

doi:10.16779/j.cnki.1003-5508.2018.04.018

基于 QUEST 决策树的火烧迹地自动提取

霍 鹏,岳彩荣

(西南林业大学,云南昆明 650224)

摘要:以研究大尺度范围内森林火灾面积提取的可行性为目的。本文以昆明市为研究区,分别提取 2005 年和 2006 年两期 Landsat TM 影像的植被指数、水体指数、火烧迹地指数、纹理特征等,并选取适于大尺度范围的最佳特征波段;利用特征融合后的影像构建决策树,实现大尺度范围内火烧迹地的自动提取。研究表明,QUEST 算法的分类总体精度达 84.5%,在 3 种分类方法中精度最高;采用 QUEST 算法的决策树分类提取“3·29”火烧迹地面积占实际的 97.9%,但大尺度范围内火烧迹地提取效果并不理想,仍需人工剔除提取产生的较小的斑块;最后用 2015 年“3·02”的火烧迹地验证了 QUEST 决策树方法的具有一定的普适性。该模式对大尺度范围内森林火灾面积统计的具有重要应用价值。

关键词:火烧迹地;QUEST;CRUISE;决策树;自动识别

中图分类号:S762.2 **文献标识码:**A **文章编号:**1003-5508(2018)04-0073-06

Automatic Extraction of Burned Area Based on QUEST Decision Tree

HUO Peng YUE Cai-rong

(Southwest Forestry University, Kunming 650224, China)

Abstract: Kunming was taken as the study area to study the feasibility of forest burned area extraction in large-scale range. The vegetation, water, burned site index and texture features were extracted from Landsat TM images in 2005 and 2006, then the best characteristics bands were chosen, which adapted for the large-scale range; Decision trees were determined using the feature-fused image, and fire burned information were automatically extracted in large-scale range. The research showed that the classification accuracy of the QUEST decision tree reached 84.5%, which was better than CRUISE 2D algorithm and maximum likelihood classification; The decision tree classification based on the QUEST algorithm was used to extract 97% of “3·29” burned area, but it was not ideal on large-scale. The smaller patches need to be manually removed; Finally, the “3·02” burned area in 2015 verified that the QUEST decision tree method had certain universality. This model had important application value for the statistics of forest burned areas in large scale.

Key words: Burned area, QUEST, CRUISE 2D, Decision tree, Automatic identification

森林火灾是影响森林生态系统的一个重要的扰动因素^[1]。林火的发生不仅会直接导致森林面积减少、森林结构和生态系统平衡被破坏,而且树木燃烧的过程中向大气中释放大量的 CO₂ 以及其他有

害气体对气候产生了深远的影响,也给当地社会经济带来了严重的损失。因此,森林火灾已成为森林资源和自然环境破坏的重大威胁,其发生原因复杂且难以预防。如何对森林火灾及其遭受的损失进行

收稿日期:2018-05-10

作者简介:霍鹏(1991-),女,辽宁锦州人,硕士研究生,研究方向:GIS RS SAR。

及时、准确和有效地监测和事后损失评估,对预防火灾发生、减少灾害损失、维护生态平衡和保护森林资源都具有非常重要的现实意义^[2]。

近年来,随着遥感技术的逐步成熟,利用其空间分辨率高、监测范围广、时效性强、成本较低的特点,将森林火灾受损状况监测与遥感技术结合起来,已成为我国森林火灾监测的一个重要手段。决策树算法是一种数据挖掘监督分类方法,通过对训练样本集进行数据挖掘和知识发现建立决策树,利用规则对遥感影像进行分类^[3]。不仅直观简单、运算速度快、分类精度高,而且能自动选取特征并融入影像以外的各种知识、能高效地处理大量数据和高维数据。传统的决策树分类在决策树的建立、特征指数的选择和阈值的设定上都较为烦琐,即使有诸如 CART 决策树能自动分类的决策树,但在变量选择时都是有偏的^[4]。

