

doi: 10.6046/zrzyyg.2023166

引用格式: 张冬韵, 吴田军, 李曼嘉, 等. 地块尺度农作物遥感分类及其不确定性分析[J]. 自然资源遥感, 2024, 36(4): 124-134. (Zhang D Y, Wu T J, Li M J, et al. Remote sensing-based classification of crops on a farmland parcel scale and uncertainty analysis[J]. Remote Sensing for Nature Resources, 2024, 36(4): 124-134.)

地块尺度农作物遥感分类及其不确定性分析

张冬韵¹, 吴田军², 李曼嘉¹, 郭逸飞¹, 骆剑承^{1, 3}, 董文¹

(1. 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; 2. 长安大学土地工程学院, 西安 710064; 3. 中国科学院大学资源与环境学院, 北京 100049)

摘要: 利用遥感技术开展农作物空间分布的快速调查与精准制图是现代精细农业的一项基础工作。然而, 由于遥感影像获取、处理与分析过程中的局限性, 传统农作物种植结构遥感制图精度受到一定影响, 亟须面向农作物分类开展不确定性的空间建模与特征分析。该文将耕地地块作为基本空间单元, 选择宁夏引黄灌区作为试验区, 利用多源遥感数据和机器学习算法实现地块尺度的农作物分类, 进而基于混合熵构建不确定性计算模型, 生成地块农作物类型不确定性空间分布, 再利用多源辅助数据对地块农作物分类不确定性进行回归建模, 探究相关地理变量的潜在影响。实验结果表明: ①耕地提取及分类环节共构建地块矢量单元 149 万个, 总体作物分类精度达 0.80, 制图结果与实际农业耕作管理单元相匹配, 分类效果较之传统的像素尺度方法更为理想; ②地块尺度的农作物分类不确定性总体较低, 存在较为显著的类别差异, 水稻、菜地、苜蓿分类不确定性较小, 单种、复种小麦不确定性较高, 玉米作物分类不确定性介于二类之间; ③地块级作物分类不确定性与种植结构、资源条件等多种环境因素有关, 且与作物类型、水源可达性的相关性最为显著。

关键词: 遥感; 农作物分类; 地块; 不确定性; 机器学习; 混合熵

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 2097-034X(2024)04-0124-11

0 引言

种植结构描述特定生产区域内农作物的种植比例及空间布局^[1], 及时、准确地获取种植结构是监测区域作物长势、估测产量和灾情评估等应用的基础^[2]。当前, 遥感已成为种植结构调查的重要手段, 但由于遥感数据自身不确定性及数据处理、分析过程中的局限, 作物分类制图精度受到显著影响。因此, 对分类过程中的不确定性进行科学表达及定量分析, 是提升农作物遥感制图精确度与可靠性的必经之路。

在地理学研究领域, 数据的精确度常被定义为观测值或计算值与真实值之间的近似程度, 因此, 在对空间数据产品进行精度评价时, 应以一个误差为零的数据作为真实值; 但由于地表对象演进的连续性和复杂性, 其绝对状态难以获取, 因而数据不确定性的概念被引入到空间数据质量评估中^[3]。不确定性是一个比误差和精度更普遍的概念, 对遥感分

类任务而言, 确定性可以看作是分类器对已分配标签的确信程度, 过去研究中常使用的误差和精度则可以理解为不确定性的 2 种表达形式^[4]。了解不确定性指标在农作物遥感分类结果中的局部强度有助于量化和评估制图质量在空间上的变化, 并允许反向追踪农作物类型识别有误的特征区域, 进一步提高后续建模的可靠性, 因此具有重要意义。

在过往研究中, 已有学者对遥感分类制图任务中不确定性的形成过程进行了解析, 指出在数据获取、影像处理、分类算法选取, 以及对分类结果的精度评价等各个环节中均会引入一定量的不确定性, 且不确定性被不断传递、叠加^[5]。专题制图的不确定性空间分布模式是非稳态、非均匀的, 与地类边界(如混合像元)、土地覆盖类型、土地利用类型、景观特征(斑块大小、海拔、坡度)、算法性能、辅助参考数据密度、可靠性和空间组成等因素^[6-9]相关。若能识别上述不确定性因素, 有助于研究人员评估制图结果的可信程度并基于不确定性来源对制图过程开展针对性优化。

