

基于时序 Sentinel-2 影像的现代农业园区作物分类研究

张东彦¹, 戴震^{1,2}, 徐新刚^{2*}, 杨贵军², 孟炀², 冯海宽², 洪琪¹, 姜飞^{1,3}

(1. 安徽大学农业生态大数据分析与应用技术国家地方联合工程技术研究中心, 安徽合肥 230601;
2. 北京农业信息技术研究中心, 北京 100097; 3. 宿州学院信息工程学院, 安徽宿州 234000)

摘要: 快速、准确地掌握作物空间分布, 估算不同作物种植面积及范围, 这对制定宏观农业政策并指导农民进行农业生产具有重要意义。以我国内蒙古自治区扎赉特旗现代农业示范园区为研究区域, 基于 2019 年 5 月至 10 月共 9 景多时相 Sentinel-2 卫星遥感影像, 通过计算并分析不同作物归一化差值植被指数 (NDVI)、比值植被指数 (RVI)、增强型植被指数 (EVI) 等多种典型植被指数和近红外波段 Ref(NIR) 的时序变化特征, 采用随机森林 (Random Forest, RF)、决策树 (Decision Tree, DT)、支持向量机 (Support Vector Machine, SVM) 和最大似然法 (Maximum Likelihood, ML) 4 种分类方法对研究区多种作物进行分类识别, 成功提取园区内主要作物 (水稻、玉米、甜叶菊、旱稻和大豆等) 空间分布情况。将 RF 结果与 DT、SVM 和 ML 分类结果对比, 结果显示, RF 总体分类精度最高, 达到 95.8%, Kappa 系数为 0.944; DT、SVM 和 ML 分类精度分别为 92.2%、91.6% 和 86.5%。上述研究结果表明, 多时相 Sentinel-2 遥感影像经过光谱指数时序变化特征提取后, 利用随机森林算法进行作物分类可得到精度较高的结果, 这为精细指导规模化园区农业生产提供了有效的技术支持。

关键词: 随机森林算法; 近红外波段; 时间序列; Sentinel-2; 作物分类

中图分类号: TP79; S127 **文献标志码:** A **DOI:** 10.3788/IRLA20200318

Crop classification of modern agricultural park based on time-series Sentinel-2 images

Zhang Dongyan¹, Dai Zhen^{1,2}, Xu Xingang^{2*}, Yang Guijun²,
Meng Yang², Feng Haikuan², Hong Qi¹, Jiang Fei^{1,3}

(1. National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei 230601, China;
2. Beijing Agricultural Information Technology Research Center, Beijing 100097, China;
3. School of Information Engineering, Suzhou University, Suzhou 234000, China)

Abstract: Quickly and accurately grasping the spatial distribution of crops, estimating the area and scope of different crops were of great significance for the country to formulate macroscopic agricultural policies and guide farmers in agricultural production. To explore an efficient and accurate crop classification method, this paper took the agricultural area of Jalaid Banner of Hinggan League in Inner Mongolia Autonomous Region of China as the study area and extracted main crop classification based on the Sentinel-2 satellite remote sensing image data from

收稿日期: 2020-12-07; 修订日期: 2021-01-19

基金项目: 安徽省科技重大专项 (18030701209); 国家重点研发计划 (2017YFD0201501); 现代农业产业技术体系建设专项 (CARS-03); 农业生态大数据分析与应用技术国家地方联合工程研究中心开放课题 (AE2019011)

作者简介: 张东彦, 男, 教授, 博士生导师, 博士, 主要从事计算机视觉、机器/深度学习、光谱信息处理及大数据、物联网等技术在农业和环境监测上的应用研究。

戴震, 男, 硕士生, 主要从事农业遥感与信息技术方面的研究。

通讯作者: 徐新刚, 男, 研究员, 博士, 主要从事农业定量遥感应用基础方面的研究。

May to October 2019. By analyzing the time-series curves of the four characteristic indexes of NDVI, RVI, EVI and Ref (NIR) in the study area, a total of four classification methods including Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and Maximum Likelihood (ML) were used to classify various crops in the study area. The RF results were compared with the classification results of DT, SVM and ML, and the spatial distribution of major crops such as rice, corn, stevia, dry rice and soybean were successfully extracted and identified. The results showed that RF had the highest overall classification accuracy of 95.8% with a Kappa coefficient of 0.944, DT, SVM and ML had classification accuracy of 92.2%, 91.6% and 86.5%, respectively. The above results indicate that the multitemporal Sentinel-2 remote sensing images can be extracted by spectral index time-varying features, and the crop classification using the random forest algorithm can obtain high accuracy results, which provides effective technical support for the fine guidance of large-scale agricultural production in the park.

Key words: random forest; Ref (NIR); time-series; Sentinel-2; crop classification

0 引言

粮食安全问题一直是国内外关注的热点。作物的种植范围、粮食产量等数据是农业生产的重要指标,是国家制定粮食政策和国民经济发展计划的重要依据。作物遥感分类识别是提取作物种植面积分布、监测作物长势和进行作物估产的前提和基础^[1-3]。

目前,作物遥感分类方法众多,优势各异,应用性不同。近年来,卫星遥感数据结合机器学习算法进行作物分类识别成为国内外学者的研究热点,常用的分类算法,如决策树 (Decision Tree, DT)^[4-6]、最大似然法 (Maximum Likelihood, ML)^[7]、支持向量机 (Support Vector Machine, SVM)^[8-9] 等,被诸多学者使用。潘琛^[10]等阐述了决策树分类器的结构与理论基础,对决策树算法的发展趋势进行归纳总结,并使用 TM 数据对徐州市地物、地貌色彩的差异进行分类实验,依据感兴趣区类别进行图像分类且分类效果较好。徐新刚^[7]等使用 Quick Bird 高分辨率多光谱遥感影像,采用最大似然法和监督分类方法并结合多尺度遥感影像信息提取原理,对绵阳实验区的冬小麦与油菜等作物进行分类实验,获得较高精度的作物分类结果。谭琨^[8]等在分析 SVM 分类方法的基础上构建了一种基于 SVM 的高光谱遥感图像分类模型,实验结果表明:这种分类器简单高效且分类精度高,是一种有效的遥感图像分类方法。然而,随着土地利用范围不断拓展深化以及区域资源环境实时变化,使得传统分类方法难以满足高精度作物分类的需求。随机森林算法^[11]自 2001 年提出伊始,就以其优越的性能在国外诸多