QUEST 算法是 Loh 和 Shih 于 1997 年提出的,该算法将分支变量和分割点的选择分开进行,使其具有兼顾适用于连续型的变量和离散型的变量的特点。此外,它还克服了其他决策树算法更倾向于选择那些具有更多的潜在分割点的预测变量^[5~8]。

CRUISE 2D 算法是 Kim 和 Loh 于 2001 年提出的,它由 FACT 算法改进而来,它融合了 FACT、QUEST 和 CART 算法的思想。CRUISE 2D 算法通过多分割减少树的深度,并采用单因子或双因子交互方法选择预测变量,同时提高了树的可读性^[9~10]。

这两种决策树分类方法皆可以快速、无偏、高效提取具有复杂地形的遥感影像的不同土地类型,与普通决策树相比具有方便、快捷、分类精度高的特点^[11]。在特征指数的适用性上也更趋近于较小的研究范围,一般仅在火灾发生地的局部区域提取火

烧迹地,对大尺度范围内的火烧迹地提取较少。根据指数适用的尺度范围、林火发生后植被减少、燃烧残留物沉积以及地形、纹理特征等,在已提出的修正植被指数、火烧迹地指数等指数中选取最佳指数,为 QUEST 决策树的建立提供坚实的基础。

针对 2006 年云南省昆明市的“3·29”林火事件,由于该燃烧区从安宁市温泉镇西北部的古朗大箐下段森林蔓延至西山区,因此本文采用 QUEST 算法、CRUISE 2D 算法的决策树以及最大似然法在大尺度范围内提取 Landsat 影像中的火烧迹地,并估测每种分类方法的精度,并选择出最适合大尺度范围内火烧迹地提取的方法。结合 2015 年云南省昆明市的“3·02”林火事件验证该方法的普适性。这为在中低分辨率的遥感影像中林火损失面积的自动提取、长时间大尺度范围内的森林火灾监测、林火发生的时空规律、林火的预防等提供数据准备和参考。

1 研究区和研究数据

1.1 研究区概况

本文实验研究区为云南省昆明市,它地处我国云贵高原中部,地理位置介于东经 102°10′~103°40′,北纬 24°23′~26°32′;平均海拔在 1 500 m~2 800 m 之间,总体地势北高南低,中部隆起,东西两侧较低(见图 1)。

昆明市属北纬低纬度亚热带高原山地季风气候,境内大多数地区夏无酷暑,冬无严寒,四季如春,气候宜人,素以“春城”之称而享誉中外。按全年降水量分布分为干、湿两季,每年的 5 月至 10 月为雨季,11 月至次年的 4 月为干季。因整个干季降水量极少仅占全年降水量的 3%~5%,每逢 3 月、4 月份蒸发旺盛致使春季为林火多发季节,昆明市曾多次

