

引文格式: 史敏伟, 李庆. 基于随机森林的 GF-2 影像土地利用分类研究[J]. 北京测绘, 2019, 33(12): 1513-1517.

DOI: 10.19580/j.cnki.1007-3000.2019.12.018

# 基于随机森林的 GF-2 影像土地利用分类研究

史敏伟 李 庆

(河南理工大学 测绘与国土信息工程学院, 河南 焦作 454003)

**[摘 要]** 高分二号卫星是我国自主研制的第一颗亚米级遥感卫星。其影像因信息丰富、纹理清晰等特点,被广泛应用于土地利用分类研究中。随机森林是一个包含多个决策树的分类器,是遥感影像土地利用分类方面的一种有效方法。针对单纯的随机森林方法在土地分类中存在的错分的情况。本文基于灰度共生矩阵提取影像纹理信息,同时利用影像的归一化差分植被指数(Normalized Difference Vegetation Index, NDVI),将影像的归一化差分植被指数和纹理信息相结合,采用随机森林的方法进行土地利用分类并得到了较高的分类精度。实验证明了该方法对高分辨率卫星遥感影像的土地利用分类具有较好的效果。

**[关键词]** 高分二号; 随机森林; 纹理信息; 土地利用分类; 归一化差分植被指数

**[中图分类号]** P237

**[文献标识码]** A

**[文章编号]** 1007-3000(2019)12-1513-5

## 0 引言

使用传统测量的方法调查土地利用情况费时费力,而遥感影像具有快速性和周期性的特点,所以利用遥感技术进行土地利用分类研究是一种重要的方法。高分二号卫星是我国发射的首颗空间分辨率小于 1 m 的民用光学卫星,分别搭载了 1 m 全色和 4 m 多光谱相机,具有亚米级空间分辨率、高精度定位和快速姿态机动能力等特点<sup>[1]</sup>。该影像的全色和多光谱影像可以获取空间信息和光谱信息,通过图像融合将会得到高分辨率的多光谱图像。目视解译中,我们一般使用图像的光谱信息。但随着影像分辨率的提高,影像波段数量较少,所以仅使用影像的光谱信息来提高地物分类提取精度具有局限性,所以对于影像的纹理信息的提取和应用就变得更加重要<sup>[2]</sup>。

由于近年来城镇化的快速发展,土地利用情况变化较大,所以及时掌握土地利用分类的情况,可以为城乡土地资源的合理规划提供依据。但是在分类过程中,对于某些地物会出现“同物异谱”和“异物同谱”,从而导致错分现象。特别是对于耕地的农作物植被和林地,常常被混淆分

类,从而影响分类精度。为解决这一现象,经大量试验发现,通过利用不同地物的 DNVI 指数和纹理特征信息,可以大大减弱此类影响。

随机森林(Random Forest, RF)是由 Leo Breiman 和 Adele Cutler 提出的利用多棵树对样本进行训练并预测的一种分类器。随机建立一个森林,这个森林由很多单棵决策树组成<sup>[3]</sup>,随机森林中的单棵决策树之间是相互独立的。当输入新的样本时,随机森林中的每一棵决策树会进行类别判断并投票,最后分类结果为票数最多的类别。随机森林的优点在于训练可以高度并行化,尤其是在大数据时代的大样本训练速度有明显优势。

## 1 研究区数据介绍

研究区域为河南省安阳市和鹤壁市的交界地区,主要为平原地形,研究区的地物信息主要为耕地、林地、水体和建筑用地。研究区数据为高分二号遥感影像,全色波段空间分辨率为 1 m,多光谱波段空间分辨率为 4 m。通过像素工厂软件对遥感影像进行空三<sup>[4]</sup>、正射校正、数据融合<sup>[5]</sup>等预处理工作,最终得到分辨率为 1 m,包含红、绿、蓝和近红外四个波段的影像数据。

