



邓镇坤,张鹏林.基于 CART 决策树的黄河源区域生态地类的遥感解译研究[J].黑龙江农业科学,2022(2):24-29,35.

基于 CART 决策树的黄河源区域生态地类的遥感解译研究

邓镇坤¹,张鹏林²

(1. 西藏大学 理学院,西藏 拉萨 850000;2. 武汉大学 遥感信息工程学院,湖北 武汉 430079)

摘要:为指导黄河源区域生态环境修复,以黄河源区域为研究区,采用 Landsat 8 影像作为数据源,以融合多种特征指数的 CART 树模型进行高寒湿地分类方法研究。结果表明:使用 CART 决策树算法进行数据挖掘处理,将得到的决策树用来分类制图,得到黄河源地区生态地类分布图,通过混淆矩阵得出分类后总精度为 88.25%,Kappa 系数为 0.834 5,而使用同样分类样本的监督分类方法中使用范围较广的最大似然法分类得到的总体精度为 84.90%,Kappa 系数为 0.788 8,分别低于 CART 决策树分类 3.35 个百分点与 0.045 7,并且各地类的精度均低于 CART 树分类方法,证明本研究构建的决策树分类模型适用于研究区的生态地类提取。

关键词:黄河源;CART 决策树;遥感影像分类;机器学习

我国生物多样性的基础纽带之一便是高寒湿地,三江源国家公园黄河源园区是高寒湿地生态系统的典型代表,黄河源地区范围广,横跨四川、甘肃与青海三地,但同时也由于海拔高且落差大导致了地形地貌的复杂多变,自然灾害频发,生态环境脆弱,而且受到夏季风的影响较为强烈,导致气候也特别敏感,所以黄河源地区是我国重要的生态涵养地区之一^[1]。近年来随着城市化进程的不断加快,全球温度不断升高,人口总数激增,高寒湿地正在面临着不断退化的危险,使得生态安全遭受到了难以想象的破坏,因此迫切需要采用更加成熟的技术和方法来对高寒湿地进行高效并且长期的监测^[2],这对科学管理土地资源和实现环境的可持续发展具有重要的现实意义。

采用遥感方法对湿地进行研究已经有几十年的历史,随着遥感图像分辨率的不断优化,湿地土地覆盖的研究向着更加细化的方向发展^[3]。分类的方法是完成这项研究最重要的因素,主要包括传统的人工目视解译和计算机自动分类,人工目视解译精度高但是效率低,任务量大^[4]。现在一

般采用计算机进行自动分类,应用较多的是监督分类方法,其中支持向量机、神经网络法以及最大似然法分类都是比较常见的。Laba 等^[5]提取了蒂沃利海湾的湿地植被,开采用最大似然法分类。张彦忠等^[6]利用 TM 数据,采用最小距离法对桔子之乡黄岩市的土地利用进行分类,充分揭示了桔园的分布。罗小波等^[7]以安徽寿县附近作为研究区域,基于改进后的 Kohonen 神经网络法进行分类试验,结果表明分类精度优于最大似然法。但是这种方法由于容易出现同谱异物或者同物异谱的情况,导致了错分漏分的情况。主要是因为这种方法只侧重于图像的光谱特征,而对于图像的纹理特征以及高程数据、坡度坡向等辅助数据没有实现很好地利用。

决策树分类方法(decision tree classifier)是通过获取各个像元的特征值,包括光谱特征以及纹理特征,以特定的知识和规则来进行分层并依次比较,从而把地物种类从影像中分离出来的方法^[8],当遥感影像的空间布局 and 分布比较难以分辨时,决策树方法就具有结构清晰、便于理解、达成简单、精度高、出结果快的优点^[9-10]。

常见的决策树方法有基于专家知识人工提取阈值的决策树法、C4.5、C5.0、和 CART 决策树法,CART 决策树方法的应用范围更加广泛,既可以用于分类又可用于回归,且计算更为简洁。高凯等^[11]基于 Landsat 数据,利用专家知识的决

