范围存在较大重叠(表 1),不易区分。

表 1 建设用地、干河床/海滩、旱田像元亮度比较

地 物	最小值	最大值	均 值
建设用地	84	179	122.638
干河床/海滩	125	210	180.065
旱 田	115	202	164.983

K-T 变换(缨帽变换)是一种经验性的多波段图像的线性变换,把 TM 数据的基本结构用三维空间中的一个植被平面、与植被平面垂直的土壤平面以及介于 2 个平面之间的过渡带来表示。TM 影像作 K-T 变换后产生 6 个波段,依次为亮度、绿度、湿度和第四、五、六分量。对影像作 K-T 变换,发现建设用地从绿度到湿度的变化趋势同旱田、干河床/海滩相反,同时,旱田从绿度分量到第四分量的变化幅度要大于干河床/海滩(图 2),可以据此对 3 者进行分类。

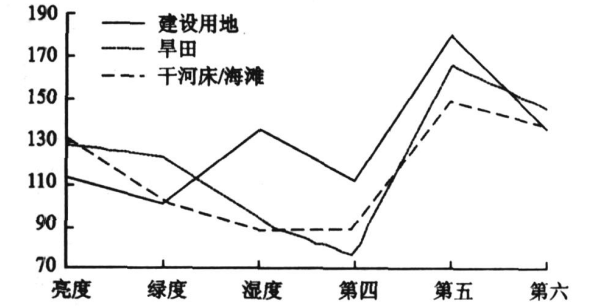


图 2 建设用地、旱田和干河床/海滩在 K-T 变换后各分量上的数值

果园与林地各波段的亮度范围及变化节律都极其相似,仅利用光谱特征很难区分。NDVI(归一化植被指数, $(NIR - R)/(NIR + R)$)利用波段间的比值,可以消除地形和阴影的影响,真实地反映地表植被覆盖的情况,是一种较为常用的区分植被信息的辅助数据。然而研究区内平原地区果园同林地 NDVI 值差别不大。在丘陵地区土地利用信息提取中,地形数据是较为理想的辅助数据^[13],研究区内的林地主要分布在南部高度大于 100 m 的山地,而在这一高度果园的盖度明显低于林地(表 2),因而可以结合 DEM 和 NDVI 对果园和林地进行区分。

表 2 不同海拔果园同林地的 NDVI 值比较

地 物	最小值	最大值	均 值
山地果园	- 0.091	0.143	0.057
平原果园	- 0.011	0.329	0.163
林 地	0.132	0.347	0.271

4.3 基于决策树方法的土地利用/覆盖信息提取

根据以上分析,建立决策树分类模型:

IF TM4 + TM5 > TM3 THEN 陆地;

IF TM3 > 82 AND TM3/ TM4 > 1.3 THEN 非植被;

IF 湿度/ 绿度 > 1 THEN 建设用地;

ELSE IF 绿度/ 第四分量 > 1.45 THEN 旱田;

ELSE 干河床/ 海滩;

END IF;

END IF;

ELSE 植被;

IF TM4 > TM3 THEN 果园或林地;

IF EDM > 100 AND NDVI > 0.11 THEN 林地;

ELSE 果园;

END IF;

ELSE IF TM5/ TM4 > 1.2 THEN 水浇地;

ELSE 滩涂;

END IF;

END IF;

END IF;
ELSE 水体、阴影;
END IF

运行模型发现有部分水浇地、滩涂和山体阴影被误分为水体。众所周知,水面是没有植被覆盖的,而阴影、水浇地和滩涂表面则有植被覆盖,因而在水体分类条件中加入 NDVI

信息可以将多提取的水浇地剔除:
IF 水体:IF NDVI> - 0.3 THEN 水体;
ELSE 非水体 END IF
被剔除的非水体像元返回到陆地这一节点重新分类。
分类流程如图 3。