领域得到广泛应用,相比而言,国内采用该方法对农业耕种区作物分类识别的研究较少。而且,传统的分类方法更多是针对单一时相影像进行分类。鉴于不同作物在同一生育期下容易出现“同物异谱、同谱异物”现象,导致上述方法的分类效果较差。经文献调研,利用多时相影像结合物候信息,可以有效提高作物的分类精度^[12-14]。魏鹏飞^[14]等利用多时相时间序列的高分一号 GF1/WFV 遥感影像数据,结合作物植被指数特征,采用决策树分类方法进行作物提取,相比较传统利用单一时相分类方法有效提高作物分类精度。

综上,笔者所在团队以作物种植种类较为复杂的内蒙古自治区扎赉特旗现代农业示范园为研究区域,探讨多时相 Sentinel-2 卫星影像和随机森林分类的机器学习模型对研究区作物分类效果。评估基于时间序列 Sentinel-2 影像的随机森林算法对农耕地作物分类的适用性,从而为作物精细遥感分类提供可用的技术方法参考。

1 数据源及研究方法

1.1 研究区概况

研究区位于我国内蒙古自治区兴安盟扎赉特旗现代农业示范园 (122.85°~123.35°E, 46.59~46.81°N), 该区域地处大兴安岭南麓向松嫩平原延伸的过度地带,位于黑龙江、吉林、内蒙古三省交接处。该区地势多为平原丘陵,平均海拔 130 m,总面积约为 493.61 km²,研究区概况如图 1 所示。园区所在位置属典型的温带大陆性气候,年平均气温 3.24 °C,年平

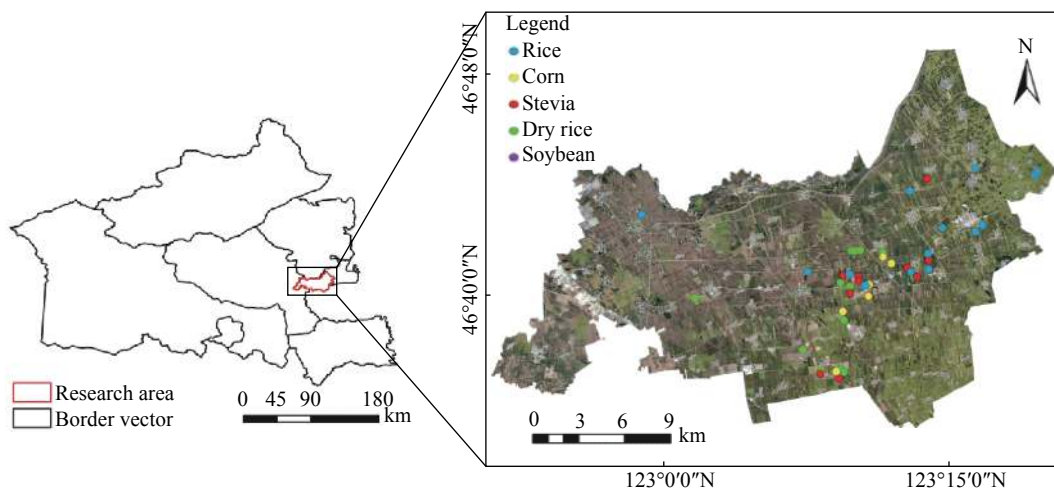


图 1 研究区地理位置及样本分布示意图

Fig.1 Geographical location and sample distributions of the study area

均降水量 400 mm, 无霜期 105~135 天。园区内种植连片的水稻、玉米、甜叶菊、旱稻和大豆, 此外还零散分布裸露的田块、林地和居民点等。对于玉米、水稻和大豆等秋季作物, 其播种和插秧的时间一般在五月

底进行。水稻在插秧前要对田块进行灌溉, 然后再将秧苗移植到田块中, 这使得水稻地块在生长早期具有水田特征。表 1 为水稻、玉米、甜叶菊、旱稻和大豆等五种主要农作物的生育期时间表。

表 1 研究区 5 种主要作物生育期

Tab.1 Growth period of five crops in the study area

| Type | May | | | June | | | July | | | August | | | September | | | October | |
|----------|---------------|---|---|-----------|---|---|-----------|---|---|-----------|---|---|-----------|---|---|----------|---|
| | E | M | L | E | M | L | E | M | L | E | M | L | E | M | L | E | M |
| Rice | Sowing | | | Tillering | | | Heading | | | Filling | | | Maturity | | | | |
| Soybean | Sowing | | | Seeding | | | Flowering | | | Podding | | | Filling | | | Maturity | |
| Stevia | Transplanting | | | | | | Branching | | | Flowering | | | Maturity | | | | |
| Corn | Sowing | | | Seeding | | | Jointing | | | Tasseling | | | Filling | | | Maturity | |
| Dry rice | Sowing | | | Tillering | | | | | | Heading | | | Filling | | | Maturity | |

Note: E means the early 10 days of a month, M is the middle 10 days, and L represents the lately 10 days.