**[收稿日期]** 2019-05-20

**[作者简介]** 史敏伟(1993—),男,河南安阳人,硕士在读,从事遥感影像土地利用分类方面研究。

**E-mail:** 913936709@qq.com

## 2 土地利用分类方法

### 2.1 NDVI 指数提取

NDVI(normalized difference vegetation index)即归一化差分植被指数<sup>[6]</sup>,它是反映农作物长势和营养信息的重要参数之一,主要在遥感领域的植被覆盖度和植被生产情况方面有着较为广泛的应用,计算公式为:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (1)$$

其中,NIR 为近红外波段,R 为红波段。

### 2.2 纹理信息提取

目前常用的纹理信息提取方法有几何法、模型法、信号处理法和统计方法。本文采用灰度共生矩阵法(Gray-level co-occurrence matrix)是建立在图像的二阶组合条件概率密度函数基础上的一种统计方法<sup>[7]</sup>。灰度共生矩阵法主要是通过研究灰度的空间相关性来描述图像纹理的方法,具有较强的适应性和抗变换性即鲁棒性。不同的实验目的选取的纹理组合也不相同,通过实验对比,本文选取了中值(Mean)、协同性(Homogeneity)、熵(Entropy)3 个特征量为图像的纹理特征。

### 2.3 随机森林分类

#### 2.3.1 随机森林分类原理

随机森林是以多棵决策树为基础组成的集成学习器。因为单棵决策树抗干扰能力低,在数据量较为复杂的情况下会出现“过拟合”现象,所以利用多棵决策树对选定样本进行训练并预测,最终投票决定最优分类结果,从而较好地解决了过拟合现象并取得较好的泛化能力。

随机森林是用随机的方法建立一个森林,森林里面包含很多棵决策树,并且随机森林中的每棵决策树相互之间是独立的。在森林组成之后,每当有一个新的样本进入森林时,森林中的每棵决策树分别自动进行下一判断<sup>[9]</sup>,从而确定这个样本应该属于哪一类别,然后统计哪一类别投票选择最多,最后预测这个样本为哪一类。如图 1 所示。

#### 2.3.2 特征选择

目前的特征选择方法主要有信息增益、卡方检验和基尼指数。本文采用的随机森林采用基尼指数<sup>[9]</sup>(GINI Index)来选择特征的。基尼指数是用来表示一个随机选中的样本在其子集中被错

分的可能性,即计算样本集合的纯度和不确定性。如果基尼指数最小,表示纯度最高,不确定性最小,数据分割越彻底,分类精度最高。反之,基尼指数越大,表明不确定性就越大,分类精度较差。对于一般的决策树,假定总共有 D 类,样本属于第 D 类的概率为  $P_D$ ,则该概率分布的基尼指数为:

$$Gini(P) = \sum_{D=1}^D P_D(1-P_D) = 1 - \sum_{D=1}^D P_D^2 \quad (2)$$

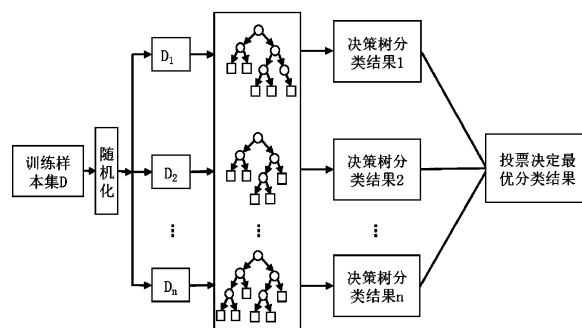


图 1 随机森林分类过程图

### 2.4 随机森林分类流程

本文基于随机森林的方法构建了土地利用分类模型。图 2 为具体的分类流程。

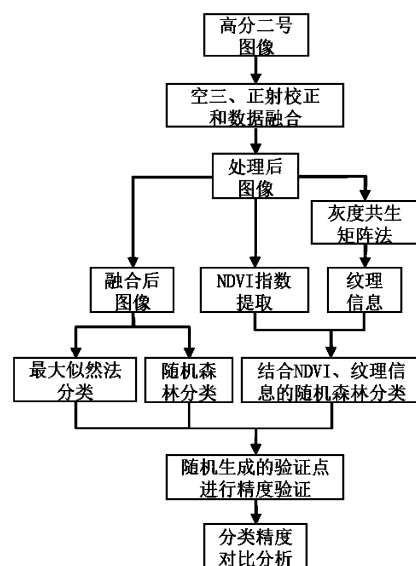


图 2 结合 NDVI 和纹理信息的随机森林分类流程

具体流程主要包括:(1)对研究区数据进行空三、正射校正和数据融合等数据预处理工作。(2)对融合后数据进纹理特征提取和 NDVI 指数提取,获得图像的均值(Mean)、协同性(Homogeneity)、熵(Entropy)三种纹理信息和 NDVI 指数信息。因为数据为 4 波段数据,提取三种纹理信

息则会产生 12 种特征信息,所以在提取纹理信息之前,首先进行主成分分析,本文主要对第一主成分进行均值、协同性、熵三种纹理信息提取。然后将原图像和 NDVI 指数、三种纹理信息进行波段组合,得到试验区增加信息后的影像数据。(3)将结合信息后的图像进行随机森林法分类,将分类结果与未结合信息的随机森林方法分类结果和最大似然分类法<sup>[10]</sup>的结果进行对比分析。

3 实验过程与分类结果分析

3.1 试验区图像

本文通过 ENVI、IDL 软件提取纹理信息、NDVI 和土地分类研究,在纹理信息提取方面,应着重考虑窗口大小和移动步长,本文选择的窗口为 3×3 像元窗口,移动步长为(1,1),所得的纹理信息较好,均值反映对象内部差异的平均值。协同性是用来表达图像纹理的同质性,衡量图像纹理局部变化的情况,其值较大表明不同区域间的图像纹理变化较小,局部比较均匀。熵是反映图像所具有的信息量,表达了图像中纹理的均匀程度和图像的复杂程度。当熵值较大时,复杂程度较高,纹理信息更加丰富,如图 3 所示。

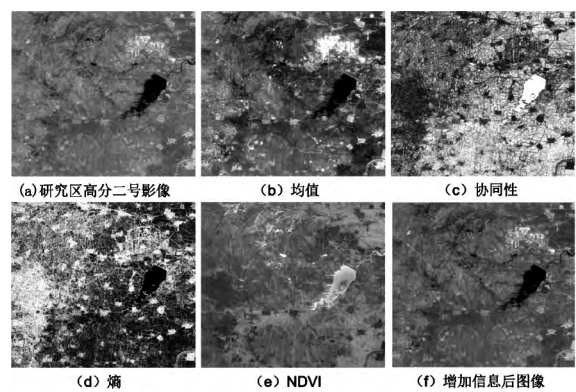


图 3 原始影像与纹理特征、NDVI 指数图

利用随机森林方法进行分类,首先需要进行主要参数设置。本文的参数设置为决策树数量为 300,特征数量(number of features)选定为 Square Root,即特征数量为输入影像波段数的平方。Impurity Function 选择为 Gini coefficient 即基尼指数。

3.2 样本可分离性

样本的可分离性是区分相近地物的一项重要指标 Jeffries-Matusita(JM)与 Transformed Divergence(TD)两个参数是分别用来表示各种类型样本之间的特征分离性和转换分离度,这两

个参数值介于 0~2 之间,若参数值在 1.9~2 之间说明样本可分离性好,属于合格样本。当小于 1.8 时,需考虑编辑或重新选择样本。最后若小于 1,则应考虑将两类样本合并。

