

引用格式: Yin Yanmin, Jia Li. Sentinel-2 Study on Crop Mapping of Shandian River Basin based on Images[J]. Remote Sensing Technology and Application, 2021, 36(2): 400-410. [尹燕旻, 贾立. 基于Sentinel-2的闪电河流域农作物分类研究[J]. 遥感技术与应用, 2021, 36(2): 400-410.]
doi: 10.11873/j.issn.1004-0323.2021.2.0400

基于 Sentinel-2 的闪电河流域农作物分类研究

尹燕旻, 贾立

(中国科学院空天信息创新研究院, 遥感科学国家重点实验室, 北京 100101)

摘要: 以内蒙古闪电河流域为研究区, 基于 Sentinel2 光学遥感影像结合随机森林和支持向量机算法, 采用 3 种方案: 基于像元的分类方法、面向对象的分类方法及改进的基于像元分类与面向对象分割相结合的集成方法, 对研究区内的农作物进行精细提取。结果表明: ① 基于随机森林采用基于像元的方法进行分类, 所有地类的总体精度为 97.8%, Kappa 系数为 0.974, 表明随机森林算法可以有效地进行农作物提取。② 改进的基于像元分类与面向对象分割相结合的集成方法分类效果较好, 所有地类的总体精度为 96.4%, Kappa 系数为 0.957, 该方法充分结合了基于像元和面向对象分类方法的优点, 可有效提升闪电河流域的作物分类效果。

关键词: Sentinel-2; 农作物分类; 面向对象; 基于像元; 随机森林; RF; SVM

中图分类号: TP79 **文献标志码:** A **文章编号:** 1004-0323(2021)02-0400-11

1 引言

农业是国民经济的基础, 是国家长治久安、百姓安居乐业的保障。农作物分布信息是进行农业结构调整, 提升种植管理水平的重要基础信息, 对于社会经济发展和粮食安全具有十分重要的意义^[1], 也是农作物需水量、水资源配置、农田灌溉系统以及水利设施配置等研究的重要背景信息。

由于具有获取信息速度快、观测范围广等优点, 遥感技术被广泛应用于农情监测中农作物分类^[2]。通过遥感进行作物分布制图, 不仅可以全面掌握作物分布信息, 还能够在此基础上进一步揭示影响作物生长的环境及气候因素, 有助于增加农作物的多样性, 帮助分析农业生产模式、调整耕作方式, 为农业灌溉系统及其他现代化生产设施的优化提供决策数据^[3-4]。

目前现有的农作物空间分布遥感产品由于数据及样本质量的制约, 最高分辨率为 30 m, 且分类方法对于多类型农作物种植结构识别能力不足, 缺

乏对于复杂种植模式、气候、地形条件的农作物分类方法^[5]。与国外已有的农作物空间分布遥感产品(如 Cropland Data Layer Data, 简称 CDL)相比, 国内尚缺乏 30 m 分辨率的农作物分布产品, 提取的农作物类型种类偏少, 主要集中于水稻、小麦和玉米等大宗作物, 限制了该类数据的应用潜力^[6-7]。

闪电河流域试验区是遥感科学国家重点实验室牵头组织的滦河—闪电河联合遥感实验飞行实验区的一部分, 在该区域已开展多项大型遥感基础实验, 而该地区地表覆盖及农作物分类数据是各项实验重要的基础数据, 开展该地区的土地分类尤其是农作物分类研究具有十分重要的意义。该实验区位于我国内蒙古与河北交界, 属于我国北方农牧交错地带, 土地利用和土地覆盖变化十分复杂, 驱动因素较多, 准确地获取较高分辨率农作物分类图是研究该区域农作物水分需求和利用的必要条件, 同时也是一个研究难题^[8]。由于哨兵 2 号(Sentinel-2)数据具有较高的空间分辨率、独特的“红边”波段、

收稿日期: 2020-06-22; 修订日期: 2021-04-04

基金项目: 中国科学院战略性先导科技专项(XDA19030203)。

作者简介: 尹燕旻(1994—), 女, 山东泰安人, 硕士研究生, 主要从事农作物分类研究。E-mail: yinym@radi.ac.cn

通讯作者: 贾立(1965—), 女, 北京人, 研究员, 主要从事水循环遥感研究。E-mail: jiali@radi.ac.cn

影像可免费获取等显著优点^[9],因此使用该数据并结合机器学习的方法进行分类研究是进行闪电河流域农作物分类的一种有效方法。

以内蒙古闪电河流域为研究区,基于欧空局 Sentinel-2 光学遥感影像,提取相应的光谱特征,纹理特征以及指数特征,利用随机森林和支持向量机算法,采用 3 种方案:基于像元的分类方法、面向对象的分类方法、改进的基于像元分类与面向对象分割相结合的集成方法,对研究区内的农作物进行精细提取,并利用采集的地面样本进行了分类精度验证。

2 研究区与数据

2.1 研究区概况

闪电河流域实验区位于我国内蒙古与河北交界处,闪电河上游(图 1),属于我国北方农牧交错地带,土地利用和土地覆盖变化十分复杂,驱动因素较多。

研究区地处内蒙古高原南缘,属于温带大陆性草原气候,年降水量 300~400 mm,约为蒸发量的 25%,汛期主要反映在 7~9 月^[10]。研究区主要土地利用类型包括耕地、草地,林地和荒地,主要作物包括:莜麦、土豆、胡萝卜、牧草和玉米等。

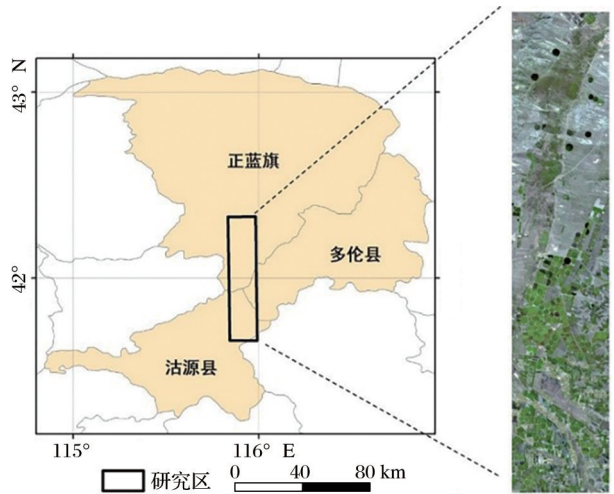


图 1 闪电河流域实验区区位图

Fig. 1 Experimental area map of Shandian river basin

2.2 研究数据与预处理

2.2.1 遥感数据

采用欧洲航天局(ESA)提供的 Sentinel-2 数据进行闪电河流域的作物提取,影像采集日期是 2018 年 8 月 1 日和 9 月 5 日,数据详细信息见表 1。Sentinel-2 数据的波段划分如下:Band1 为海岸/气溶胶波段,用来监测近岸水体和大气中的气溶胶;Band2、3、4 为可见光波段;Band5、6、7 是红边范围内波段,对监测植被健康信息非常有效^[11];Band 8 是近红外波段(宽);Band8A 是近红外波段(窄);Band 9 是水蒸气波段;Band10、11、12 是短波红外波段。