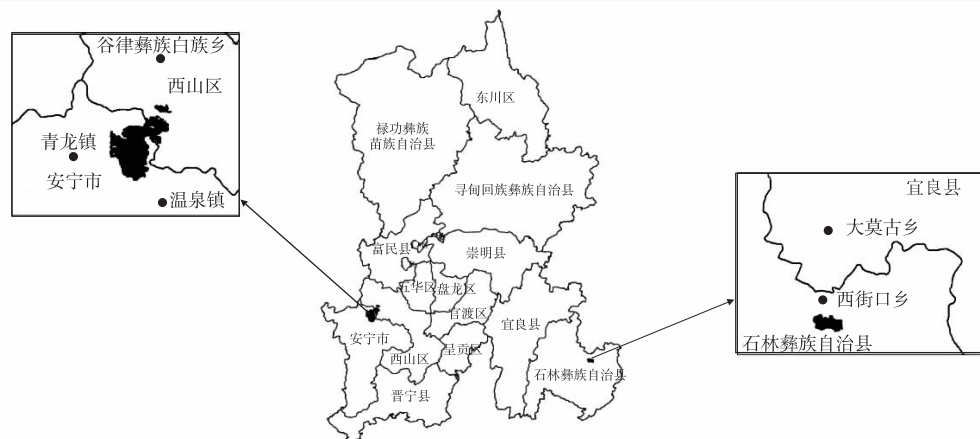


图 1 昆明市研究区图

发生森林火灾事件,因此选该地区为本实验的研究区。

1.2 研究数据

本文的研究数据主要为美国陆地卫星(Landsat)影像和 DEM 数据,所用数据皆取自于地理空间数据云平台(<http://www.gscloud.cn>),其中 Landsat 数据产品为 Level1T 的地形校正数据,其投影坐标为 UTM/WGS84,空间分辨率为 30 m;DEM 数据是 2009 年采集的同为 UTM/WGS84 投影坐标的数字高程模型,其空间分辨率为 30 m。

2 研究方法

2.1 数据预处理

实验数据选用是 2005 年和 2006 年云南省昆明市的 Landsat TM 数据,验证数据采用的是 2015 年 1 月和 3 月的 Landsat8 OLI 数据。由于从地理空间数据云获取的数据为 Level1T 级地形校正数据,已经过系统辐射校正和地面控制点几何校正,因此只需对研究区域做大气校正、影像镶嵌和裁切。

2.2 Cart 决策树火烧迹地信息提取

2.2.1 植被指数

针对大范围内的火烧迹地提取,对其他地物的分离度也要求较高。考虑到电磁波辐射、大气、植被覆盖土壤背景等因素的影响,本文主要对以下几种植被指数进行研究,包括归一化差值植被指数(Normalized Difference Vegetation, NDVI)、多时相差值植被指数(Multi-temporal Normalized Difference Vegetation, MNDVI)、土壤调节植被指数(Soil Adjusted Vegetation Index, SAVI)、修改型土壤调节植被指数(Modified Soil Adjusted Vegetation Index, MSAVI)、全球环境监测指数(Global Environment Monitoring Index, GEMI)以及亮度植被指数(Bright Vegetation Index, BVI)、绿度植被指数(Green Vegetation Index, GVI)、湿度植被指数(Wetness Vegetation Index, WVI)。

经过对不同指数进行比较发现,NDVI 对绿色植被表现最为敏感,即使在低密度的植被区域仍能很好地提取出来^[12]。MNDVI 是将同一区域的两个不同时期的 NDVI 进行波段相减。对于小范围内的发生过火灾的区域,能很好的将植被和其他地类区分开来;但对于大范围的地物提取,即使整景图像中地类变化较小,但也可能会受到影像拍摄时间不同以及阴影等其他因素的影响,降低了整个区域中植

被的提取精度。几种植被指数中仅有 MSAVI 不受云层影响。MSAVI 对火烧迹地区域的区分较为明显,但是植被和城镇会难以区分,且受阴影影响较大,而 SAVI 能将植被和城镇较好的区分开,总体受阴影的影响较小。GEMI 与 NDVI 相比,既保存了比 NDVI 指数相对低密度至浓密度覆盖更大的动态范围,又在不用改变植被信息的前提下减小了大气影响,但仍受到裸土的亮度和颜色相当大的影响,对于稀疏或中密度植被覆盖的区域对植被的提取效果较差^[13]。而根据经验方法得到土壤亮度指数(BVI)、绿度植被指数(GVI)、黄度植被指数(YVI),虽然忽略了大气、土壤、植被间相互作用,但就不同地区区分度来说,BVI 和 WVI 植被和水体严重混分,GVI 与 GEMI 区分效果相似,但受传感器的影响因素较大。

2.2.2 水体指数

水体指数是同植被指数相似的在提取水体领域的一种指数。常见的水体指数主要包括归一化差值水体指数(Normalized Difference Water Index, NDWI)、修改型归一化差异水体指数(Modified Normalized Difference Water Index, MNDWI)、增强型水体指数(Enhanced Water Index, EWI)、新型水体指数(New Water Index, NWI)。