收稿日期: 2023-06-06; 修订日期: 2024-03-25

基金项目: 国家重点研发计划项目“地理空间智能核心技术与软件系统”(编号: 2021YFB3900905)资助。

第一作者: 张冬韵(1997-), 女, 硕士研究生, 从事农业遥感研究。Email: zhangdongyun19@mails.ucas.ac.cn。

通信作者: 骆剑承(1970-), 男, 博士, 研究员, 从事遥感大数据智能计算研究。Email: luojc@aircas.ac.cn。

在定性分析的基础上,已有多种方法用于遥感分类不确定性的定量评价。综合而言可划分为影像整体尺度、地物类型尺度和像元尺度 3 类^[10]。其中,影像整体尺度和地物类型尺度方法均为基于误差矩阵的统计分析,存在以下局限性:其一,误差矩阵无差别综合了所有像元的分类结果,而忽略不同像元对于总体不确定性的贡献度差异;其二,误差矩阵及其衍生指标不具备提供不确定性空间分布信息的能力,与空间属性明显的遥感地学分析情境不符^[11-12]。对此,Goodchild 等^[13]在贝叶斯分类过程中,曾提出一种基于后验概率向量的计算方法,用于模拟整个分类区域像元尺度的不确定性。在此基础上,Brown 等^[14]结合信息熵理论提出了概率熵;Comber^[15]开发了一种地理加权方法来估计局部精度;Khatami 等^[16]利用机器学习分类算法中的分类置信度表达类型不确定性;Immitzer 等^[17]使用随机森林算法进行分类,通过输出类型的投票数与分类树总数的比例来表达类型不确定性,并指出随机森林算法可以提供有意义的类型不确定性信息。较之混淆矩阵中得出的准确度,上述方法基于像素或区域尺度提供了类型不确定性的空间信息,有助于理解误差的空间分配。

然而,在空间表达过程中,传统基于像元或多尺度分割对象的农业遥感制图结果与真实地理实体相

脱节,导致计算过程不可控、结果难以验证,是不确定性的主要来源之一。与之相比,耕地地块作为人类利用土地资源开展农耕活动的基本单元,可真实表征耕作区域空间形态、作物类型及生长状况。因此,基于不规则地块的表达方式有望弥补上述缺陷,促使计算过程收敛、结果可信。然而,目前鲜有将地块作为整体研究对象的不确定性分析研究。为探索该问题,本文参考像元尺度的不确定性分析思路,面向农作物遥感分类问题尝试开展地块尺度的不确定性分析。

1 研究方法

基于上述研究目标,本文提出如下总体研究思路:首先通过机器学习方法获取地块边界和农作物类型信息,然后以混合熵原理为基础,综合随机性和模糊性开展地块尺度的农作物分类不确定性评价,并在此基础上分析不同地理环境变量对该不确定性空间分布模式的影响。具体研究路线如图 1 所示。先从高空间分辨率影像上提取耕地地块,再结合多期高时间分辨率影像构建地块尺度的归一化植被指数(normalized difference vegetation index, NDVI)时序曲线,进而结合实地调查样本进行农作物类型识别,最后据此开展不确定性的分析评价。