Sentinel-2 数据大气上层表观反射率产品,已经经过几何校正和辐射校正,所以只需要进行大气校正即可^[12]。实验中使用 SNAP 软件中的 Sen2cor 模型对其进行大气校正,大气校正后将 20 m 分辨率波段重采样为 10 m 分辨率。本文采用了波段 2、3、4、5、6、7、8、8a、11、12 反射率数据。

表 1 闪电河流域研究区 Sentinel2 遥感影像信息
Table 1 Sentinel2 image acquisition time in Shandian river basin study area

获取日期	获取卫星	波段数	影像质量	作物生长期
2018-08-01	S2A	13	无云	胡萝卜、土豆:块茎膨大期 莜麦:抽穗期
2018-09-05	S2B	13	无云	胡萝卜、土豆:成熟期 莜麦:收获期

2.2.2 地面样本数据

采用的地面样本数据为 2018 年夏季使用手持 GPS 进行实地采集,并且结合 Google Earth 手动获取。每类地物实地标注约 200 个样方,每个样方内部勾画 5 个子样方进行存储^[13]。在进行基于像元的分类时,每个 Sentinel-2 图像的 10 m 像元为 1 个样本;在进行面向对象的分类时,包含样本像元的单个对象为 1 个样本。表 2 统计了基于像元分类和面向对象分类不同尺度下的各类别样本信息。样本集通过 python 语言的 Sklearn 库进行自动随机划分,80% 为训练集,20% 为测试集。

表 2 闪电河研究区农作物分类样本统计表

Table 2 Sample statistics for crop classification in Shandian river basin study area

方案	尺度	玉米	土豆	胡萝卜	莜麦	林地	草地	建设用地	荒地
方案一	像元	9 672	34 752	19 501	11 709	22 649	8 716	10 560	3 679
	25	264	1 018	577	432	515	584	1 531	285
方案二	50	100	423	235	175	152	321	400	107
	100	46	193	111	86	63	190	153	49

3 研究方法

3.1 实验方案介绍

共设计3种分类方案进行对比研究,分别是:基

于像元的分类方法、面向对象的分类方法、基于像元分类与面向对象分割相结合的集成方法,目的是通过不同方案间的对比,探索闪电河流域农作物提取的最佳分类方案。实验总体流程如图2所示。

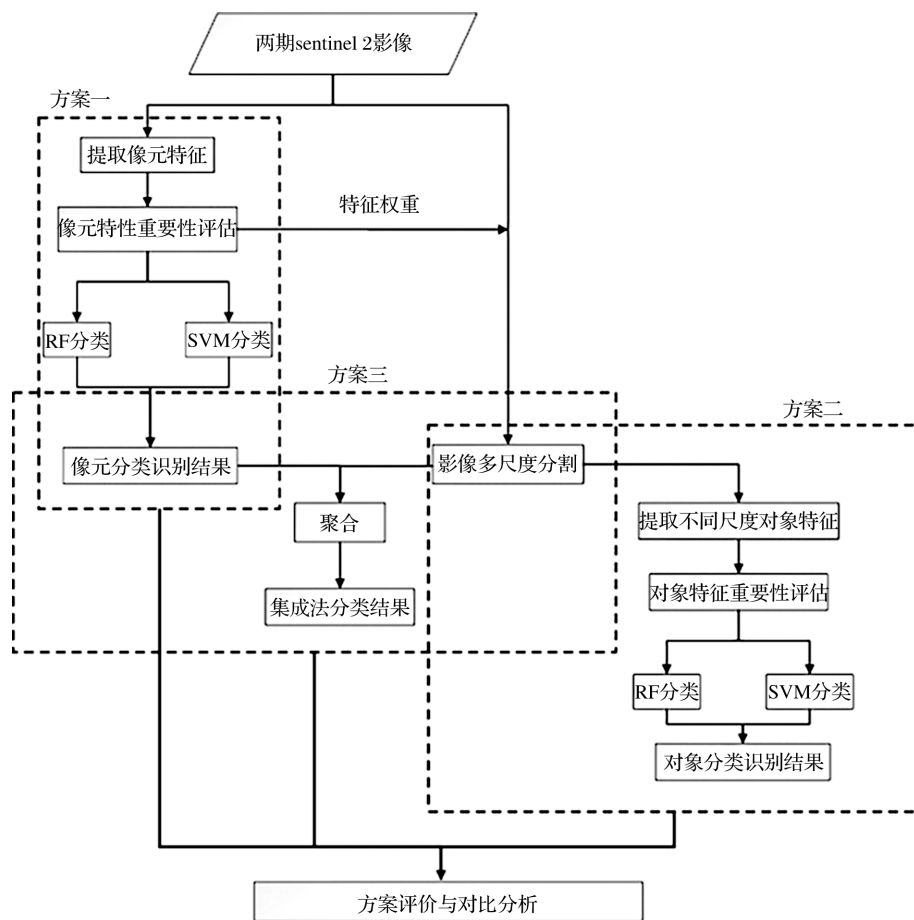


图2 闪电河流域作物分类实验方案流程图

Fig.2 Flow chart of crop classification experiment scheme in the Shandian river basin

方案一:基于像元的分类。对预处理后的两期Sentinel-2影像提取像元级特征,通过随机森林重要性评估算法获得不同像元级特征的权重并进行特征选择,使用随机森林和支持向量机进行分类,对其进行精度评定及对比分析,得到该方案的最佳分类结果。

方案二:基于对象的分类。依据第一步中提取的像元级特征权重进行多尺度影像分割,提取3种不同尺度的对象级特征;通过随机森林重要性评估算法获得对象级特征的权重并进行特征选择,使用随机森林和支持向量机进行分类,对其进行精度评定及对比分析,得到该方案的最佳分类结果。