这四种水体指数中 NDWI、EWI 受云的影响较大。本研究采用影像大约为 4 月、5 月左右,滇池的湖水边缘生长了许多漂浮的水生植物,而 EWI、NWI 对水体中的植被比较敏感,在提取水体时对水质要求较高。MNDWI 在几个植被指数中提取水体的效果最好,受云、城市影响较小,不受水质的限制,但对阴影以及覆有大棚的农田影响仍难以剔除^[14]。除以上的水体指数以外,缨帽变换中的湿度植被指数对面积较大的水体提取较好,但对于大范围中的水体提取,效果并不理想易与城市等其他地物发生混淆。

2.2.3 火烧迹地指数

用于林火迹地提取的遥感指数主要有归一化燃烧率(Normalized Burn Ratio, NBR)、差值归一化燃烧率(Difference Normalized Burn Ratio, dNBR)、燃烧面积指数(Burn Area Index, BAI)、碳土壤指数(The Char Soil Index, CSI)。对于颜色较深的水体,这几种火烧迹地指数都难以将之与火烧迹地区分开来。dNBR 比 NBR 受其他地物影响较少,区分度较高^[15]。

2.2.4 特征融合

将全球环境监测指数(GEMI)、修改型归一化差异水体指数(MNDWI)差值归一化燃烧率(dNBR)、灰度共生矩阵的对比度(contrast)以及数字高程数据(DEM)同Landsat的多光谱影像做波段融合。

2.3 基于QUEST算法的分类

本实验将研究区分为6种主要土地类型,包括林地、农地、水体、城市、火烧迹地和裸地,并采用目视解译的方法,在特征融合后的影像上选取训练样本(550个感兴趣区,总计82836个像元)。目视解译是在解译者具备影像解译专业知识、区域背景知识及遥感系统知识的基础上,根据图像基本解译要素和解译标志来识别目标地物。训练样本的选取是构建决策树的关键部分,它的好坏直接影响后期分类精度。因此,在选取训练样本时,考虑了其代表性、典型性及完备性,也就是说训练样本代表了整个区域内不同地类间的光谱特征差异。

3 研究结果与分析

3.1 火烧迹地提取精度评价

为了验证火烧迹地自动识别的精度,根据野外采集数据和目视解译随机选取550个验证样本点,分别计算QUEST决策树、CRUISE决策树和极大似然分类方法的分类精度,建立混淆矩阵进行精度评价,如表1所示。结果表明,QUEST决策树的分类总体精度为84.5%,Kappa系数为0.8161;CRUISE 2D决策树的分类总体精度为83.5%,Kappa系数为0.8041;最大似然分类的分类总体精度为80.0%,Kappa系数为0.7624。分类总体精度是分类器正确

分类的像元总数与总像元数之比,若总体精度越大,分类器分类精度越高;Kappa系数是综合了生产精度和用户精度的指标,若Kappa系数越大,分类结果与真实参考之间差异越小,则分类器分类精度越高。相比其他两种分类方法,火烧迹地的生产者精度和用户精度也是3种方法中最高的。因此,QUEST决策树在大尺度范围内提取火烧迹地的效果最为理想(图2和表1)。

表1 QUEST决策树、CRUISE决策树和最大似然分类精度评价表

分类方法	地类	用户精度	生产者精度	Kappa系数	总体精度
QUEST决策树	森林	61%	99%	0.8161	84.5%
	水体	99%	91%		
	城市	86%	79%		
	火烧迹地	99%	98%		
	裸地	86%	93%		
	农地	88%	82%		
CRUISE决策树	森林	62%	99%	0.8041	83.5%
	水体	95%	93%		
	城市	90%	84%		
	火烧迹地	98%	96%		
	裸地	87%	95%		
	农地	96%	80%		
最大似然分类	森林	63%	99%	0.7624	80.0%
	水体	99%	88%		
	城市	97%	62%		
	火烧迹地	86%	98%		
	裸地	69%	96%		
	农地	90%	83%		