Jeffries-Matusita(JM)计算公式为:

$$a = \frac{1}{8}(u_i - u_j)^T (\frac{C_i + C_j}{2})^{-1} (u_i - u_j) + \frac{1}{2} \ln (\frac{C_i + C_j}{\sqrt{|C_i|} \times \sqrt{|C_j|}}) \tag{3}$$
$$JM = 2(1 - e^{-a}) \tag{4}$$

式中, $i$  和  $j$  为两个不同类别; $C_i$  和  $C_j$  为类别特征值的方差, $u_i$  和  $u_j$  为类别特征值均值。

Transformed Divergence (TD) 的计算公式为:

$$D_{ij} = \frac{1}{2}tr[(C_i - C_j)(C_i^{-1} - C_j^{-1})] + \frac{1}{2}tr[(C_i^{-1} - C_j^{-1})(u_i - u_j)(u_i - u_j)^T] \tag{5}$$
$$TD_{ij} = 2[1 - \exp(\frac{-D_{ij}}{8})] \tag{6}$$

式中, $D_{ij}$  为两个类别的离散度, $C_i$  和  $C_j$  分别是类别  $i$  和  $j$  的协方差矩阵, $u_i$  和  $u_j$  为类别的平均矢量。

在 ENVI 中使用 Compute ROI Separability 工具即可以计算样本可分离性。从表 1、2 中我们可以看出,经过增加信息的图像,其分离性明显优于未增加信息的图像,特别是林地与耕地、建筑用地与耕地之间的可分离性有了明显的提高。从而可以一定程度上减小相近地物类型混淆的情况。

表 1 未增加信息图像的样本可分离性

类别	特征分离性	转换分离度
林地与耕地	1.122 6	1.197 8
建筑用地与耕地	1.570 5	1.898 9
建筑用地与林地	1.764 2	1.947 1
建筑用地与水体	1.827 8	1.999 5
林地与水体	1.958 9	1.997 8
水体与耕地	1.999 1	1.999 9

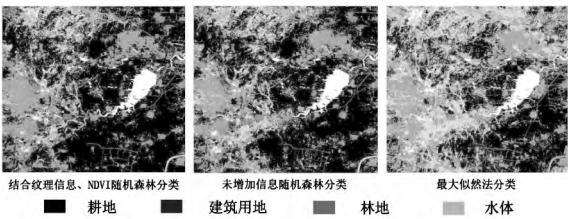


图 4 分类结果图

表 2 增加信息图像的样本可分离性

类别	特征分离性	转换分离度
林地与耕地	1.620 2	1.921 4
建筑用地与耕地	1.888 7	1.996 9
建筑用地与林地	1.815 7	1.987 1
建筑用地与水体	1.952 1	1.999 9
林地与水体	1.921 4	1.999 2
水体与耕地	1.999 1	2.000 0

### 3.3 分类结果及精度评价

通过设定分类器参数,得到分类结果如图 4 所示,从图中我们看出,增加信息后的分类结果效果最好,各种地物都能较好的区分出来,比较于其他两种分类结果,林地和耕地错分的情况明显减少。

研究区影像大小为  $20\ 160 \times 20\ 160$  像素,随机生成  $500 \times 500$  像素共 2 500 个验证点,对随机生成的验证点进行目视解译,并标注其属性标签,然后将其用于评价分类精度,本文采用的评价方法为总体分类精度(Overall Accuracy)和 Kappa 系数。随机生成验证点如图 5 所示。

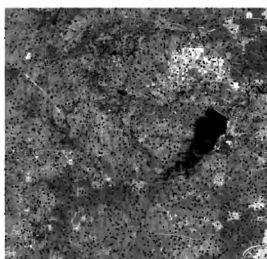


图 5 随机生成的验证点

通过对三种分类结果进行统计,并进行混淆矩阵计算得出各分类结果的 Kappa 系数和总体精度。通过表 3 可以看出,结合纹理信息和 NDVI 指数信息的分类精度高于另外两种方法。实验表明,结合纹理信息和 NDVI 指数信息可以丰富高分二号影像数据的波段信息来提高地物分类的精度。随机森林分类精度高于最大似然法分类精度,也表明了随机森林这种基于机器学习的分类算法在高分辨率影像分类中优于基于概率的最大似然法。