收稿日期:2021-11-25

基金项目:国家重点研发计划(2018YFF0215006);安徽省自然资源调查监测评价体系研究(2021-K-1)。

第一作者:邓镇坤(1997—),男,硕士研究生,从事生态遥感监测研究。E-mail:942983228@qq.com。

通信作者:张鹏林(1970—),男,博士,教授,从事计算机视觉研究。E-mail:47437044@qq.com。

策树方法对武汉市的景观进行了分类并制图。邹文涛等^[12]基于 SPOT 影像数据提取了它的纹理特征以及其他辅助数据并选用 QUEST 决策树的方法对高寒湿地的植被进行分类。侍昊等^[13]基于随机森林模型,采用 Landsat 8 的影像数据并提取了图像不同的光谱指数以及图像变换后的其他指数来构建特征变量,最终提取了水生植物的空间分布信息,结果精度较高。针对研究区域地理位置的特殊性以及气候条件恶劣导致的生态环境脆弱的问题,本文以遥感解译的方法来完成对研究区生态地类分布的监测,主要以融合了 DEM 参数的 Landsat 8 影像为基础,再综合提取影像的光谱特征指数与纹理特征指数,利用 CART 模型训练生成决策树并分类得到研究区的生态地类分布图,与传统监督分类中的最大似然法分类进行精度相比,本文选用的研究方法有着明显的优势,此次研究旨在提高分类的精度,以探讨该方法

在实际运用中的可行性。分类的精度越高,影像在之后的定性定量分析以及专题信息获取时的精度与质量就会越高,可为今后的科研工作奠定坚实的基础。

1 研究区域概况及数据来源

1.1 研究区域概况及自然资源条件

图 1 使用的影像位于果洛州玛多县境内的黄河源园区,位于 33° 55′ 5″ N ~ 35° 28′ 15″ N, 97° 1′ 20″ E ~ 99° 14′ 57″ E,扎陵湖、鄂陵湖和星星海在内的两个保护分区位于其中,占地 19 100 km²,包括扎陵湖乡、玛坟墓差查里镇及玛多县黄河乡等 19 个行政村。主要气候特征为雨热同期出现,夏季短冬季长,太阳直射时间长,每年的温差较小但是日较差大,年均温度为-5.6~7.8℃,园区的平均海拔高度为 4 200 m 左右区域的植被类型较简单,以高寒草甸、草原为主^[14-15]。