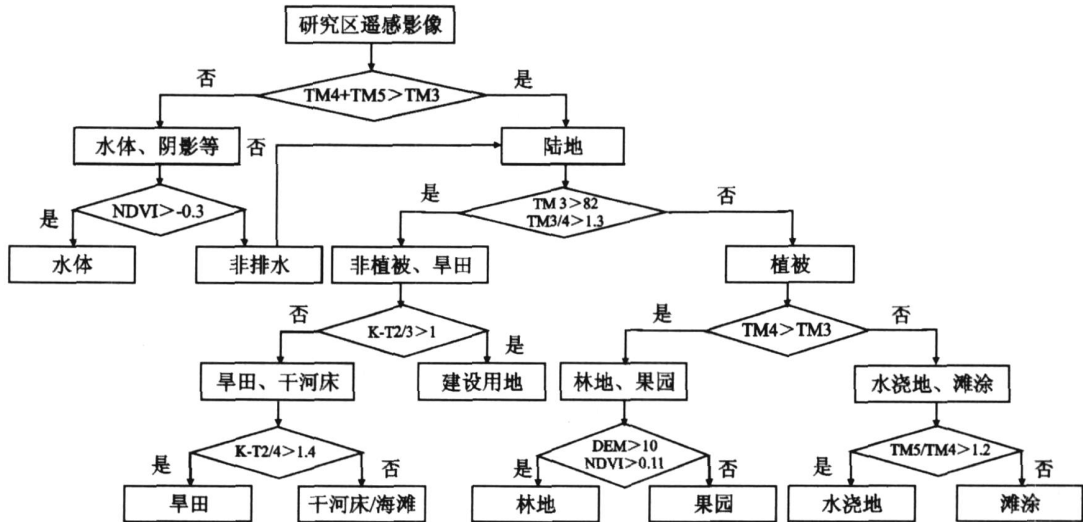


图 3 土地利用/覆盖自动分类决策树

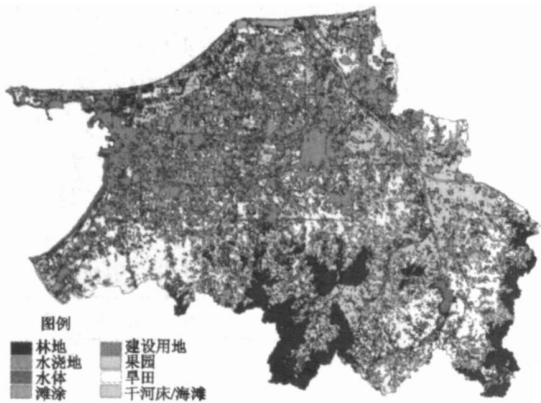


图 4 龙口土地利用/覆盖分类结果

表 3 土地利用/覆盖自动分类精度评价

地 物	参考数	分类数	正确 分类	生产 精度/ %	用户 精度/ %
干河床/海滩	50	43	39	78.00	90.70
建设用地	135	137	122	90.37	89.05
滩 涂	23	14	13	56.52	92.86
旱 田	208	227	192	92.31	84.58
水 浇 地	104	107	97	93.27	90.65
水 体	27	30	25	92.59	83.33
林 地	106	100	88	83.02	88.00
果 园	197	192	170	86.29	88.54
总 数	850	850	746	-	-
总精度/ %					87.76

4.4 精度评价

由于精度评价是以像元为单位,因此为了保证评价结果的真实性,在评价之前不再对分类结果作任何降噪处理。将分类影像同 SPOT - 5 全色波段融合,目视解译融合影像,以

目视解译结果为参考数据,采用随机方法生成一定数目评价样本像素,通过比较试验样本的分类数据和参考数据而计算出误差矩阵,由误差矩阵最终得到 Kappa 系数,Kappa 系数的大小反映了分类的精度。在随机抽取的 850 个样本像元中共 87 个被错分,总分类精度达到 87.76 % (表 3),满足实际应用的要求。

5 结 论

通过精度评价的结果可以看出,建设用地、水浇地的提取精度较高,而果园和林地的提取精度则要低一些,滩涂被漏提最多。遥感影像的空间分辨率总是有一定的限度的,尺寸小于像元空间分辨率的地物在影像上将以混合像元的形式出现。从原始影像中找到错分的样本像元发现,错分的像元大部分为混合像元或地物的边缘地带。

综上所述,在决策树分类模型中综合应用各种地物的光谱特征、植被指数、空间特征以及对影像进行变换产生的派生特征,对龙口市的土地利用/覆盖信息进行自动提取,由于决策树分类法在每次分类中的运算几乎都是简单的算术运算和比较大小,且只对一种地物进行识别而对非目标地物进行淹没,从而在很大程度上避免了非目标地物的干扰,取得了较好的分类效果。混合像元的存在从一定程度上影响了分类精度,如在分类模型中加入对混合像元的分类条件,则可以进一步提高分类精度。

参考文献:

[1] 巴雅尔,王建华,祁元. 基于 GIS 的遥感影像土地利用与覆盖信息提取研究[J]. 干旱区资源与环境,2004,3 (2):106-109.
[2] 李弼程,彭天强,彭波,等. 智能图像处理技术[M]. 北京:电子工业出版社,2004.

(下转第 121 页)

表 10 不同密度油松林产流量和产沙量				
样地号	林分密度/ (株·hm ⁻²)	降雨强度/ (mm·min ⁻¹)	产流量/ mm	产沙量/ (t·km ⁻²)
阳 1	800	2.55	24.11	55.36
阳 2	1000	2.54	21.56	42.35
阳 3	1200	2.57	19.51	31.75
阳 4	1500	2.47	16.33	20.78
阳 5	2000	2.50	12.57	11.23

4 讨 论

在密度为 800 株/hm² 的阳坡油松林地内,林分密度通过改变林分内土壤、林分内温湿度、光照强度及其改变林内土壤营养元素分配等小环境条件,来限制林下植物种类、数量等。林木生长状况良好,林下植被层各种生物多样性指数较其它林分高,并伴随着出现辽东栎等其它栎类更树种的生长,这说明该林分土壤、光照、水分、温湿度等各种环境条件较为适宜,已经能够满足阔叶林树种的生长。