方案三:基于像元和对象集成的分类。依据方案一得到的最优像元分类结果和方案二得到的3种尺度的分割结果,进行对象内的像元聚合来确定对

象的地物类别,获得3种分类结果,对其进行精度评定及对比分析,得到该方案的最佳分类结果。

最后,对3种方案的分类效果从总体精度、各类别分类精度、目视解译效果、分类性能等方面进行对比分析。

上述3种实验方案共设计了5个子实验(表3),目的是:①对比分析不同方案中,随机森林分类器与SVM分类器的分类性能;②探索不同方案中面向对象分割的最佳尺度;③通过方案之间的对比,探索提升闪电河流域农作物分类效果的最佳方法。

3.2 特征说明

3.2.1 像元特征集

本文设计选取两个时相的光谱特征、植被指数、水体指数和纹理特征共36个特征构建像元特征集,表4详细描述了各种特征及其表达方式。实验

表 3 闪电河研究区作物分类实验方案信息

Table 3 Crop Classification experiment program information sheet in Shandian river basin

方案	子实验	详细信息
1	a	基于像元的随机森林分类
	b	基于像元 SVM 分类
2	c	25、50、100 分割尺度下基于对象的随机森林分类
	d	25、50、100 分割尺度下基于对象的 SVM 分类
3	e	依据方案 1 中优选像元分类结果与方案 2 中不同尺度分割结果,进行对象内聚合

对每个时相影像各选取 10 个波段的反射率作为光谱特征,共提取 20 个光谱特征;对影像进行处理后提取植被指数 NDVI 和水体指数 NDWI 两个指数特征,共提取 4 个指数特征^[14]。

不同农作物之间物候差异可能比较小,由于本文选用的影像分辨率较高,因此在光谱特征、植被指数、水体指数的基础上,借助纹理信息可显著提高农作物的分类精度^[15]。本文选取灰度共生矩阵 GLCM (Gray-Level Co-occurrence Matrix) 进行纹理分析,其数学表达模型为:

$$P(i,j;d,\theta) = \# \left\{ (x_1,y_1)(x_2,y_2) | f(x_1,y_1) = i, f(x_2,y_2) = j, \left| (x_1,y_1) - (x_2,y_2) \right| = d, \angle((x_1,y_1)(x_2,y_2)) = \theta \right\} \quad (1)$$

在该式中, # 表示在该集合中的元素的数目, $f(x_1,y_1) = i, f(x_2,y_2) = j$ 表示图像中 (x_1,y_1) 与 (x_2,y_2) 位置像元的灰度级为 i 和 j , d 表示两个像元之间的距离, θ 表示两个像元之间的方向角,通常取 $d = \{1,2,3,4\}$, $\theta = \{0^\circ,45^\circ,90^\circ,135^\circ\}$ 。

研究表明高分一号等高分辨率影像红外波段所提取的灰度共生矩阵可有效提升农作物分类精度^[16],因此本文对两个时相影像的 B8 波段分别提取角二阶矩、相关性、对比度、熵、同质度、非相似性 6 个标量 4 个方向统计值叠加的均值^[17],共提取 12 个纹理特征。

表 4 作物分类像元特征集描述

Table 4 Description of pixel-wise features for crop classification

特征变量	简称	说明
光谱特征	B	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12
指数特征	NDVI	(B8a-B4)/(B8a+B4)
	NDWI	(B3-B8a)/(B3+B8a)
纹理特征	GLCM-A	角二矩阵
	GLCM-cor	对比度
	GLCM-con	相关性
	GLCM-E	熵
	GLCM-dis	非相似性
	GLCM-H	同质度

3.2.2 对象特征集

本文使用 eCognition 软件提供的多尺度影像分割算法 (Multiresolution Segmentation) 进行影像分割,需要设定波段权重、分割尺度参数、异质性相关因子 (color、shape、compactness 因子) 等参数^[18]。将上文提取的像元特征集作为输入数据,通过随机森林算法可获得像元特征集的各特征重要性排序,来确定波段权重;采用目视法,初步设置分割尺度参数为 25、50、100;所选异质性相关因子, shape 为 0.1、0.3、0.5、0.7、0.9, compactness 为 0.1^[19]。得到 3 种不同尺度的分割对象后,提取对象内部像元特征的均值、标准差、贡献率 3 种统计特征作为对象特征集^[20],共提取 104 个特征,表 5 详细描述了各种特征及其表达方式。

3.3 分类算法

在进行基于像元分类和面向对象分类实验的时候,选用了随机森林和支持向量机两种分类器进行分类。

3.3.1 随机森林算法

随机森林是由 Breiman 提出的基于 bagging 方法使用一组分类回归树 (CARTs) 进行预测的一种集成学习的方法^[21]。其具有许多优点,包括速度快、并行性强、噪声鲁棒性强和固有的多类性质。由于这些特点,在遥感图像分析中得到了广泛的应用^[22]。

随机森林方法建立步骤如下:

第一步:在原始样本中,随机抽取 N 个训练样本 (bootstrap 方法) 放回原始样本,共抽取 N 个树次,形成 N 个树训练样本集。

第二步:基于训练样本集,建立 n 棵 CART 决策树,形成随机森林。在决策树的生长过程中,每棵树的每个节点随机抽取 m 个特征 (特征总数为 M, $m \leq M$)。根据最小基尼系数原理,选择一个分类能力最强的特征来分割决策树中的节点。

第三步:生成多个决策树,构造一个随机森林分类器,并通过投票法决定新样本的类别对数据进

表 5 作物分类对象特征集描述

Table 5 Description of object-wise features for crop classification

统计特征	简称	像素特征类别	说明
均值	Mean	光谱特征	求一个对象内所有像素的像素特征的均值
		植被指数	
		水体指数	
		纹理特征	
标准差	Std	光谱特征	求一个对象内所有像元的像元特征的标准差
		植被指数	
		水体指数	
		纹理特征	
贡献率	Ratio	光谱特征	求同一期影像内,一个对象的某一光谱特征平均值与所有光谱特征平均值总和的比率
		纹理特征	

行分类。

3.3.2 支持向量机分类算法

支持向量机算法(SVM)是一种基于统计学习的新的机器学习的方法,该算法采用结构风险最小化的原则,在确保误差最小的同时,缩小模型泛化误差的上限,从而提高分类模型的泛化能力^[23]。

SVM在解决分类问题时,使用分离超平面作为分离训练数据的线性函数,其求取最优分类面的优化函数如下:

$$\max W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j a_i a_j (x_i, x_j) \quad (2)$$

其中: x 为分类样本, l 为样本个数, y 为类别。 a_i 是函数优化的Lagrange系数,其判别函数为:

$$g(x) = \text{sign}(f(x)) = \text{sign}(\sum y_i a_i \cdot (x_i, x) + b^*) \quad (3)$$