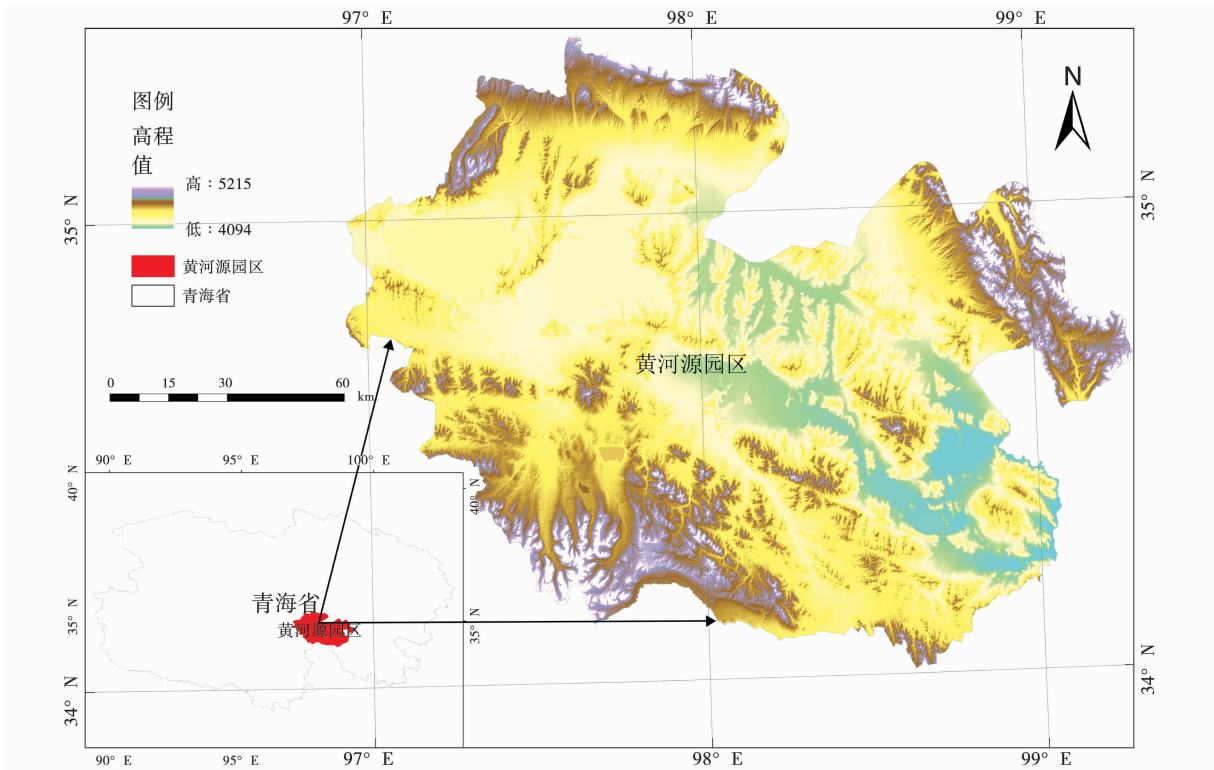


图 1 黄河源地理位置

1.2 数据来源

研究所采用的遥感影像数据为在美国地质勘测局官网 (<http://glovis.usgs.gov/>) 所获取的 3 张 Landsat 8 影像,选取的月份是云较少的 8—

10 月的影像,分别是 2020 年 8 月 25 日轨道编号为 133/36、2020 年 9 月 17 日轨道编号为 134/36、2020 年 10 月 10 日轨道编号为 135/36、云量均小于 10%,全色波段空间分辨率为 15 m,多光谱波

段分辨率为 30 m。详细数据如表 1 所示。

表 1 Landsat 8 OLI 影像数据

成像时间	轨道号	波段	波段名称	波长范围/nm
2020-08-25	133/36	B2-B7	B1-Coastal	0.43~0.45
			B2-Blue	0.45~0.51
			B3-Green	0.53~0.59
2020-09-17	134/36	B2-B7	B4-Red	0.64~0.67
			B5-NIR	0.85~0.88
			B6-SWIR1	1.57~1.65
2020-10-10	135/36	B2-B7	B7-SWIR2	2.11~2.29
			B8-PAN	0.50~0.68
			B9-Cirrus	1.36~1.38

2 影像数据处理

2.1 数据预处理

将采集到的影像进行预处理,由于 Landsat 8 影像数据已经进行了辐射定标与几何校正,所以只需要进行大气校正、数据的镶嵌融合以及裁剪等预处理操作。搜集覆盖研究区的1:10万地形图,以 30 m DEM 数据作辅助数据^[16]。

2.2 生态地类确定

通过对试验地区的调查取证与资料分析,根据 LUCC 分类体系与实地情况确定相应的土地利用分类系统,建立解译标志库^[17],将试验区域的生态地类划分为 6 种,分别是草地、林地、水体、滩地、裸土地与裸岩石砾地(表 2)。

表 2 黄河源区土地覆被分类系统

编号	类型	分类标准
1	林地	指生长乔木、灌木、竹类等林业用地
2	草地	指以生长草本植物为主,覆盖度为 5% 的各类草地,包括以牧草为主的灌丛草地和郁闭度在 10% 以下的草地
3	水体	指天然陆地水域和水利设施用地
4	黑土滩	高寒草甸草场严重退化后形成的一种大面积黑色裸地景观
5	裸土地	指地表土质覆盖,植被覆盖度低于 5% 的土地
6	裸岩石砾地	指地表为岩石或石砾,并且覆盖度大于 5% 的土地

2.3 样本选取

为了切实掌握研究区的实际土地覆被类型与

分布特征,综合实地调查数据与谷歌地球高分辨率影像、Landsat 8 影像数据,总共选取了 698 个样本点,并从中抽取 30% 作为验证样本,每类样本点详见表 3,样本点的空间分布如图 2 所示。