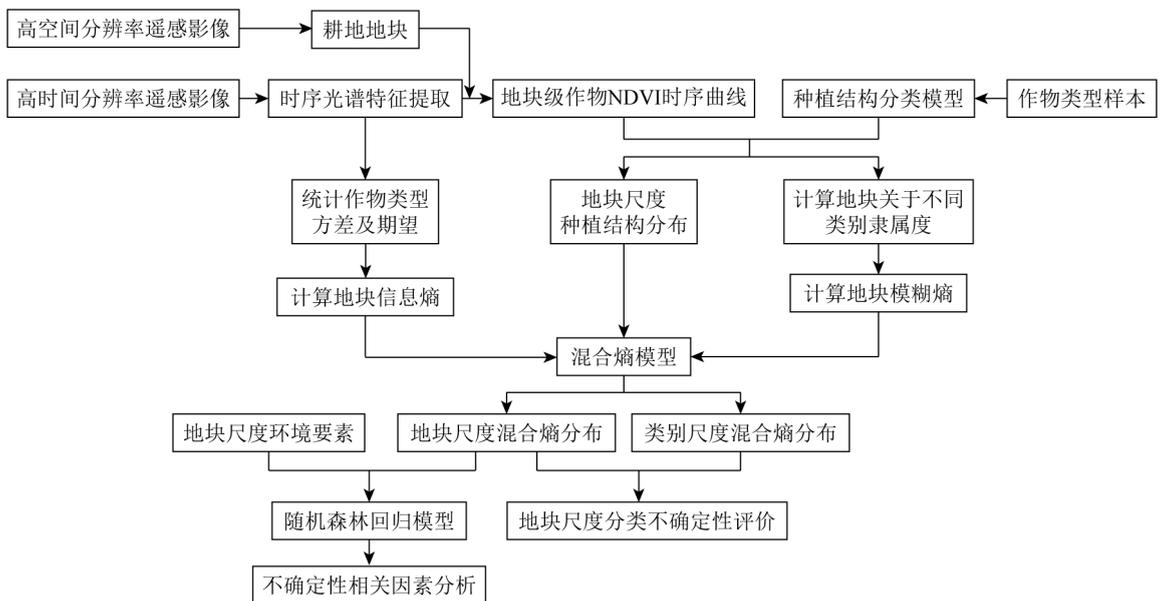


图 1 本文研究方法的总体流程图

Fig.1 Overall flow chart of the research method in this paper

1.1 基于机器学习的耕地地块提取与农作物分类

首先开展耕地地块的提取。平原地区耕地地块大多几何形态较为规则、边缘较为清晰、内部作物类型较为单一,在高分遥感影像上表现为排列紧凑、分布整齐、内部较均匀的四边形面状地物^[18]。基于上

述特点,本研究利用更丰富卷积特征(richer convolutional features, RCF)网络^[19]构建耕地地块提取模型(图 2)。RCF 由主干网络、深监督模块及特征融合模块组成,实验中删除原网络第 4 层,并将所有激活函数替换为 tanh 函数。在训练样本方面,本研究

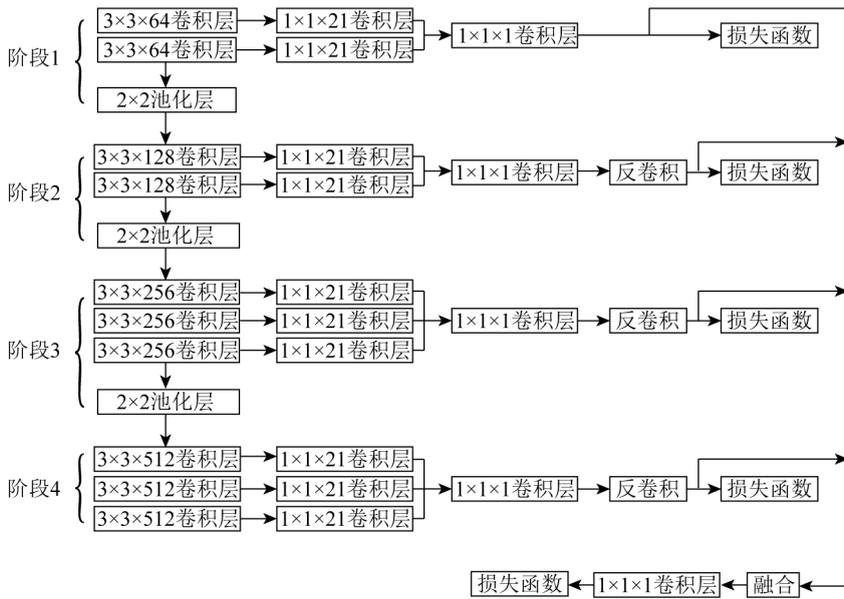


图 2 改进的 RCF 网络结构

Fig.2 Machine learning networks for farmland boundary extraction and crop type recognition

在高分影像上随机选取 80 个 1 000 像素×1 000 像素的子区域,通过目视解译对耕地边界进行数字化,构建样本集。

在地块单元提取的基础上,本研究利用多时相的 Sentinel-2 影像,通过波段计算得到不同时相上每个像元对应的 NDVI,并对地块边界内的所有像元统计均值,经异常值剔除后生成一条相对平滑的地块级 NDVI 时序曲线,用于描述农作物生长期内的生长状态变化,以表示其类型的时空分异^[20]。为了挖掘输入信号序列在时间维度上的长期依赖特征,本实验选用长短期记忆网络(long short-term memory, LSTM)^[21]作为推测模型。该模型在隐藏层中设置包含 3 个门控组件的记忆元,借此选择需要记忆或遗忘的信息,并通过记忆元的连接实现处理时间序列信息上的连贯性。实验中基于 LSTM 的上述基本结构,构建了包含 4 个隐藏层、36 个隐藏神经元的推测模型,结合实地获取的地面样本和地块级 NDVI 时序数据进行农作物类型的分类识别。