针对于非线性分类问题,可以通过某种非线性变换将其转化到高维空间中,在高维特征空间中学习低维空间的分类法则。由于在线性支持向量机学习的对偶问题里,目标函数和分类决策函数都只涉及实例和实例之间的内积,所以不需要显式地指定非线性变换,而是用核函数替换当中的内积。核函数表示,通过一个非线性转换后的两个实例间的内积。具体地, $K(x, z)$ 是一个函数,或正定核,意味着存在一个从输入空间到特征空间的映射 $\theta(x)$,使得对任意输入空间中的 (x, z) 有:

$$K(x, z) = \Phi(x) \cdot \Phi(z) \quad (4)$$

在线性支持向量机学习的对偶问题中,用核函数 $K(x, z)$ 替代内积,求解得到的就是非线性支持向量机如下:

$$g(x) = \text{sign}\left(\sum_{SVM} y_i a_i \cdot K(x, z) + b^*\right) \quad (5)$$

本文使用的核函数为高斯核函数,其形式为:

$$K(x, x') = \exp(-r \|x - x'\|^2) \quad (6)$$

3.4 基于像元的分类与面向对象分割的集成方法

除传统的基于像元和面向对象的分类方案外,还提出改进的集成方法,即将基于像元的分类和面向对象的分割相结合^[24],以求探索一种分类效果较好且运算量较小的分类方案。

首先,利用随机森林和SVM等分类算法对影像进行基于像元的分类,每个像元都被标记为一个类别;然后,基于多尺度分割方法对影像进行分割,得到的对象均包含一组具有唯一ID(区域标签)的像元。为合并基于像元的分类和面向对象的分割结果,根据以下规则重新分配每个对象内的像元类别:

如果一个对象内部大于阈值A的像素被归为某种地物,则整个对象被分类为该地物。通过统计分割尺度为25的所提取的训练样本中,基于像元进行分类得到的各地类像素所占的百分比,确定本实验阈值为80%。

如果不满足以上条件,则基于像元的分类结果在最终分类结果图中将保持不变,从而在一个对象中产生不同地物类别的像元。

这样可以解决农田异质性和相邻像元间的光谱污染带来的作物像元的缺失,同时可以提升运算效率。

3.5 精度评价

混淆矩阵(Confusion Matrix)主要用于比较分类结果和实际测量值之间的混淆程度进行精度评估^[25]。本文对各个方案进行评价时使用的客观评价指标主要包括总体精度、Kappa系数、生产者精度和用户精度。

4 结果与分析

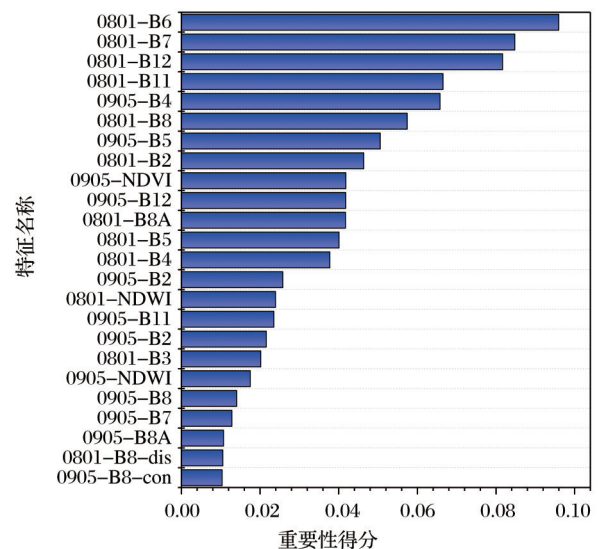
4.1 特征优选分析

本文使用随机森林中所有决策树得到的平均不纯度衰减来度量特征的重要性^[26],选取重要性等级排名靠前的24个特征生成像元特征集重要性分布图(图3(a)),选取重要性等级排名靠前的35个特征生成对象特征集(分割尺度为25)重要性分布图(图3(b))。其中,纵轴为特征名称,横轴为重要性得分(所有特征的重要性得分之和为1)。

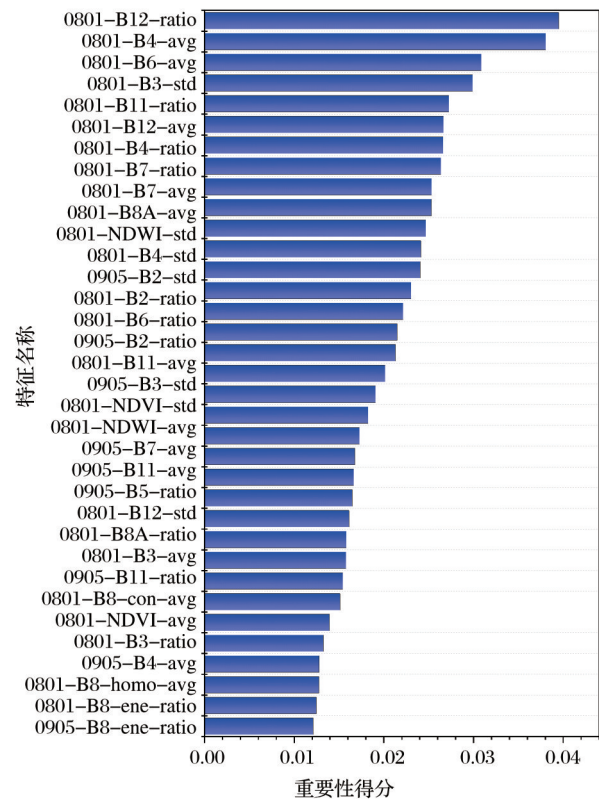
像元特征集中得分排名前4的特征为0801-B6、0801-B7、0801-B12、0801-B11,表明红边波段和短波红外波段为最具判别效果的特征,此时马铃薯和胡萝卜正处于块茎膨大期,莜麦处于抽穗期,在此作物生长期提取的光谱特征在判别效果上优于成熟期。纹理特征重要性排名较低,均低于0.01,这表明基于 Sentinel-2 影像进行作物提取时,加入纹理特征对分类效果提升效果较小,这是由于10 m分辨率在提取不同作物的纹理特征时较为粗糙,提取效果低于高分一号(GF-1)、高分二号(GF-2)、航空影像等高分辨率影像。排名前12个特征的重要性之和为0.69,前24个特征的重要性之和为0.96,因此将前24个特征作为方案一随机森林分类、SVM分类和方案二基于对象的多尺度分割中的输入数据,并按照重要性得分赋予多尺度分割输入特征相应权重(表6)。对象特征集中,排名前35个特征的重要性之和为0.75,前72个特征的重要性之和为95%,因此将前72个特征作为方案二随机森林分类、SVM分类的输入特征,且第28、33、34、35个特征为纹理特征,这表明基于对象的特征提取提升了纹理特征的判别前效果。