表 3 样本数量统计

类别	草地	林地	水体	黑土滩	裸土地	裸岩石砾地
样点数	148	116	170	104	92	68

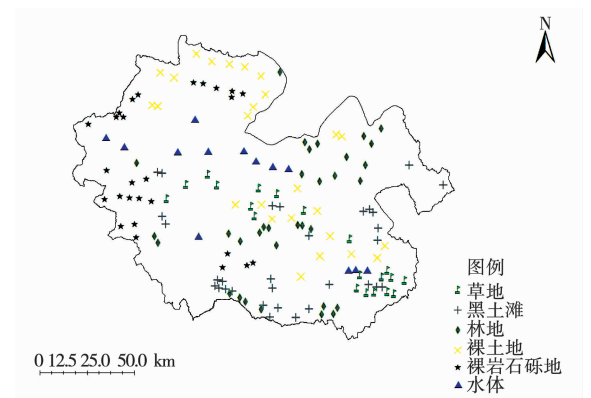


图 2 样点空间分布图

3 研究方法

3.1 模型选择

利用 CART 决策树分类方法建立分类模型,对 Landsat 8 影像进行土地覆被分类,与传统的基于专家知识的决策树分类方法不同,CART 决策树分类方法不需要反复试验确定分类阈值,可以自动确定阈值,使得人为的因素减少,降低了分类的不确定性,提高了分类的精度^[18]。CART 决策树算法使用基尼系数来选择特征,基尼系数所代表的意思便是模型的不纯度,它基于 0~1 之间的数,值越小代表样本集合纯净度越高,反之越大则代表样本集合越杂乱。概括来说,基尼系数反映了从数据集中随机抽出两个样本不一致的概率,在不考虑最终的剪枝情况下,得到数据集中某个特征所有取值的基尼指数之后,就可以得到该特征的基尼增益系数,模型的建立过程就是选择系数最小的节点作为分叉点直到最后一个节点不能再分支下去,得到的数据集都属于同一类。对于一个数据集 T ,将某一个特征作为分支,那么第 i 个取值的基尼系数的计算公式见式(1)^[19]。

$$Gini(T_i) = 1 - \sum_{j=1}^n p_j^2 \quad (1)$$

式中: n 表示数据类别数, p_j 表示第 j 个类别的概率, 计算出某种特征所有取值的基尼指数后得到基尼增益系数, 计算公式见式(2)^[19]。

$$Gini_{split}(T) = \sum_{i=1}^2 \frac{N_i}{N} gini(T_i)$$

(2)

式中: N 表示样本总数, N_i 表示属于特征第 i 个属性的样本数。由于在它的模型建立过程中, 训练数据不可避免的会出现一些噪音或者孤立点, 会使得分类的准确性不高, 所以通过预剪枝和后剪枝的方式来去掉噪音和孤立点, 预剪枝的原理是及早地停止树的生长, 后剪枝是在已经生成好的树上删除节点分支。

3.2 分类特征变量选取

3.2.1 光谱特征变量 各类地物由于自身组成与成分的不同, 对光的反射各不相同, 这种现象就成为遥感图像分类的重要依据, 因此本研究选用 Landsat 8 融合后的 Band 1 到 Band 7 总共 7 个波段的光谱数据。整个图像归一化处理后得到归一化植被指数 NDVI, NDVI 的计算公式见式(3)。

$$NDVI = (NIR - R)/(NIR + R)$$

(3)

式中: R 、 NIR 分别表示红波段与近红外波段的反射率^[20]。

3.2.2 纹理特征分量 将图像数据进行主成分分析和降维, 再选择处理后图像的第一主成分作为数据源, 采用灰度共生矩阵算法计算纹理特征, 计算方法见式(4)。

$$H(s, t, \theta, G) = \{(p, q), (p + G, q + G) \in N \cdot N \mid f(p, q) = s, f(p + G, q + G) = t\}$$

(4)

式中: H 代表灰度共生矩阵, (p, q) 代表任意一点的像素坐标, θ 、 G 代表像素点偏移的角度和距离; s 与 t 表示发生偏移前后处像素点的灰度, $f(p, q)$ 表示 (p, q) 的灰度, c 代表移动步长为 1, 在提取图像纹理特征时并不直接使用灰度共生矩阵, 而是在矩阵的基础上统计特征值, 所以本试验采取的纹理特征变量及其公式如表 4 所示^[20]。

3.2.3 其他分量 除了研究区影像数据外, 还在地理空间数据云官网收集了遥感影像的 DEM 数据辅助分类。

表 4 纹理特征变量及公式

纹理特征变量	公式
均值(ME)	$Mean = \sum_{s=1}^n \sum_{t=1}^n l(s, t) \cdot s$
方差(VA)	$Variance = \sum_s \sum_t (s - u)^2 \cdot l(s, t)$
协同性(HO)	$Homogeneity = \sum_s \sum_t \frac{1}{1 + (s - t)^2} \cdot l(s, t)$
对比度(CO)	$Contrast = \sum_s \sum_t (s - t)^2 \cdot l(s, t)$
相异性(DI)	$Dissimilarity = \sum_s \sum_t l(s, t) \cdot s - t $
信息熵(ENT)	$Entropy = - \sum_s \sum_t l(s, t) \cdot \lg[l(s, t)]$
二阶矩(SM)	$Asm = \sum_s \sum_t \{l(s, t)\}^2$
相关性(COR)	$Correlation = \sum_s \sum_t (s, t) \cdot l(s, t) - u_p u_q / \delta_p \delta_q$