1.2 数据选择及预处理

1.2.1 遥感影像数据

Sentinel-2 遥感卫星于 2015 年 6 月由欧空局成功发射, 该系列卫星具有时空分辨率高、光谱信息丰富, 且获取成本低廉的优点。为农耕区作物分类提供了新的数据源。此研究中 Sentinel-2 遥感影像是通过 ESA 数据共享网站 (<https://scihub.copernicus.eu/>) 免费下载。研究区种植作物结构多样化, 为了充分覆盖作物生长周期, 获取作物有效信息, 选取成像时间为 2019 年 5 月 22 日至 2019 年 10 月 4 日, 共 9 景影像, 影像数据列表如表 2 所示。

表 2 Sentinel-2 影像数据列表

Tab.2 Data lists of Sentinel-2 images

| Data time | Sensor | Quality |
|------------|-------------|---------|
| 2019-05-22 | Sentinel-2A | Best |
| 2019-06-11 | Sentinel-2A | Good |
| 2019-06-21 | Sentinel-2A | Best |
| 2019-07-01 | Sentinel-2A | Good |
| 2019-08-15 | Sentinel-2B | Best |
| 2019-08-30 | Sentinel-2A | Best |
| 2019-09-14 | Sentinel-2B | Best |
| 2019-09-24 | Sentinel-2B | Best |
| 2019-10-04 | Sentinel-2B | Best |

为确保研究区影像质量,所选取影像云量控制在 30% 以内。为了更好地观察研究区作物光谱指数时序变化趋势,保持影像时序性效果,所选取的影像成像时间间隔尽量保持一致。Sentinel-2 遥感卫星搭载的多光谱传感器 (MSI) 可以提供从可见光、近红外到短波红外等 13 个波段的光谱信息,如表 3 所示。包含 10 m、20 m 和 60 m 中、高空间分辨率的遥感影像。此研究选取 10 m 分辨率的红 (Red)、绿 (Green)、蓝 (Blue) 和近红外 (NIR) 4 个波段的影像数据进行分析。

表 3 Sentinel-2 传感器光谱波段 (S2A 和 S2B)

Tab.3 Spectral bands of the Sentinel-2 sensors (S2A & S2B)

| Sentinel-2 bands | Wavelength/ μm | Reflection/m |
|----------------------------|---------------------------|--------------|
| Band1-Coastal aerosol | 0.443 | 60 |
| Band2-Blue | 0.490 | 10 |
| Band3-Green | 0.560 | 10 |
| Band4-Red | 0.665 | 10 |
| Band5-Vegetation red edge | 0.705 | 20 |
| Band6-Vegetation red edge | 0.740 | 20 |
| Band7-Vegetation red edge | 0.783 | 20 |
| Band8-NIR | 0.842 | 10 |
| Band8A-Vegetation red edge | 0.865 | 20 |
| Band9-Water vapour | 0.945 | 60 |
| Band10-SWIR-Cirrus | 1.375 | 60 |
| Band11-SWIR1 | 1.610 | 20 |
| Band12-SWIR2 | 2.190 | 20 |

1.2.2 园区数据采集

外业调查数据:2019 年 6 月和 8 月在研究区进行野外实地调查,利用手持 GPS 采集研究区主要作物的样本点坐标。前后共获取样本点 84 个,其中水稻 11 个、大豆 8 个、甜叶菊 38 个、玉米 2 个以及旱稻 25 个。采集的样本点数据通过软件制作矢量图层,为后续选取样本区提供地面数据支持。

1.2.3 数据预处理

为了提高卫星影像质量,保证分类精度,需要对下载的 Sentinel-2 影像进行预处理。下载的 Level-1C 数据是经过正射校正和几何精校正的大气表观反射率产品,并没有进行大气校正处理。此研究利用 ESA 发布的大气校正处理器插件 Sen2Cor 进行大气校正,11C 级数据大气校正之后最终得到 L2A 级数据,处理后的数据包括水蒸汽图像 (Water vapor)、大

气底层反射率图像 (Bottom of Atmosphere, BOA) 等。大气校正后的 L2A 级数据,使用 SNAP 软件对分辨率为 10 m 的红、绿、蓝和近红波段进行导出,并保存为 ENVI 格式,完成波段融合、影像拼接和研究区裁剪等步骤后,最终得到研究区范围内的 Sentinel-2 影像数据集。

利用野外采集的作物样本点,通过目视解译的方法,随机选择作物感兴趣区域,作为样本集,样本集分布如图 1 所示。通过对样本区域进行统计,总共获取了样本点像元 83279 个,并随机抽取训练样本集 (总样本点 70%),用来对研究区作物进行分类识别,随机抽取验证样本集 (总样本点 30%),用来做分类后的精度验证。

1.3 研究方法

1.3.1 分类指标选择

遥感影像数据中像元光谱值的大小会因为地物类型的改变而产生变化,地物间光谱值的差异性和其独特的光谱特性构成了遥感识别地物的基础^[1]。植被指数根据植被的光谱特性,将卫星可见光和近红外波段进行组合,它在一定条件下能用来定量说明植被的生长状况^[2]。经实验观察,研究区内的作物在不同生长时期,其光谱曲线有明显的变化,为研究分类模型构建提供了有效的理论支持。

此研究通过对比同一时期不同作物的光谱指数特征,找出差异最明显时期,确定作物划分的阈值区间,利用决策树分类模型提取研究区不同作物地块。再选取光谱特征差异性明显的时期,构建光谱指数时序特征影像集,采用随机森林、支持向量机和最大似然法对研究区作物进行分类识别。

此研究选取归一化差值植被指数^[15](Normalized Difference Vegetation Index, NDVI)、比值植被指数^[16](Ratio Vegetation Index, RVI)、增强型植被指数^[17-18](Enhanced Vegetation Index, EVI) 和近红外波段^[19] Ref(NIR) 作为分类指标。其中 NDVI 和 EVI 在植被信息提取和地物识别方面应用最为广泛和有效。水稻的 Ref(NIR) 波段与其他作物光谱差异性最为明显,可以有效区分水稻地块,故引用 Ref(NIR) 波段。旱稻、玉米的 NDVI、EVI 和 Ref(NIR) 时序特征曲线差异性不明显,为更好地识别研究区旱稻、玉米的种植区域,引入 RVI 指数补充。各类指标描述如表 4 所示。

表 4 笔者研究中分类指标列表

Tab.4 Classification indicators used in the study

| Indicator | Description | Source |
|--|---|--------------|
| Normalized Difference Vegetation Index(NDVI) | $NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$ | Ref.[15] |
| Ratio Vegetation Index(RVI) | $RVI = \frac{\rho_{NIR}}{\rho_{RED}}$ | Ref. [16] |
| Enhanced Vegetation Index(EVI) | $EVI = 2.5 \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + 6.0\rho_{RED} - 7.5\rho_{BLUE} + L}$ | Ref. [17-18] |
| Near Infrared Ray(Ref(NIR)) | The reflection of Band-8 in Tab.3 | Ref. [19] |

Notes: In the formula, ρ_{NIR} is the near-infrared band reflectivity, ρ_{RED} is the red band reflectivity, ρ_{BLUE} is the blue band reflectivity and L is the soil adjustment coefficient of 1.