油松林 1 200 株/hm² 时林分蓄积量最大为 95.53 m³/hm²。但从单株蓄积量来看,不同密度的油松林单株蓄积量变动范围在 0.032 1~0.1017 m³/hm²,且随着油松林密度的减小而增大,油松 800 株/hm² 的单株蓄积是 2 000 株/hm² 单株蓄积的 3.17 倍,可见,低密度林分促进了林木个体的生长发育。

不同密度油松林生态系统碳密度范围为 78.87~142.13 t/hm²,平均碳密度为 119.14 t/hm²,且随着密度增大而减小。油松林碳密度主要有 3 个部分组成:植被层、枯落物层和土壤层,其空间分布为土壤层>植被层>枯落物层。植被层碳密度平均为 29.57 t/hm²,占总碳密度的 24.82%;枯落物的变动在 8.63~11.12 t/hm²,占 6.07%~14.10%;林地土壤的碳密度平均为 79 t/hm² 以上,地上部分碳密度与地下(包括土壤、树根和死地被物)碳密度之比平均为 1/3.30。

稳渗速率为 0.098 mm/min,是密度为 2 000 株/hm² 时

稳渗速率的 2.51 倍;产流量随着林分密度的增大而减小,当阳坡油松密度为 800 株/hm² 时,其产流量最大为 24.11 mm。可见,当密度为 800 株/hm² 对增加当地区的径流量具有重大意义。产沙量也随密度的增加而减少,当油松密度为 800 株/hm² 时产沙量最大为 55.36 t/km²。从改良土壤渗透性能、增加该地区森林水源涵养能力的角度来说,32 a 油松林的林分密度应在 800 株/hm² 为宜。

总体来说,32 a 油松林密度为 800 株/hm² 时,生物多样性、碳密度、单株蓄积量、水源涵养功能最大。为此,对于 32 a 阳坡油松健康经营的角度来说,间伐后 800 株/hm² 为首选密度。

参考文献:

[1] 余新晓,于志民.水源保护林培育、经营、管理、评价[M].北京:中国林业出版社,2001.

[2] 刘国华,傅伯杰,方精云.中国森林动态及其对全球碳平衡的贡献[J].生态学报,2000,20(5):734-739.

[3] 廖福霖,陈光水,谢锦升,等.栽杉留阔模式群落结构及其多样性研究[J].福建林学院学报,2000,20(4):329-333.

[4] 邹双全,何宗明,陈光水,等.老杉木人工林与松阔混交林群落多样性比较[J].福建林学院学报,2002,22(2):124-128.

[5] 项文化,湛小勇,蔡宝玉.湿地松人工林生产量的时空特征[C]//刘焯章.森林生态系统定位研究.北京:中国林业出版社,1993.

[6] 中国土壤学会农业化学专业委员会.土壤农业化学常规分析方法[M].北京:科学出版社,1984.

[7] Brown S. Present and potential roles of forests in the global climate change debate[J]. Unsylva,1996,185:3-10.

[8] 王树森.华北土石山区基于森林植被演替规律的森林健康研究[D].北京:北京林业大学,2005.

(上接第 116 页)

[3] 姜青香,刘慧平.利用纹理分析方法提取 TM 图像信息[J].遥感学报,2004,9(5):458-464.

[4] Huete A,Didan K,Miura T,et al. Overview of the radiometric and biophysical performance of the MODIS vegetation index[J]. Remote Sensing of Environment, 2002,83:195-213.

[5] 陈华丽,陈植华,丁国平.用基于知识的决策树方法分层提取矿区土地类型:以湖北大冶为例[J].国土资源遥感,2004,9(3):49-53.

[6] John D,蔡竟峰,蔡自兴.决策树技术及其当前研究方向[J].控制工程,2005,12(1):15-8.

[7] 李爽,张二勋.基于决策树的遥感影像分类方法研究[J].地域研究与开发,2003,2(1):17-21.

[8] Fried M A,Brodeley C E. Decision tree classification land cover from remotely sensed data [J]. Remote Sensing & Environ.,1997,61:399-409.

[9] 孙家丙,舒宁,关泽群.遥感原理、方法和应用[M].北京:测绘出版社,1997.

[10] 史培军,宫鹏,李晓冰,等.土地利用/覆盖变化研究的方法与实践[M].北京:科学出版社,2000.

[11] Clawson M,Stewart C L. Land use information ,a critical survey of US statistics including possibilities for greater uniformity ,resource for the future[M]. Baltimore :John Hopkins University Press ,1965.

[12] 钱乐祥.遥感数字影像处理与地理特征提取[M].北京:科学出版社,2004.

[13] 王周龙,冯学智,刘晓玫,等.秦淮河丘陵地区土地利用遥感信息提取及制图[J].遥感学报,2003,7(2):131-135.