注: 表中 u 代表 $l(s, t)$ 的均值; $\delta_p \delta_q$ 为 $l(s, t)$ 的相关系数; n 值是灰度值的阶数; $l(s, t)$ 是 $n \times n$ 的归一化共生矩阵。

4 结果与分析

4.1 训练区样本分析与决策树构建

分离度是评价样本选择是否合理的重要参考因素, 只有分离度达到了要求才能合理有效地进行分类研究, 一般来说地物间分离度达到了 1.8 以上便可使用这组训练样本^[20]。影像中部分分离度如表 5 所示。

表 5 部分地类分离度

土地类型	分离度
草地与裸土地	1.976
草地与灌木	1.889
草地与裸岩地	1.975
草地与黑土滩	1.996
草地与水体	1.997
裸土地与林地	1.998
裸土地与裸岩地	1.842
裸土地与水体	1.998
裸土地与黑土滩	1.979
裸土地与草地	1.976
林地与水体	1.995
林地与裸土地	1.998
林地与草地	1.889
林地与裸岩地	1.998
林地与黑土滩	1.801

共有 698 个样本作为训练样本与验证样本，其中验证样本占比 30%，通过 ENVI 5.1 软件建立基于 CART 的决策树。

特征指数中，归一化植被指数的加入使得植被与其他地物得以区分开来，ISODATE 非监督分类是一种不需要先验知识的分类方法，能够为最终的分类结果提供参考，纹理特征中均值能区分长条形态的河流、河边滩地与草地；纹理特征中的信息熵使裸岩石砾地和草地更容易区分；高程 DEM 数据的加入对草地和林地的区分起着积极的作用^[2]。

4.2 分类结果

运行上述决策树提取得到遥感影像的生态地类初步分布图，然后继续进行重编码、聚类 and 去除分析等分类后处理操作，目的是剔除小的图斑碎斑块，将小的同类型的多边形进行合并处理，使得分类效果更好，得到黄河源区域最终的生态地类分布图(图 3)。

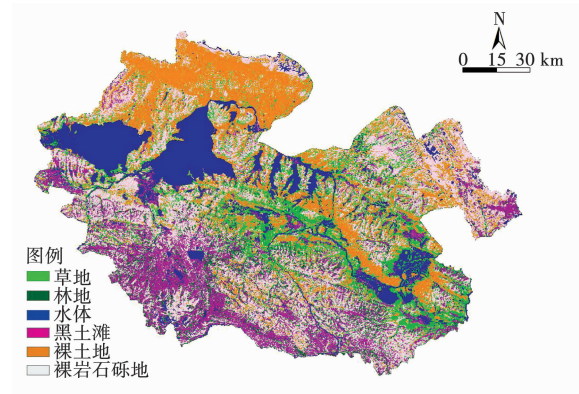


图 3 基于 CART 决策树的黄河源区生态地类分布图

4.3 精度验证与对比

为了检验本文基于 CART 决策树的分类方

法的精度，以构建混淆矩阵的方法进行验证，混淆矩阵可在一个表中展示分类后模型的归错类及归对类观测值的个数和百分比。主要的评价指标是以分类精度与 Kappa 系数来完成的。Kappa 系数是基于混淆矩阵得以实现的，它是一种多元的离散型方法，计算方法见式(5)。

$$Kappa = \frac{p_0 - p_e}{1 - p_e}$$
(5)

式中： p_0 指对角线所有元素的和比上矩阵中所有的元素， p_e 指所有类别分别对应的“实际与预测数量的乘积”之总和除以样本总数的二次方。由于在实际问题中各个样本的数量会有一定的差别，模型往往会放弃小类别而更加倾向于样本多的类别，通过 Kappa 系数的公式可以看出，越不平衡的矩阵 p_e 越高，Kappa 值越低，刚好能够平衡。