3.2 火烧迹地提取精度分析

3种分类方法中QUEST和CRUISE 2D分类精度相差较小,但在提取火烧迹地的面积上存在很大

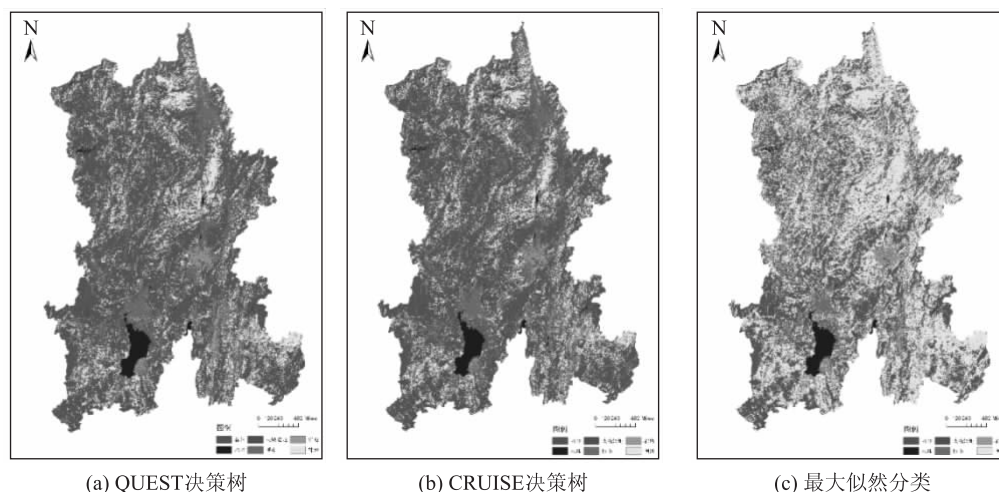


图2 分类结果

的差异。仅针对“3·29”事件发生区域,利用火烧迹地提取精度验证公式(如公式(1)和公式(2))计算3种方法的火烧迹地提取精度。其中,QUEST决策树的火烧迹地提取精度为97.9%;CRUISE 2D决策树的火烧迹地提取精度为80.5%;最大似然分类的火烧迹地提取精度为64.8%。QUEST算法在提取局部火烧迹地与实际面积最为吻合,如图3(a)所示。

识别出的火烧迹地面积 < 实际面积时:

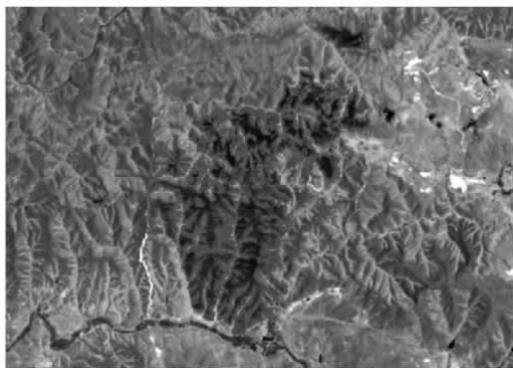
$$\text{识别精度} = \text{识别面积} / \text{实际面积} \times 100\% \quad (1)$$

识别出的火烧迹地面积 > 实际面积时:

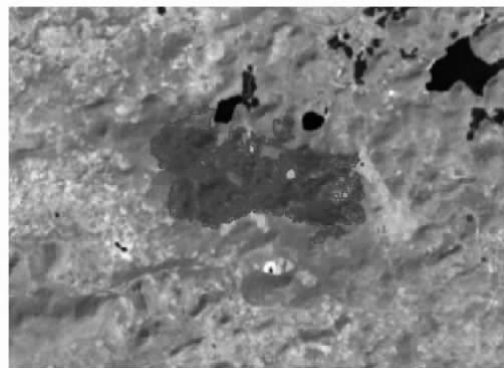
$$\text{识别精度} = 1 - \left(\frac{\text{识别面积} - \text{实际面积}}{\text{实际面积}} \right) \times 100 \quad (2)$$