4.2 面向对象的分割尺度分析

为得到最优的分类效果,本文所提出的方案二、方案三均需要确定面向对象的最优分割尺度,因此,在使用 eCognition 软件进行多尺度分割时,通过目视观察,初步确定尺度参数为25、50、100。为更加直观地呈现分割尺度对分类效果的影响,统计不同分割尺度进行分类得到的制图精度和用户精度与各地物类别的关系,得到方案二、方案三各尺度地物分类精度分布图(图4)。考虑到闪电河流域



(a) 像素特征重要性分布



(b) 对象特征重要性分布

图3 特征重要性分布

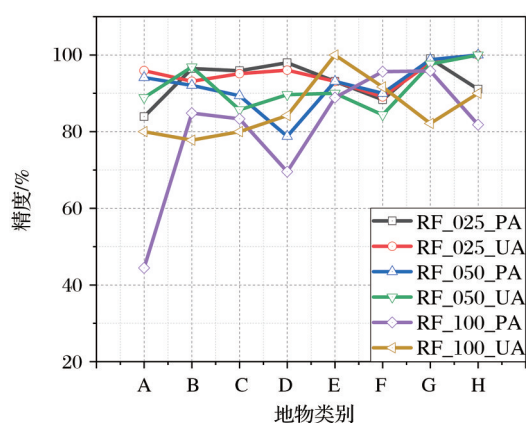
Fig.3 Feature importance distribution

主要作物为土豆、胡萝卜、莜麦,除作物以外的主要地物为草地,在进行确定最优尺度时,将优先考虑这4种地物的分类精度。

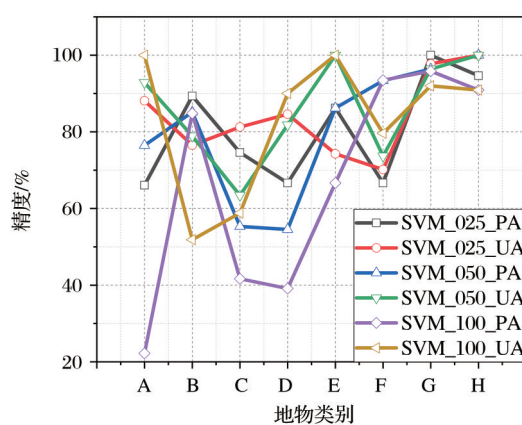
表6 对象分割输入特征权重对应表

Table 6 Feature weight of object segmentation

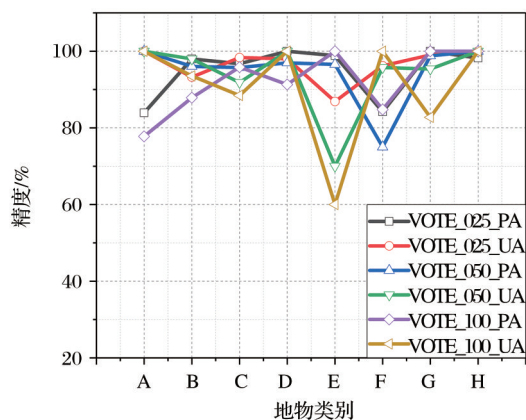
重要性得分	0.00~0.02	0.02~0.03	0.03~0.05	0.05~0.07	0.07~0.09
特征权重	1	2	4	6	8



(a) 方案二随机森林分类方法精度统计



(b) 方案二支持向量机分类方法精度统计



(c) 方案三集成法精度统计

(横坐标A、B、C、D、E、F、G分别表示玉米、土豆、胡萝卜、苜蓿、林地、草地、建设用地、荒地)

图4 不同方案精度统计图

Fig.4 Precision statistics of different schemes

方案二对两种分类器分别进行实验。图4(a)表明,使用随机森林分类器,应用尺度为25时,3种主要作物(土豆、胡萝卜、苜蓿)的分类效果最好,均超过了90%;草地的分类精度略低于尺度为100的分类效果,接近85%;其他地类的分类精度也都高于80%。因此,在该子实验中,最佳分割尺度确定为25。图4(b)表明,使用SVM分类器,应用100和

50两种尺度时,土豆和草地两种地类分类效果最好,均大于80%,但胡萝卜和苜蓿两种地类的制图精度均低于60%;而应用尺度为25时,除玉米外其他地类的精度均大于70%。考虑整体的分类效果,该子实验的最优尺度为25。

由于方案一在进行基于像元的分类时,随机森林有更好的分类效果及分类性能,为简化实验步骤,在确定方案三的最优分割尺度时,只对随机森林的分类结果进行分析。图4(c)表明,应用不同的分割尺度,4种主要地物均有良好的分类效果;在提取树林这一地类时,分割尺度为25的分类效果更好,用户精度远高于其他尺度。因此,方案三的最优尺度为25。

4.3 分类结果分析与精度评价

本文使用3种不同分类方案进行的5个子实验的分类结果如图5和图6所示,通过目视解译可以看出,方案一(图5(a)、5(b)、6(a)、6(b))椒盐噪声较多,种植单一作物的田块内部有很多单个像素被错分为草地或其他作物;有大量低矮稀疏的草原灌木丛被分为林地(如图5(a)中的方框所示)。方案三(图5(e)和方案一(图5(a))相比,田块内部的椒盐噪声大量减少,但更多含灌木的草地被误分为林地((如图5(e)中的方框所示))。对5个子实验中结果差异较大的区域与Sentinel-2原始影像和实地调查数据进行粗略的目视对比,可以看出方案一子实验b(图5(b)、6(b))和方案二(图5(c)、5(d)、6(c)、6(d))的分类效果不理想,存在着明显的错分漏分现象,如:大量草地被误分为林地,大量玉米、苜蓿、胡萝卜被错分为其他作物(如图5(b)、5(c)、5(d)中的方框所示)。从目视效果来看,方案一子实验a(图5(a)、6(a))和方案三子实验e(图5(e)、6(e))的制图效果较好。

为了定量准确地评价不同方案及其子实验对农作物提取精度的影响,利用验证数据,对不同方案进行精度评价分析,其中,方案2选取最优分割尺度25进行实验,方案三选取最优分割尺度25并应用随机森林分类器进行实验。主要评价指标包括总体精度、Kappa系数、生产者精度以及用户精度,具体结果见表7,其中方案子实验名称与表4中相同。