为了检验本文中用于分类的 CART 决策树分类方法的精度，采用基于同样样本对影像进行监督分类时公认使用范围最广、精度较高的最大似然法进行分类，两种方法得到的混淆矩阵分类精度对比如表 6 所示，监督分类的总体精度为 84.90%，Kappa 系数为 0.788 8，而 CART 决策树分类的精度为 88.25%，增加了 3.35 百分点，Kappa 系数增加了 0.045 7，各个地类的生产者精度都有所提高，分类效果较差的林地也提升 4.43 百分点。总体来看，实际运用中纹理特征和其他辅助特征指数的加入使得基于 CART 决策树的分类方法在实际运用上无论是总体精度、Kappa 系数还是单个地类的精度都较之于监督分类有一定的提升，能够更好地满足工作中对于所得数据精准性的要求。

表 6 监督分类与 CART 决策树分类各地精度以及 Kappa 系数对比

类别	监督分类			CART 决策树分类		
	生产者精度/%	用户精度/%	总精度/%	生产者精度/%	用户精度/%	总精度/%
水体	99.84	97.87	84.90	99.34	99.47	88.25
草地	69.61	86.65		74.40	83.43	
林地	44.48	62.03		48.91	58.82	
裸土地	91.78	90.65		92.17	95.37	
裸岩石砾地	78.96	56.57		96.64	53.14	
黑土滩	65.86	77.45		69.75	74.24	
Kappa 系数			0.7888			0.8345

5 讨论与结论

基于纹理特征和其他辅助特征(高程 DEM 数据等)的 CART 决策树分类方法,能够在高寒湿地的地类分类中有较好的精度表现,能够有效地对其进行分类,本研究采用的遥感影像光谱特征、纹理特征、主成分变换降维的第一主成分以及其他辅助特征指数作为数据集,利用 CART 决策树构建的方法,针对高寒湿地分类得到的精度为 88.25%,Kappa 系数为 0.834 5,所构建的决策树模型能够较好地应用于研究区的生态地类提取。

最大似然法分类与 CART 决策树分类相比,林地的分类精度虽有提高,但是仍然有不足之处,主要原因是林地与草地混合比较多,其光谱特征与归一化植被指数较为相近,虽加入了高程 DEM 数据进行区分后取得了一定效果,但仍需要探究更多影响进一步区分效果的因素。

除了本次基于灰度共生矩阵得到的 8 个常用纹理特征分量外,其他测度是否对高寒湿地生态系统分类有帮助还需要进一步研究。

参考文献:

[1] 郑子彦,吕美霞,马柱国.黄河源区气候水文和植被覆盖变化及面临问题的对策建议[J].中国科学院院刊,2020,35(1):61-72.

[2] 李旭,程涛,曹卫星,等.基于 QUEST 决策树的 Landsat 8 遥感影像的南京市土地分类研究[J].湖北农业科学,2017,56(1):35-38.

[3] 隋瑜,徐泮林,谷彦斐.基于遥感数据的土地资源分类方法比较研究[J].测绘与空间地理信息,2021,44(2):107-110.

[4] 张熙,鹿琳琳,王萍,等.基于决策树的漓江上游土地覆盖分类[J].测绘科学,2016,41(3):100-104.

[5] LABA M,BLAIR B,DOWNS R. Use of textural measurements to map invasive wetland plants in the Hudson River National Estuarine Research Reserve with IKONOS satellite imagery[J]. Remote Sensing of Environment, 2010, 114(4):876-886.

[6] 张彦忠,张福祥. TM 数据用于亚热带桔果林区的土地利用分类和专题信息提取[J]. 国土资源遥感,1994(2):28-33.

[7] 罗小波,邓磊,杨富平. 基于 Kohonen 神经网络的遥感影像监督分类[J]. 重庆邮电大学学报(自然科学版),2011,23(5):616-620.

[8] 张磊. 近 32 年黄河源区土地利用和植被指数时空动态变化及其驱动因子分析[D]. 西安: 西安理工大学,2017.

[9] 齐红超,祁元,徐瑱. 基于 C 5.0 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.

[10] 陈静秋,王莉,姜小三. 决策树方法在云贵高原典型区域 ALOS 影像土地利用分类中的应用[J]. 南京农业大学学报,2013,36(6):45-50.

[11] 高凯,周志翔,滕明君. 基于知识分类法的 TM 遥感影像景观分类制图——以武汉市为例[J]. 山东建筑大学学报,2010,25(4):456-459.

[12] 邹文涛,张怀清,鞠洪波,等. 基于 QUEST 树的高寒湿地植被覆盖类型遥感分类研究[J]. 中南林业科技大学学报,2011,31(12):138-144.