但是 QUEST 决策树在大尺度范围内对火烧迹

地提取的效果并不理想(如图4),在分类时会产生与其他地类混分的现象,如火烧迹地与裸地混分较为严重,特别是因水土流失后的裸土,与林火发生后裸露的土壤区域在目视解译上颜色较为相近;春季农地焚烧的区域也会被提取出来,判断是否是林火迹地还需有相关地类调查信息,如图4中的(a)和(b)的红框为以上两种情况误分的区域。此外,这3种方法在提取火烧迹地的过程中皆会产生通过分类后处理无法去除的大量的碎斑,这些碎斑需根据国家规定的森林火灾等级的面积,采用手动的方式对碎斑进行剔除。与其相比,最大似然分类虽在总体分类精度上与 QUEST 决策树相当,且混分的地类较少,但提取局部区域的火烧迹地面积与实际面积差距巨大。



(a) 2006年“3.29”火烧迹地



(b) 2015年“3.02”火烧迹地

图3 火烧迹地提取结果

由于自然修复或人为种植会使得火烧迹地在一定时间范围内缓慢恢复,虽无法将生物量恢复为原来的数值,但超过一定的时间限制会使得在目视解译上难以辨认,这就造成了火烧迹地提取的时效性。因此,本实验方法仅可以识别出较短时间范围内火烧迹地的状况,时间范围的选取需根据研究区域的气候环境以及林火灾后修复手段加以判断。

3.3 普适性验证

本文选取2015年云南省昆明市的为验证区域采用同样的方法对研究区进行分类(如图4),研究发现2015年的总体分类精度也达到97%,Kappa系数为0.957,提取火烧迹地的面积占实际精度的89%。可见虽采用不同传感器的同系列卫星的产品,与Landsat TM影像相比的Landsat8 OLI影像多了一个波段,不仅能保证分类的精度,而且Landsat8

OLI数据的多种假彩色波段的组合更有利于通过目视解译区分不同的地类。

两个研究区在时间选取上的差异性对整个影像的分类具有一定的影响。Landsat5 TM影像采用的是2006年4月初的数据,而Landsat8 OLI采用的是2015年3月初的数据。虽仅存在一月之差,但对于中国西南部的气候来说,进入四月气候转暖快滇池等一些区域的水质明显下降,水生生物增多造成水体边界模糊;又正值农耕播种的季节,3月初明显未更耕种的农地会被分成裸地,使得与火烧迹地间隙颜色接近的裸地被误分的可能性增加,这使得在同一区域的不同时期的土地分类产生了差异。又因为决策树的精度对感兴趣区的选取有较高的依赖性,因此这对目视解译人员有较高的要求。