1.3.2 主要分类方法和精度评价

此研究使用到的分类方法有随机森林 (Random Forest, RF)、决策树、支持向量机和最大似然法 4 种。

基于决策树算法改进的随机森林算法拥有广阔的应用前景^[20-22]。它是利用多个决策树对样本进行训练、分类并预测的一种算法,主要应用于回归和分类场景。它的基本单元是决策树,而它的本质属于机器学习的一大分支—集成学习 (Ensemble Learning) 方法。随机森林中随机是核心,通过随机选择样本,降低了决策树之间的相关性。随机在原始训练数据中有放回的选取等量的数据作为训练样本,建立决策树时,随机选取一部分特征建立决策树。从另一个角度解释,每棵决策树都是一个分类器,随机森林生成 N 棵树会有 N 个分类结果。对所有分类器的结果进行投票,投票次数最多的类别为最终结果。

决策树是一种归纳推理的分类方法,通过建立树形结构,在其中每个内部节点表示属性测试,对图像各类信息值进行比较,得出新的分支当作新的节点更新规则,继续分类直到最终得到想要的分类结果。适合于分层分样的样本分类统计工作,而且在一定范围内分类精度很高。

支持向量机是一类按监督学习方式对数据进行二元分类的广义线性分类器,其决策边界是对学习样本求解的最大边距超平面,解决复杂数据分类问题,它适用于高维特征空间、小样本统计学习。

最大似然法也称为最大似似估计或最大似然估计,是求估计的另一种方法。该方法通过对感兴趣区

域里需分类像元的统计,得出均值等参数后,确定一个专门分类的函数,将图像中待分类的像元点带入该分类函数后进行扫描,把函数返回值最大的类别作为被归属类别,通过统计分类像元,得到分类的效果。

此研究采用的随机森林分类模型、决策树、支持向量机和最大似然法 4 种分类方法都是基于 ENVIS.3 来实现。通过对比 4 种方法的分类结果,突出随机森林分类的优势。对研究区作物分类识别后,利用随机选取的验证样本集,通过计算混淆矩阵,得出每种作物的制图精度和用户精度,并将 4 种分类方法的总体分类精度和 Kappa 系数作为分类结果的精度评价指标,精度验证的计算公式如表 5 所示。

表 5 精度验证的计算公式

Tab.5 Formulas of accuracy evaluation

| Type | Calculation formula |
|-------------------|---|
| Mapping accuracy | $PA = \frac{X_{ii}}{X_j} 100\%$ |
| User accuracy | $UA = \frac{X_{ii}}{X_i} 100\%$ |
| Overall accuracy | $OA = \sum_{i=1}^k \frac{X_{ii}}{N} 100\%$ |
| Kappa coefficient | $K = \frac{N \sum_{i=1}^k X_{ii} - \sum_{i=1}^k X_i X_j}{N^2 - \sum_{i=1}^k X_i X_j}$ |

Notes: where k represents the number of rows and columns of the confusion matrix, X_{ii} represents the value on the diagonal of the confusion matrix that is the number of pixels correctly classified, N represents the total number of pixels verified, X_i represents the i row of the confusion matrix. The sum of elements, X_j represents the sum of elements in the j column of the confusion matrix.

2 结果与分析

2.1 光谱时序特征分析

2.1.1 主要地物 NDVI 特征分析

研究区主要作物的 NDVI、EVI、RVI 和 Ref(NIR) 时序特征曲线如图 2 所示。由图 2(a) 可以看出水体、道路和居民区等非耕地的 NDVI 时间序列曲线处于稳定状态, 平均值在 0.10~0.20 之间, 在 8 月和 9 月与其他作物的值差异最为明显, 通过实验发现 8 月 15 日影像的 NDVI 值区分非耕地效果较好, 此时非耕地的 NDVI 最大、最小值分别为 0.75 和 -0.20, 通过与其他作物相同时期 NDVI 最值对比, 最后设置阈值 $NDVI_{8/15} > 0.48$ 将水体、道路及居民区剔除掉。5 月份研究区作物处于幼苗阶段, 植物地表覆盖度低, 此时林地生长较为茂盛, 其 NDVI 值明显高于其他作物, 可以利用林地 NDVI 值, 将其剔除。5 月 22 日林地 NDVI 最小值为 0.63, 大于同时期其他作物 NDVI 值,

因此设置阈值 $NDVI_{5/22} \geq 0.63$, 可以将林地剔除。将识别出的水体、道路、居民区和林地地块进行掩膜, 方便后续进行研究区主要作物的分类识别。

2.1.2 主要作物光谱特征分析

从图 2 可以看出, 作物前期生长较为缓慢, 光谱指数特征差异不明显。进入 6 月下旬, 作物对叶绿素吸收能力开始增强, 生长迅速, Ref(NIR) 逐渐增加, 作物长势明显, 植被指数快速增长且光谱指数特征增长趋势较为一致。玉米、水稻和大豆在 7、8 月份, 4 种光谱指数达到饱和状态, 并保持稳定。8 月份除了早稻, 其余作物的 Ref(NIR) 达到峰值。甜叶菊在这一时期仍处于生长阶段, EVI 和 Ref(NIR) 变化规律相似, 上升到峰值后, 保持稳定状态。8 月下旬开始, 作物进入灌浆期, 营养器官能量吸收衰减, 吸收的能量更多向生殖器官转移。作物的 NDVI、EVI 和 Ref(NIR) 开始降低。随着水稻、早稻、玉米进入开花期, 大豆进入结荚期, 这 4 种作物的 EVI 和 RVI 值达到最大。