[13] 侍昊,李旭文,牛志春. 基于随机森林模型的太湖水生植被遥感信息提取[J]. 湖泊科学,2016,28(3):635-644.

[14] 张滔. 黄河源区土地覆盖变化的遥感监测及时空预测[D]. 南京:河海大学,2011.

[15] 孙璐,张友静,张滔. 黄河源区土地利用/覆盖生态服务功能价值评价及时空预测[J]. 地理与地理信息科学,2014,34(5):99-104.

[16] 白晓娟,赵雨霖,王伟东. 基于 DEM 的哈尔滨市土地利用类型动态变化[J]. 东北林业大学学报,2010,38(12):84-87.

[17] 白娥,薛冰. 土地利用与土地覆盖变化对生态系统的影响[J]. 植物生态学报,2020,44(5):543-552.

[18] 薛斯文,周旗. 基于 CART 决策树的陕北气候变化风险感知信息通道研究[J]. 河南科学,2020,38(10):1641-1649.

[19] 熊慎,李华,赵鑫,等. 基于决策树方法的混凝土泵车液压油超温分析与识别[J]. 建设机械设计与管理,2021,34(5):103-106.

[20] 刘玉婷,李子阳,李恒凯,等. 融合物候和纹理特征的南方柑橘林地提取方法[J]. 测绘科学,2021,46(9):83-93.

Effects of Different Fertilization Methods and Ratios on Agronomic Characters of *Anemarrhena asphodeloides*

TIAN Meng-ni, LI Wei, LIU Rong, REN Yu-min, ZHANG Zi, REN Hong-meng, ZHANG Cheng-hua, JIANG Chang-yu

(Beijing Aibidi Biotechnology Limited Company, Beijing 102299, China)

Abstract: In order to explore the best fertilization method and ratio in the process of *Anemarrhena asphodeloides* cultivation, the annual *Anemarrhena asphodeloides* seedlings produced in Yi County, Hebei Province were transplanted into flower pots for unified slow seedling management, and then the fertilizer experiment was carried out to study the effects of different fertilizer ratios on the plant height, aboveground and root agronomic characteristics of *Anemarrhena asphodeloides*. The results showed that when the fertilizer dilution concentration of foliar spraying treatment was 200 times, the application effect was the best. When 30.0 mL of mixed solution of 2.0 g·L⁻¹ urea+1.0 g·L⁻¹ potassium sulfate, 7.5 mL of amino acid water-soluble fertilizer and 22.5-30.0 mL of GGR 8 were added, which was characterized by higher plant height, wider and thicker leaves, more numbers nitrogen content and chlorophyll SPAD value was higher, the fresh weight and dry weight of leaves were heavier, the root length was longer, the number of fibrous roots was more, the fresh weight and dry weight of fibrous roots were heavier, and the drying rate was higher. When the dilution concentration of fertilizer for root irrigation treatment was 1 000 times, the application effect of adding 20.0 g urea+10.0 g K₂SO₄, 10.0 mL amino acid water-soluble fertilizer and 10.0-20.0 mL GGR 8 was the best. In general, the application effect of foliar spraying treatment is better than that of foliar spraying treatment root irrigation treatment group, and had certain advantages.

Keywords: *Anemarrhena asphodeloides*; water soluble fertilizer; plant growth regulator; GGR

(上接第 29 页)

Research on Remote Sensing Interpretation of Ecological Land Types in the Source Region of the Yellow River Based on CART Decision Tree

DENG Zhen-kun¹, ZHANG Peng-lin²

(1. College of Science, Tibet University, Lhasa 850000, China; 2. College of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China)

Abstract: To guide the restoration of the ecological environment in the Yellow River source region, the alpine wetland classification study was conducted using Landsat 8 imagery as the data source and a CART number model incorporating multiple feature indices, using the Yellow River source region as the study area. The results showed that the data mining process using CART decision tree algorithm and the obtained decision tree were used for classification and mapping to obtain the distribution map of ecological land types in the Yellow River source area, and the total accuracy after classification was 88.25% with a Kappa coefficient of 0.834 5 by confusion matrix, while the overall accuracy obtained by the supervised classification method using the same classification sample with a wider range of maximum likelihood was 84.90% and the Kappa coefficient was 0.788 8, 3.35 percentage points and 0.045 7 were lower than the CART decision tree classification, and the accuracy of each class was lower than the CART tree classification method.

Keywords: Yellow River source; CART decision tree; remote sensing image classification; machine learning