由表7可知,提取像元特征并使用随机森林分类器进行作物提取(方案一子实验a)取得了最优分类精度,总精度为97.8%,Kappa系数为0.97;提取对象特征并使用SVM分类器进行作物提取(方案

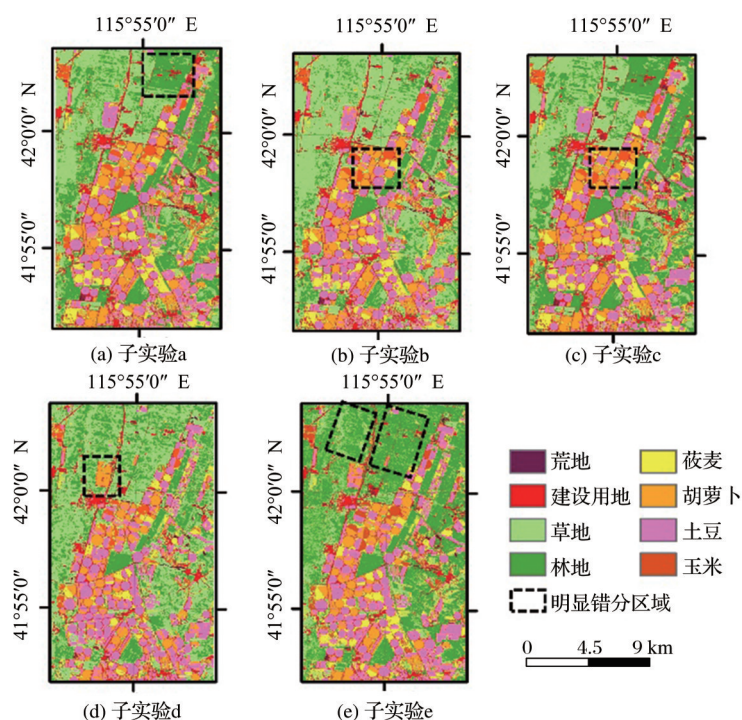


图 5 闪电河流域研究区内不同分类方案局部地区作物分布图 (2018 年)

Fig.5 Crop distribution map of different classification schemes in the study area of Shandian river basin (2018)

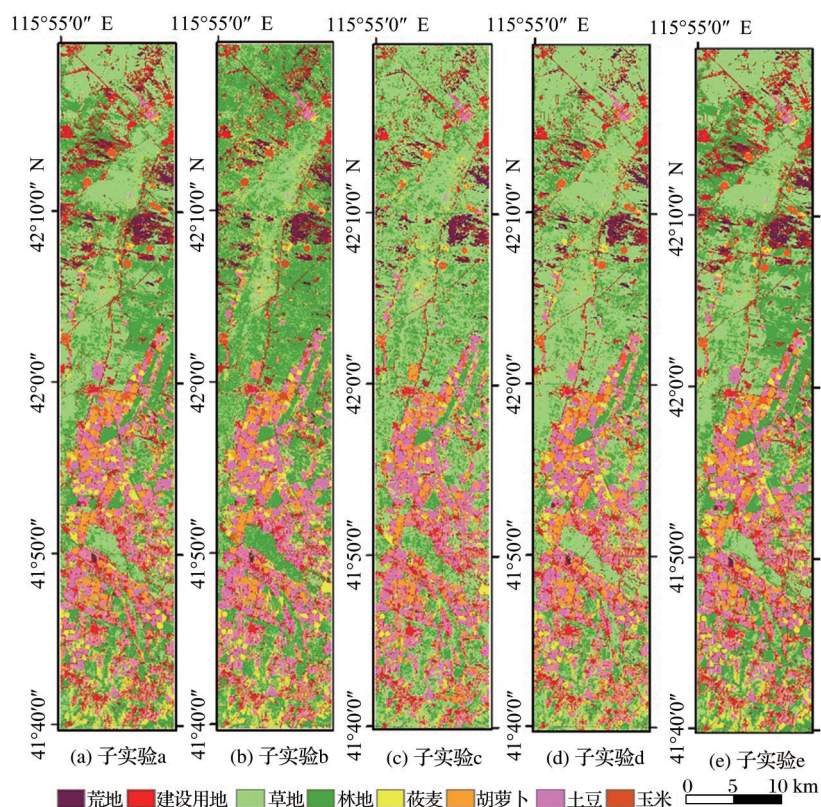


图 6 闪电河流域研究区内不同分类方案作物分布图 (2018 年)

Fig.6 Crop distribution map of different classification schemes in the study area of Shandian river basin (2018)

二子实验 d) 精度最差, 总精度仅为 84.8%, Kappa 系数为 0.82。当对相同的特征使用不同的分类器进行分类时, 使用随机森林分类器可取得更高的分类精

度: 在基于像元的分类中 (方案一子实验 a), 比 SVM 分类器的分类精度提高 6.3%, Kappa 系数提高了 0.08; 在面向对象的分类中 (方案二子实验 c),

其分类精度提升了 10.4%, Kappa 系数提高了 0.13。这说明随机森林分类器更适合进行高维小样本的多类别农作物提取, 且运算量远远小于 SVM, 更适合进行大面积的作物提取。

当使用相同的分类器时, 基于像元的方法分类精度均优于面向对象的分类方法。这说明, 由于闪电河流域农田田块规整, 农业规模经营发展较成熟, 在使用 Sentinel-2 数据进行农作物提取时, 像元特征即可具有较好的表达效果; 且基于像元进行分类时, 可获得更多的训练样本, 这对于提升分类器的分类效果具有很重要的作用。

分割尺度为 25、使用随机森林作为分类器的方

案三子实验 e 分类精度仅次于使用相同分类器的方案一, 高于方案二, 总精度为 96.4%, Kappa 系数为 0.96。从单个类别的分类效果来看, 方案三在居民地和裸土这两类地物的提取中有最佳分类效果; 提取玉米效果最差, 制图精度仅为 83.9%; 除玉米外其他作物的分类效果均接近最佳分类精度。由此可以看出, 方案三在提取呈自然分布的细碎地物有优势; 在提取农作物时基本效果同方案一相差不大, 但受分割尺度和阈值的影响, 对于非大规模种植的作物(如玉米), 会由于分割尺度较大或阈值较低出现错分; 且由于其样本数量较大, 分类效果明显优于方案二。

表 7 闪电河流域研究区内作物分类精度统计

Table 7 The statistics of classification accuracy in Shandian river basin study area