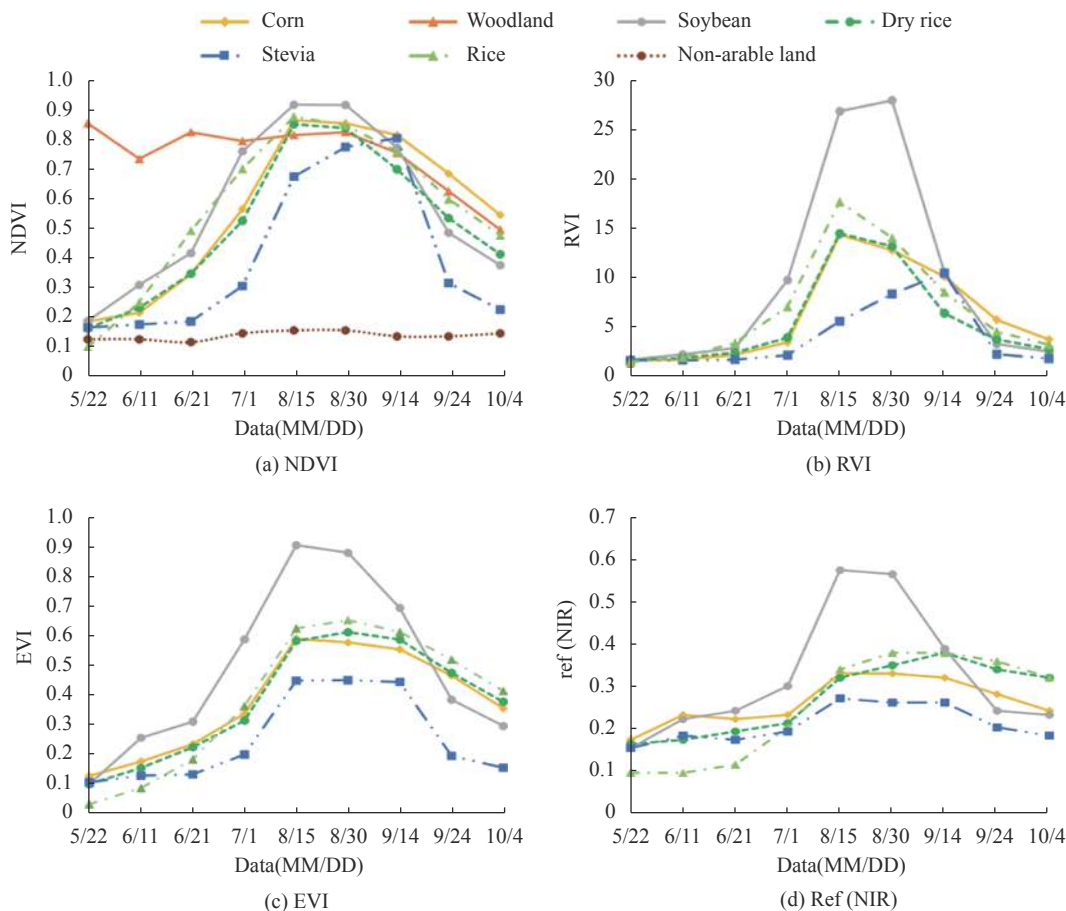


图 2 各作物 NDVI (a)、RVI (b)、EVI (c)、Ref(NIR) (d) 时序变化

Fig.2 Temporal changes of NDVI (a), RVI (b), EVI (c) and Ref(NIR) (d) for main crops

9 月中上旬甜叶菊的 RVI 和 NDVI 时间序列曲线出现峰值,这是因为此时甜叶菊处于开花期,导致甜叶菊 RVI 和 NDVI 峰值出现的时间较为靠后。9 月后,各作物的指数特征都出现明显下降的趋势。

2.2 决策树分类过程

通过对比作物的光谱时间序列曲线变化特征,发现 Ref (NIR) 可以用来区分大豆和水稻,并且对水稻区分效果较好。Ref (NIR) 结合 EVI,可以提高大豆识

别精度。甜叶菊 9 月 14 日和 8 月 15 日的 NDVI、RVI 值做差作为筛选条件,可以较好地识别出甜叶菊地块。玉米和旱稻地块混种现象严重,区分难度较大,用到了 NDVI、RVI、EVI 和 Ref (NIR) 4 种光谱特征指数来识别玉米和旱稻地块。此研究选取基于时间序列影像的多种特征指数,采用决策树方法进行主要作物分类提取过程如图 3 所示。