表 3 分类精度表

分类方法	总体精度/%	Kappa 系数
结合信息随机森林法分类	93.40	0.844 7
随机森林法分类	89.32	0.755 5
最大似然法分类	80.32	0.626 7

## 4 结束语

本文以高分二号影像为数据源,利用波段计算提取 NDVI 指数、灰度共生矩阵法提取纹理信息,采用随机森林的方法对研究区进行土地利用分类,结论如下:

(1) 结合纹理信息、NDVI 指数的随机森林分类精度达到 93.4%,高于未增加信息的随机森林分类的 89.32%和极大似然法分类的 80.32%。表明了纹理信息和 NDVI 指数是高分辨率影像的重要信息,可以有效地提高土地利用分类精度。

(2) 在没有增加信息的情况下,从总体精度和 Kappa 系数可以得出:随机森林分类法优于最大似然分类法,说明随机森林方法在高分影像土地利用分类研究中具有一定的优势。

## 参考文献

- [1] 王忠武,刘顺喜,戴建旺,等. 高分二号卫星数据在新增建设用地监测中的应用分析[J]. 航天器工程,2015,24(6):134-139.
- [2] 郑淑丹,郑江华,石明辉,等. 基于分形和灰度共生矩阵纹理特征的种植型药用植物遥感分类[J]. 遥感学报,2014,18(4):868-886.
- [3] 李垒,任越美. 基于随机森林的高光谱遥感图像分类[J]. 计算机工程与应用,2016,52(24):189-193.
- [4] 邢著荣,冯幼贵,赵青,等. 基于 QuickBird 数据的山区正射影像制作及精度分析[J]. 北京测绘,2015(3):5-8,4.
- [5] 李珊珊,洪友堂,胡超,等. 基于高分二号影像的融合方法对比研究[J]. 北京测绘,2019,33(3):261-265.
- [6] 王金亮,邵景安,李阳兵. 基于多种遥感指数综合应用的典型地物分类——以重庆市主城 6 区为例[J]. 地球信息科学学报,2013,15(6):925-931.
- [7] 侯群群,王飞,严丽. 基于灰度共生矩阵的彩色遥感图像纹理特征提取[J]. 国土资源遥感,2013,25(4):26-32.
- [8] 马慧娟,高小红,谷晓天. 随机森林方法支持的复杂地形区土地利用/土地覆被分类研究[J]. 地球信息科学学报,2019,21(3):359-371.
- [9] 韩燕,王玲,罗冲. 随机森林算法在干旱区土地利用遥感分类中的应用研究[J]. 石河子大学学报(自然科学版),2017,35(1):95-101.
- [10] 田绍鸿,张显峰. 采用随机森林法的天绘数据干旱区城市土地覆盖分类[J]. 国土资源遥感,2016,28(1):43-49.

## Land Use Classification of GF-2 Image Based on Random Forest

SHI Minwei, LI Qing

(School of Surveying and Mapping Land Information Engineering, Henan Polytechnic

University, Jiaozuo Henan 454003, China)

**Abstract:** The GF-2 satellite is the first sub-meter-level remote sensing satellite independently developed by China. It has been widely used in land use classification research because of the rich information and clear texture. Random forest is a classifier with multiple decision trees and has become an effective method for remote sensing land use classification. The problem of the misclassification in the land classification for the simple random forest method. This paper extracts the image texture information based on the gray level co-occurrence matrix, and uses the normalized difference vegetation index (NDVI) of the image, combining the image with the texture information and NDVI, and using the random forest method for the land use classification. The higher classification accuracy was obtained. Experiments show that the method has good effect on land use classification of high resolution satellite remote sensing images.

**Key words:** GF-2; random forest; texture information; land use classification; Normalized Difference Vegetation Index (NDVI)