类别	方案一				方案二				方案三	
	子实验 a		子实验 b		子实验 c		子实验 d		子实验 e	
	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%
玉米	92.28	99.83	79.57	96.20	83.93	95.92	66.07	88.10	83.93	100
土豆	99.11	97.17	96.94	92.11	96.45	93.14	89.34	76.52	97.97	93.24
胡萝卜	99.22	98.65	94.11	94.78	95.90	95.12	74.59	81.25	96.72	98.33
莜麦	99.58	99.53	91.26	94.46	97.98	96.04	66.67	84.62	100	98.02
树林	99.67	95.57	97.57	86.02	93.10	93.10	86.21	74.26	98.85	86.87
草地	87.08	98.00	54.57	76.33	88.33	89.08	66.67	70.18	84.17	96.19
居民地	99.18	99.03	97.39	97.34	99.02	98.05	100	97.76	100	99.03
裸土	98.43	99.60	96.99	97.63	91.07	100	94.64	100	98.21	100
总精度%	97.82		91.55		95.11		84.74		96.35	
Kappa 系数	0.974		0.897		0.941		0.815		0.957	

注: 表中 PA、UA 分别代表各类分类的生产者精度和用户精度

5 结 论

本文基于 Sentinel-2 数据, 采用 3 种分类方案, 结合随机森林和 SVM 分类器对闪电河流域农作物进行提取, 主要得出以下结论:

随机森林分类器相对于 SVM 分类器, 在基于 Sentinel-2 影像对闪电河流域进行多种农作物提取时, 可以发挥更好的分类效果, 对比方案一和方案二的各子实验发现, 采用随机森林进行分类后, 分类精度有大幅改善, 并且其分类精度受尺度的影响较小, 以上分析说明: 随机森林分类器受样本数量影响更小、在样本数量不足的情况下分类精度保持较高水平, 且具有模型运算速度快的优点, 因此更适用于中高分辨率遥感影像基于小样本的多类作物提取。

对于闪电河流域, 部分较为细碎的地物(如草地、林地、荒地、建设用地), 地物特征提取受面向对

象方法中的分割尺度影响较大; 由于该地区种植模式较为成熟, 单个地块内作物种类单一且面积较大, 因此作物特征提取受分割尺度影响较小, 但是随着分割尺度变大, 样本对象变少, 因此当尺度变大时, 作物的分类精度仍迅速降低, 以上分析说明, 在样本数量不足的情况下, 应尽可能使用基于像元方法或面向对象方法小分割尺度进行分类。

对比分析各实验方案, 改进的基于像元分类与面向对象分割相结合的集成方法充分结合前两种方案的优点: 使用基于像元的方法提取地物特征进行初步分类, 保证了训练样本的数量, 缩短了模型运行时间; 面向对象分割后进行像素聚合, 可有效减少基于像元分类的椒盐噪声, 提升田块边缘的分类效果。以上分析说明, 该集成方法可有效提升闪电河流域的作物分类效果。

本研究的主要优势在于对新数据源 Sentinel-2

数据的引入以及不同特征提取方法、分类方案的评价分析,并且本文也验证了新的集成分类方案的可行性。同时,本文对闪电河流域的作物分布研究也可为该地区遥感综合实验提供宝贵的基础数据。但在后续研究中仍需在以下方面进行改进:①对于作物分类,样本的数量和质量是制约分类精度的重要因素,本次实验中由于样本数量不足影响了面向对象方法的分类精度,未来应通过进一步地面调查,扩大地面样本采集范围,为样本库扩容。②在特征选取方面,不同的特征组合对各作物的提取效果有很大影响,这方面缺乏深入的研究。③受数据质量(包括云覆盖)的影响,Sentinel-2数据可选取的时相较少,未充分利用时相特征对作物进行分类,后续的工作重点将主要研究如何利用多源数据的多时相信息进行闪电河流域的农作物分类。

参考文献(References):

- [1] Chen Zhongxin, Ren Jjianqiang, Tang Huajun, *et al.* Progress and Perspectives on Agricultural Remote Sensing Research and Applications in China[J]. *Journal of Remote Sensing*, 2016, 20(5): 748-767. [陈仲新, 任建强, 唐华俊, 等. 农业遥感研究应用进展与展望[J]. *遥感学报*, 2016, 20(5): 748-767.]
- [2] Zhao Chunjiang. Advances of Research and Application in Remote Sensing for Agriculture[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(12): 277-293. [赵春江. 农业遥感研究与应用进展[J]. *农业机械学报*, 2014, 45(12): 277-293.]
- [3] Wu Bingfang, Zhang Miao, Zeng Hongwei, *et al.* Agricultural Monitoring and Early Warning in the Era of Big Data[J]. *Journal of Remote Sensing*, 2016, 20(5): 1027-1037. [吴炳方, 张森, 曾红伟, 等. 大数据时代的农情监测与预警[J]. *遥感学报*, 2016, 20(5): 1027-1037.]
- [4] Meng Jihua, Wu Bingfang, Du Xin, *et al.* A Review and Outlook of Applying Remote Sensing to Precision Agriculture[J]. *Remote Sensing for Land and Resources*, 2011, 23(3): 1-7. [蒙继华, 吴炳方, 杜鑫, 等. 遥感在精准农业中的应用进展及展望[J]. *国土资源遥感*, 2011, 23(3): 1-7.]
- [5] Patil P, Murali K G, Gumma, *et al.* A Review of the Available Land Cover and Cropland Maps for South Asia[J]. *Agriculture*, 2018, 8(7): 111. doi:10.3390/agriculture 8070111.
- [6] Song Qian, Zhou Qingbo, Wu Wenbin, *et al.* Recent Progresses in Research of Integrating Multi-source Remote Sensing Data for Crop Mapping[J]. *Scientia Agricultura Sinica*, 2015, 48(6): 1122-1135. [宋茜, 周清波, 吴文斌, 等. 农作物遥感识别中的多源数据融合研究进展[J]. *中国农业科学*, 2015, 48(6): 1122-1135.]
- [7] Liu Zhe, Liu Diyou, Zhu Dehai, *et al.* Review on Crop Type Fine Identification and Automatic Mapping Using Remote Sensing[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(12): 1-12. [刘哲, 刘帝佑, 朱德海, 等. 作物遥感精细识别与自动制图研究进展与展望[J]. *农业机械学报*, 2018, 49(12): 1-12.]
- [8] Liu Bingjing, Yang Yanzhao, Li Yi. Quantitative Analysis of Land Use Structure Characteristics over the Farming—Pastoral Zone in the West Liaohe River basin, Northern China[J]. *Journal of Arid Land Resources and Environment*, 2018, 32(6): 64-71. [刘冰晶, 杨艳昭, 李依. 北方农牧交错带土地利用结构特征定量研究——以西辽河流域为例[J]. *干旱区资源与环境*, 2018, 32(6): 64-71.]
- [9] Han Tao, Pan Jianjun, Zhang Peiyu, *et al.* Study on Differences between Sentinel-2A and Landsat 8 Images in Rape identification[J]. *Remote Sensing Technology and Application*, 2018, 33(5): 890-899. [韩涛, 潘剑君, 张培育, 等. Sentinel-2 与 Landsat 8 影像在油菜识别中的差异性研究[J]. *遥感技术与应用*, 2018, 33(5): 890-899.]
- [10] Chen Mingye, Chen Lei, Zhou Xun. A Remote Sensing Study of Spatio-temporal Changes of Ecological Environment of Shandian River[J]. *Remote Sensing for Land and Resources*, 2017, 29(4): 166-172. [陈明叶, 陈磊, 周询. 闪电河上游生态环境时空变化遥感研究[J]. *国土资源遥感*, 2017, 29(4): 166-172.]
- [11] Zhao Y, Potgieter A B, Zhang Miao, *et al.* Predicting Wheat Yield at the Field Scale by Combining High-resolution Sentinel-2 Satellite Imagery and Crop Modelling[J]. *Remote Sensing*, 2020, 12(6): 1024. doi:10.3390/rs12061024.
- [12] Drusch M, Del Bello U, Carlier S, *et al.* Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services[J]. *Remote Sensing of Environment*, 2012, 120: 25-36. doi: 10.1016/j.rse.2011.11.026.
- [13] Li Qiliang, Fan Jinlong, Xu Qi, Liu Shaojie, *et al.* Ground Sample Survey based on GPS Photos Processing System[J]. *Journal of Geomatics*, 2019(3): 113-116. [李启亮, 范锦龙, 许淇, 等. 基于GPS照片数据处理系统的地面样方调查[J]. *测绘地理信息*, 2019(3): 113-116.]
- [14] Zhang X, Sun Y L, Shang K, *et al.* Crop Classification based on Feature Band Set Construction and Object-Oriented Approach Using Hyperspectral Images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, 9(9): 4117-4128. doi:10.1109/JSTARS.2016.2577339.
- [15] Zhou Zhuang, Li Shengyang, Zhang Kang, *et al.* Crop Mapping Using Remotely Sensed Spectral and Context Features based on CNN[J]. *Remote Sensing Technology and Application*, 2019, 34(4): 694-703. [周壮, 李盛阳, 张康, 等. 基于CNN和农作物光谱纹理特征进行作物分布制图[J]. *遥感技术与应用*, 2019, 34(4): 694-703.]
- [16] Song Q, Zhou Q B, Wu W B, *et al.* Mapping Regional Cropping Patterns by using GF-1 WFV Sensor Data[J]. *Journal of Integrative Agriculture*, 2017, 16(2): 337-347. doi: 10.1016/