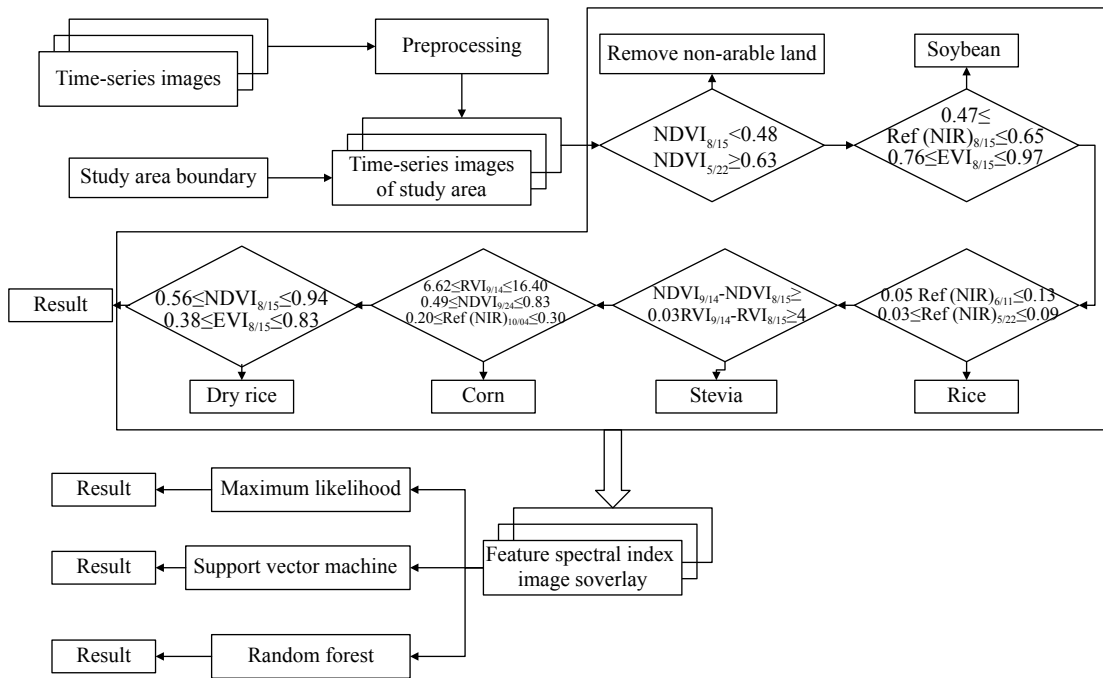


图 3 作物分类提取流程图

Fig.3 Flow chart of crop classification extraction

通过 $NDVI_{8/15} < 0.48$ 剔除水体、道路和建筑等,利用 $NDVI_{5/22} \geq 0.63$ 剔除林地,将这些区域统一划分为非作物区域进行剔除,然后进行主要作物的分类。

大豆在整个生育期内,除了 NDVI 生长曲线相比较其他作物差异性不明显外,其他光谱指数特征曲线生长趋势明显高于其它作物。大豆的 EVI 和 Ref(NIR) 整体趋势相似,与其他作物有一定的区分度,因此选取 EVI 和 Ref (NIR) 对大豆地块进行提取。图 2(c)、(d) 中可以看出大豆的 EVI 和 Ref(NIR) 数值在 8 月 15 日明显高于甜叶菊、旱稻、水稻和玉米,因此选取 $Ref (NIR)_{8/15}$ 和 $EVI_{8/15}$ 用来提取大豆种植区域。大豆 $Ref (NIR)_{8/15}$ 最大值为 0.65,最小值为 0.40,根据样本地块统计 Ref(NIR) 和 EVI 数值,经反复实验最终

设定满足 $0.47 \leq Ref (NIR)_{8/15} \leq 0.65$ 的像元为大豆, $0.76 \leq EVI_{8/15} \leq 0.97$ 的像元作为补充。

水稻、玉米、旱稻生长趋势基本吻合,尤其是水稻和旱稻的差异性不明显。在 6 月份前后,水稻刚开始插秧,水稻田块里有水体存在,使水稻的 Ref (NIR) 数值高于其他地物同时期 Ref (NIR) 数值,因此选择 6 月 11 日的水稻 Ref (NIR) 提取水稻地块。受 6 月 11 日影像质量影响,选取 5 月 22 日水稻的 Ref (NIR) 作为补充。将水稻样本地块在这两个时期的 Ref (NIR) 统计值,与其他地物统计的 Ref (NIR) 最大值进行比较,反复实验,最终确定满足 $0.05 \leq Ref (NIR)_{6/11} \leq 0.13$ 的像元为水稻,并满足 $0.03 \leq Ref (NIR)_{5/22} \leq 0.09$ 的像元为补充。

从甜叶菊的 NDVI 和 RVI 时间序列曲线可以看出, 在 8 月至 9 月其他作物植被指数缓慢降低时, 甜叶菊出现了增长。通过对比不同时期的指数差值, 提取甜叶菊种植区域。经过反复试验, 最终确定满足 $NDVI_{9/14} - NDVI_{8/15} \geq 0.03$ 和 $RVI_{9/14} - RVI_{8/15} \geq 4$ 的像元为甜叶菊。在剔除大豆、水稻、甜叶菊后, 剩下玉米和旱稻两种作物, 两种作物的光谱特征曲线的变化趋势相似, 但在 9 月 14 日玉米的 RVI 值大于旱稻, 9 月

24 日玉米的 NDVI 值明显大于旱稻, 而在 10 月 4 日玉米的 Ref (NIR) 值小于旱稻, 经过反复试验, 最终确定满足 $6.62 \leq RVI_{9/14} \leq 16.40$ 和 $0.49 \leq NDVI_{9/24} \leq 0.83$ 以及 $0.20 \leq Ref(NIR)_{10/04} \leq 0.30$ 的像元为玉米。最后剩下旱稻, 选取旱稻关键生育期 8 月 15 日的 NDVI 和 EVI 指数进行提取, 经过反复试验, 最终选定满足 $0.56 \leq NDVI_{8/15} \leq 0.94$ 和 $0.38 \leq EVI_{8/15} \leq 0.83$ 的像元为旱稻。决策树分类结果如图 4(a) 所示。