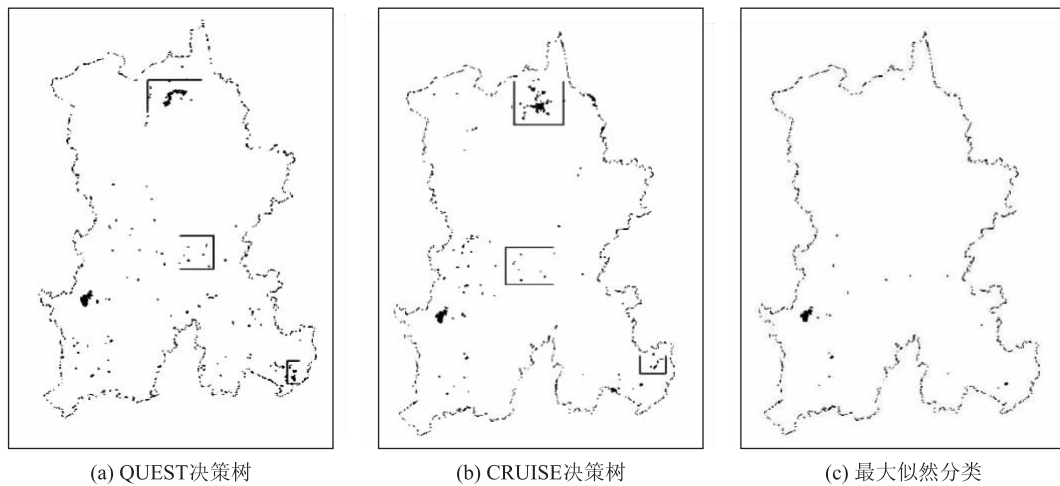


图4 大范围火烧迹地提取图

4 讨论

本文采用 QUEST 算法、CRUISE 算法的决策树以及最大似然分类法对大范围内的火烧迹地进行提取,经实验验证 QUEST 算法的决策树在三种分类方法中对分类精度最高、火烧迹地的提取的效果上最为理想。在实现大尺度范围内对火烧迹地自动提取方面,QUEST 决策树虽在这三种方法中效果最好,但仍有许多待改进之处。

但本实验中的不足之处包括以下几点:根据森林火灾中火种的蔓延特点,在提取火烧迹地的结构中会有许多零散的误分区域,还需要在分类后人工剔除,尚未实现火烧迹地提取的完全自动化;仅利用地物自身的生物特性和整体的生长特征,提取指标单一,检测结果易受干旱、植被季节变化等影响;能否可以通过增加其他特征来降低某些颜色的裸土与火烧迹地混分的状况;为了降低决策树的深度,本文并未采用过多指数的波段融合,仅选取最佳指数,过少的特征是否会影响分类提取的精度;在决策树分到多少级时分类效果最佳等问题还需要以后的研究中逐步验证。

参考文献:

- [1] 杨伟,张树文,姜晓丽.基于 MODIS 时序数据的黑龙江流域火烧迹地提取[J].生态学报,2015,35(17):5866~5873.
- [2] Li Z, Cihlar J, Moreau L, et al. Monitoring Fire Activities in the Boreal Ecosystem. J. Geophys. Res.,1997(103).
- [3] 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.
- [4] Gray J B, Fan G. Classification tree analysis using TARGET[J]. Computational Statistics & Data Analysis, 2008, 52(3): 1362~1372.
- [5] 吴健生,潘况,彭建,等.基于 QUEST 决策树的遥感影像土地利用分类——以云南省丽江市为例[J].地理研究,2012,31(11):1973~1980.
- [6] 李振源,许捍卫,王金龙.基于 QUEST 决策树的大丰湿地信息提取[J].地理空间信息,2014(5):35~37.
- [7] Lim T S, Loh W Y, Shih Y S. A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms[J]. Machine Learning, 2000, 40(3): 203~228.
- [8] 那晓东,张树清,李晓峰,等.基于 QUEST 决策树兼容多源数据的淡水沼泽湿地信息提取[J].生态学杂志,2009,28(2):357~365.
- [9] 杨超.基于混合像元分解决策树的土地利用/覆被分类研究[D].云南师范大学,2017.
- [10] 李双,杨宝华.基于改进 CRUISE 算法的遥感影像分类研究[J].滁州学院学报,2014(2):75~79.
- [11] 白秀莲,巴雅尔,哈斯其其格.基于 C5.0 的遥感影像决策树分类实验研究[J].遥感技术与应用,2014,29(2):338~343.
- [12] 梅安新.遥感导论[M].高等教育出版社,2001:45~50.
- [13] 田庆久,闵祥军.植被指数研究进展[J].地球科学进展,1998,13(4):327~333.
- [14] Baret F, de, ric. Contribution au suivi radiométrique de cultures de céréales[J]. Bibliogr,1986.
- [15] 余哲修,何超,李昔纯,等.基于归一化燃烧率火烧迹地遥感提取识别规则研究[J].西北林学院学报,2018(1):216~224.