- S2095-3119(16)61392-8.
- [17] Zhang Lei, Gong Zhaoning, Wang Qiwei, *et al.* Wetland Mapping of Yellow River Delta Wetlands based on Multi-feature Optimization of Sentinel-2 Images[J]. Journal of Remote Sensing, 2019, 23(2): 313-326. [张磊, 宫兆宁, 王启为, 等. Sentinel-2 影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 23(2): 313-326.]
- [18] Du S H, Guo Z, Wang W Y, *et al.* A Comparative Study of the Segmentation of Weighted Aggregation and Multiresolution Segmentation[J]. Giscience & Remote sensing, 2016, 53(5): 651-670.
- [19] Shi Peirong, Chen Yongfu, Liu Hua, *et al.* Parameters of Multi-segmentation based on Segmentation Evaluation Function[J]. Remote Sensing Technology and Application, 2018, 33(4): 628-637. [施佩荣, 陈永富, 刘华, 等. 基于分割评价函数的多尺度分割参数的选择[J]. 遥感技术与应用, 2018, 33(4): 628-637.]
- [20] Ma L, Li M C, Ma X X, *et al.* A Review of Supervised Object-based Land-cover Image Classification[J]. Elsevier Bes-loten Vennootschap, 2017, 130. doi: 10.1016/j.isprsjprs.2017.06.001.
- [21] Breiman L. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32. doi: 10.1023/A:1010933404324.
- [22] Belgiu M, Drăguț L. Random Forest in Remote Sensing: A Review of Applications and Future Directions[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24-31. doi: 10.1016/j.isprsjprs.2016.01.011.
- [23] Mountrakis G, Im J, Ogole C. Support Vector Machines in Remote Sensing: A Review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 66(3): 247-259. doi: 10.1016/j.isprsjprs.2010.11.001.
- [24] Xiong J, Thenkabail P S, Tilton J C, *et al.* Nominal 30 m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-based Algorithms Using Sentinel-2 and Landsat 8 Data on Google Earth Engine[J]. Remote Sensing, 2017, 9(10): 1065. doi: 10.3390/rs9101065.
- [25] Li Xiang, Liu Kai, Zhu Yuanhui, *et al.* Study on Mangrove Species Classification based on ZY-3 Image[J]. Remote Sensing Technology and Application, 2018, 33(2): 360-369. [李想, 刘凯, 朱远辉, 等. 基于资源三号影像的红树林物种分类研究[J]. 遥感技术与应用, 2018, 33(2): 360-369.]
- [26] Carolin S, Anne-Laure B, Achim Z, *et al.* Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution[J]. BMC Bioinformatics, 2007, 8: 25. doi: 10.1186/1471-2105-8-25.

Sentinel-2 Study on Crop Mapping of Shandian River Basin based on Images

Yin Yanmin, Jia Li

(State Key Laboratory of Remote Sensing Science, Institute of Aerospace Information Innovation, Chinese Academy of Sciences, Beijing 100101)

Abstract: In this study, Sentinel-2 data combined with Random Forest method (RF) and Support Vector Machine method (SVM) were used to extract crop information in the Shandian River Basin in Inner Mongolia. Three schemes are proposed: pixel-based classification algorithm, object-based classification algorithm and improved integration algorithm based on pixel-based classification and object-based segmentation. Results are as follows: (1) pixel-based classification with RF gets the best extraction accuracy, with an overall accuracy up to 97.8% and Kappa coefficient of 0.974. This result shows that RF has evident advantages in crop extraction. (2) The improved integration algorithm also has good extraction accuracy. The overall accuracy is 96.4%, and kappa coefficient is 0.957. This method fully combines the advantages of pixel-based and object-based classification methods, which effectively improves the crop classification effect in Shandian River region.

Key words: Sentinel-2; Crop extraction; Pixel-based classification; Object-based classification; RF; SVM