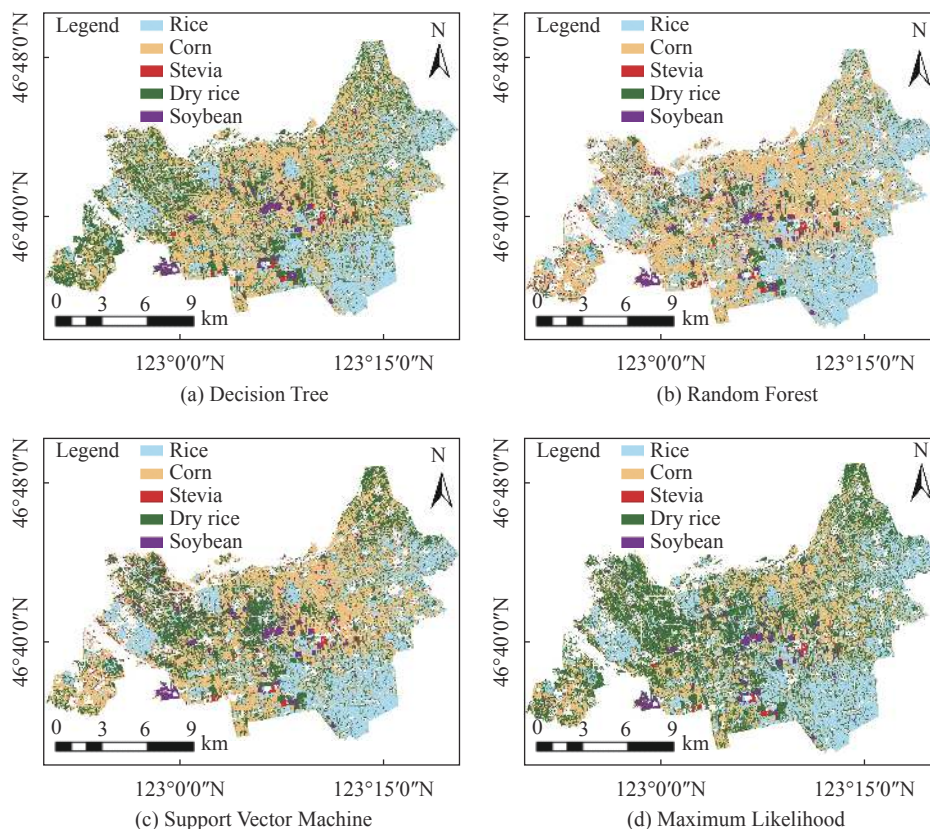


图 4 采用决策树 (a)、随机森林 (b)、支持向量机 (c)、最大似然法 (d) 的作物分类结果

Fig.4 Crop classification results using Decision Tree (a), Random Forest (b), Support Vector Machine (c), Maximum Likelihood (d)

2.3 随机森林模型分类过程

选取作物分类提取过程中关键时期的植被指数和近红外光谱影像, 其中包括 NDVI: 5 月 22 日、8 月 15 日、9 月 14 日和 9 月 24 日 4 个时期影像, Ref(NIR): 5 月 22 日、6 月 11 日、8 月 15 日和 10 月 4 日 4 个时期影像, RVI: 8 月 15 日、9 月 14 日 2 个时期影像, EVI: 8 月 15 日 1 个时期影像。将上述 11 个时期影像叠加。通过 En MAP-BOX^[23] 工具, 将合成的影像与训练样本集输入到随机森林分类器中进行分类, 对于 Random Forest 分类参数设置, 一般保持默认即可, 随

机森林树的数量设置为 100, 特征数量默认使用“Square Root”方法, 最小节点样本选择 1, 设置好参数后, 直接运行分类器, 获得最终分类结果如图 4(b) 所示。

2.4 分类结果比较及分析

此研究采用随机森林分类模型对研究区主要农作物进行分类识别并对比决策树、支持向量机和最大似然法分类结果。各方法分类结果展示如图 4 所示。从图 4 中可以看出, 在研究区范围内主要种植玉米和水稻 2 种作物, 且地块较大, 较为集中。其中水

稻主要分布在园区东南、东北地带,玉米主要分布在研究区中部及西南部。大豆和甜叶菊种植区域主要集中在研究区中部,并且地块较小。早稻种植区域零星分布,地块较小,与玉米的间种现象比较严重。4 种分类方法对水稻、大豆和甜叶菊的分类效果较好,对于玉米早稻混种区域,随机森林分类效果最好,决策树次之,最大似然法的分类效果最差。

决策树分类效果较好,是因为增加了对不同作物间光谱指数阈值选取,并逐步分层识别提取作物,作物光谱信息被更加深入的挖掘,分类具有针对性,解

决单一生育期,不同作物光谱特征差异性不大的问题,而随机森林算法建立多支决策树,并对分类器的结果进行投票,投票次数最多的类别为最终结果,通过增加分类器数量,改善了分类精度。支持向量机和最大似然法都只根据有限的样本信息进行地物分类提取,很难划分细碎地块和间种地块,从而导致它们的错分概率相对较高。

将分类后的结果与验证样本集计算混淆矩阵,评估分类后精度。决策树和随机森林分类模型混淆矩阵结果见表 6 和表 7。

表 6 决策树混淆矩阵结果

Tab.6 Confusion matrix result of Decision Tree

| Type | Soybean | Rice | Stevia | Corn | Dry rice | Total | User accuracy |
|------------------|---------|--------|--------|--------|----------|-------|---------------|
| Soybean | 3288 | 4 | 0 | 0 | 320 | 3612 | 91.03% |
| Rice | 0 | 5680 | 16 | 33 | 176 | 5905 | 96.19% |
| Stevia | 0 | 0 | 2039 | 0 | 59 | 2098 | 97.19% |
| Corn | 0 | 1 | 0 | 8451 | 221 | 8673 | 97.44% |
| Dry rice | 8 | 9 | 147 | 1055 | 4820 | 6039 | 79.81% |
| Mapping accuracy | 99.76% | 99.75% | 92.60% | 88.57% | 86.13% | | |

表 7 随机森林混淆矩阵结果

Tab.7 Confusion matrix result of Random Forest

| Type | Soybean | Rice | Stevia | Corn | Dry rice | Total | User accuracy |
|------------------|---------|--------|--------|--------|----------|-------|---------------|
| Soybean | 3294 | 0 | 0 | 0 | 393 | 3687 | 89.34% |
| Rice | 0 | 5692 | 19 | 183 | 216 | 6110 | 93.16% |
| Stevia | 0 | 0 | 1993 | 0 | 8 | 2001 | 99.60% |
| Corn | 0 | 0 | 2 | 9301 | 41 | 9344 | 99.54% |
| Dry rice | 4 | 2 | 142 | 102 | 4938 | 5188 | 95.18% |
| Mapping accuracy | 99.88% | 99.96% | 92.44% | 97.03% | 88.24% | | |

4 种分类方法总体精度和 Kappa 系数见表 8。结合图 4 对研究区作物分类识别结果进行分析,随机森林分类总体精度最好为 95.8%, Kappa 系数为 0.944。决策树分类总体精度为 92.3%, Kappa 系数为 0.897。

支持向量机分类总体精度为 91.6%, Kappa 系数为 0.890。最大似然法总体分类精度最低为 86.5%, Kappa 系数为 0.823。

表 8 各方法的分类总体精度和 Kappa 系数

Tab.8 Overall accuracy estimation and Kappa coefficient of classification based on each method

| Classification method | Overall accuracy | Kappa coefficient |
|------------------------|------------------|-------------------|
| Maximum Likelihood | 86.5% | 0.823 |
| Support Vector Machine | 91.6% | 0.890 |
| Decision Tree | 92.2% | 0.897 |
| Random Forest | 95.8% | 0.944 |

3 结 论

此研究采用时间序列 Sentinel-2 卫星遥感影像数据集, 结合研究区作物物候特点, 通过 NDVI、EVI 以及 RVI 等植被指数和近红外波段 Ref(NIR) 的时序变化特征, 选取作物关键生育期差异明显的光谱指数特征影像集, 采用随机森林分类方法成功提取研究区不同作物的种植区域。相比其他传统分类方法, 提高了分类精度, 简化了分类流程。采用随机森林方法的作物总体分类精度为 95.8%, Kappa 系数达到 0.944。通过该方法可以实现对扎赉特旗现代农业示范园区作物的精确提取。

此研究还存在一些不足需要改进, 尽管提出的方法在所选研究区域有较高的分类精度, 其普适性还需要开展更加深入的研究。另外, 随机森林算法中树的分支数带来的影响并没有考虑进去, 下一步将对不同分支数对分类精度开展更加深入的分析 and 探讨。

参考文献:

- [1] 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. (in Chinese)
- [2] Zhao Zijuan, Liu Dong, Hang Zhongqiao. Research status and prospects of crop remote sensing recognition methods [J]. *Jiangsu Agricultural Sciences*, 2019, 47(16): 45-51. (in Chinese)
- [3] Yang Chenghai, James H Everitt, Dale Murden. Evaluating high resolution SPOT 5 satellite imagery for crop identification [J]. *Computers and Electronics in Agriculture*, 2010, 75(2): 347-354.
- [4] Shen Wenming, Wang Wenjie, Luo Haijiang, et al. Classification methods of remote sensing image based on decision tree technologies [J]. *Remote Sensing Technology and Application*, 2007, 22(3): 333-338. (in Chinese)
- [5] Tian Jing, Wang Juanle, Li Yifan, et al. Land cover classification in mongolian plateau based on decision tree method: A case study in Tov Province, Mongolia [J]. *Journal of Geo-information Science*, 2014, 16(3): 460-469. (in Chinese)
- [6] Friedl M A, Brodley C E, Strahler A H. Maximizing land cover classification accuracies produced by decision trees at continental to global scales [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 1999, 37(2): 969-977.
- [7] Xu Xingang, Li Qiangzi, Zhou Wancun, et al. Classification application of quick bird imagery to obtain crop planting area [J]. *Remote Sensing Technology and Application*, 2008, 23(1): 17-23. (in Chinese)
- [8] Tan Kun, Du Peijun. Hyperspectral remote sensing images classification based on support vector machine [J]. *Journal of Infrared and Millimeter Waves*, 2008, 27(2): 123-128. (in Chinese)
- [9] Gu Xiaohe, Han Lijian, Zhang Jinshui, et al. Monitoring of paddy rice plant area based on similar index by multi-resolution remote sensing data [J]. *Scientia Agricultura Sinica*, 2008, 41(4): 978-985. (in Chinese)
- [10] Pan Chen, Du Peijun, Zhang Hairong. Decision tree classification and application in remote sensing image processing [J]. *Science of Surveying and Mapping*, 2008, 33(1): 208-211. (in Chinese)
- [11] Brfiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [12] Li Cunjun, Wang Jihua, Liu Liangyun, et al. Land cover mapping of winter wheat and clover using multi-temporal Landsat NIR band in a growing season [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2005, 21(2): 96-101. (in Chinese)
- [13] Li Xinchuan, Xu Xingang, Wang Jihua, et al. Crop classification recognition based on time-series images from HJ satellite [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(2): 169-176. (in Chinese)
- [14] Wei Pengfei, Xu Xingang, Yang Guijun, et al. Remote sensing classification of crops based on the change characteristics of multi-phase vegetation index [J]. *Journal of Agricultural Science and Technology*, 2019, 21(2): 54-61. (in Chinese)
- [15] Xu Qingyun, Yang Guijun, Long Huiling, et al. Crop information identification based on MODIS NDVI time-series data [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2014, 30(11): 134-144. (in Chinese)
- [16] An Qin, Chen Shengbo, Sun Shichao, et al. Study on corn yield estimation based on multi-temporal MODIS-RVI [J]. *Geospatial Information*, 2018, 16(3): 14-16, 8. (in Chinese)
- [17] Zhang Xia, Jiao Quanjun, Zhang Bing, et al. Preliminary study on cropping pattern mapping using MODIS_EVI image time

- series [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2008, 24(5): 161-165. (in Chinese)
- [18] Pan Yaozhong, Li Le, Zhang Jinshui, et al. Winter wheat area estimation from MODIS-EVI time series data using the Crop Proportion Phenology Index [J]. *Remote Sensing of Environment*, 2012, 119(3): 232-242.
- [19] Wang Limin, Liu Jia, Yang Fugui, et al. Rice recognition ability basing on GF-1 multi-temporal phases combined with near infrared data [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33(23): 196-202. (in Chinese)
- [20] Liu Yi, Du Peijun, Zheng Hui, et al. Classification of China small satellite remote sensing image based on random forests [J]. *Science of Surveying and Mapping*, 2012, 37(4): 194-196. (in Chinese)
- [21] Balogun A L, Yekeen S T, Pradhan B, et al. Spatio-temporal analysis of oil spill impact and recovery pattern of coastal vegetation and wetland using multispectral Satellite Landsat 8-OLI Imagery and Machine Learning Models [J]. *Remote Sensing*, 2020, 12(7): 1225-1225.
- [22] Yu Haiyang, Zhao Gengxing, Chang Chunyan, et al. Random forest classifier in remote sensing information extraction: A review of applications and future development [J]. *Remote Sensing Information*, 2019, 34(2): 8-14. (in Chinese)
- [23] Sebastian van der Linden, Andreas Rabe, Matthias Held, et al. The en MAP-Box —A toolbox and application programming interface for en MAP data processing [J]. *Remote Sensing*, 2015, 7(9): 11249-11266.