1.2 基于熵理论的农作物分类不确定性计算

对于农作物分类任务而言,不确定性的来源包括随机不确定性和模糊不确定性 2 类。其中,随机不确定性主要来源于自然现象本身固有的不确定性,在遥感影像上表现为“同物异谱”和“同谱异物”^[22]。在熵理论中,常用信息熵对随机不确定性进行量化,度量系统中信息的有序程度^[23],即不同类型地物对应光谱特征的稳定性。本研究基于地块尺度时序光谱特征,构建如下熵模型:

$$H_i(\mathbf{x}) = E[-\lg p(\mathbf{x}_i)] = - \sum_{i=1}^j p(\mathbf{x}_i) \lg p(\mathbf{x}_i), \quad (1)$$

$$p(\mathbf{x}_i) = \frac{p'(\mathbf{x}_i)}{\sum_{i=1}^j p'(\mathbf{x}_i)}, \quad (2)$$

$$p'(\mathbf{x}_i) = \prod_{n=1}^m \frac{1}{\sigma_{in} \sqrt{2\pi}} \exp \left[- \frac{(\mathbf{x}_n - E_{in})^2}{2 \sigma_{in}^2} \right], \quad (3)$$

式中: \mathbf{x} 为地块的光谱特征矩阵; \mathbf{x}_n 为 \mathbf{x} 的第 n 维 ($n \in [1, m]$) 特征(特征值为内含所有像元值的平均); E_{in} 和 σ_{in} 分别为第 i 类 ($i \in [1, j]$) 地物训练样本第 n 维特征值的期望和标准差; $p'(\mathbf{x}_i)$ 为 \mathbf{x} 对应第 i 类地物的概率密度分布函数,由于不同类型农作物并不满足相同正态分布,因此需对其进行归一化处理得到 $p(\mathbf{x}_i)$; $H_i(\mathbf{x})$ 为 \mathbf{x} 对应地块的信息熵,反映了描述地块种植作物类型归属的概率分布所需信息量。

模糊不确定性则由数据质量、分类方法等各个环节存在的不完备性共同造成。例如,由于传感器的分辨率限制和地表系统的复杂性,影像上的单个像元和地块对象内部可能混杂多类地物,混合像元或混合地块与纯净地物间的光谱特征存在一定差异,引发类型归属的模糊不确定性^[24-25]。为此,本研究进一步引入模糊熵用于评价分类结果的模糊不确定性,构建如下计算公式:

$$\mu(\mathbf{x}) = \begin{bmatrix} p(y = 1 | \mathbf{x}, \boldsymbol{\theta}) \\ p(y = 2 | \mathbf{x}, \boldsymbol{\theta}) \\ \vdots \\ p(y = j | \mathbf{x}, \boldsymbol{\theta}) \end{bmatrix} = \frac{1}{\sum_{i=1}^j e^{\theta_i^T \mathbf{x}}} \begin{bmatrix} e^{\theta_1^T \mathbf{x}_L^T} \\ e^{\theta_2^T \mathbf{x}_L^T} \\ \vdots \\ e^{\theta_j^T \mathbf{x}_L^T} \end{bmatrix}, \quad (4)$$

$$H_f(\mathbf{x}) = - \sum_{i=1}^j \{ \mu_i(\mathbf{x}) \lg \mu_i(\mathbf{x}) + [1 - \mu_i(\mathbf{x})] \lg [1 - \mu_i(\mathbf{x})] \}, \quad (5)$$

式中: y 为输入地块特征光谱 \mathbf{x} 对应的预测类型; $e^{\theta^T} \mathbf{x}_L^T$ 为 Bi-LSTM 最后一层的输出向量; \mathbf{x}_L 为网络最后一层输入; $\theta = (w_0, w_1, \dots, w_n)$ 为模型参数; $\mu(\mathbf{x})$ 为对 $e^{\theta^T} \mathbf{x}$ 进行 SoftMax 处理后的概率密度函数; $\mu_i(\mathbf{x})$ 为 \mathbf{x} 属于类型 i 的概率值, 等价于模糊集

$$H_h(R, F) = - \sum_{i=1}^j p_i \lg p_i - \sum_{i=1}^j \{ \mu_i \lg \mu_i(\mathbf{x}) + [1 - \mu_i(\mathbf{x})] \lg [1 - \mu_i(\mathbf{x})] \} + \sum_{i=1}^j (1 - p_i) \{ \mu_i(\mathbf{x}) \lg \mu_i(\mathbf{x}) + [1 - \mu_i(\mathbf{x})] \lg [1 - \mu_i(\mathbf{x})] \} = H_r + H_f - H_{rf}, \quad (7)$$

式中: H_r 和 H_f 分别为由式(1)和式(5)定义的信息熵和模糊熵; H_{rf} 为交叉熵, 可看作是随机性和模糊性的交叉效应。式(7)表明, 在由随机空间和模糊空间所定义的整体空间中, 混合熵等于信息熵与模糊熵之和减去交叉熵, 当随机不确定性与模糊不确定性某一者消失时, 混合熵退化为其中一者。根据混合熵定义及式(7)可知, 其值越接近于 0, 该地块作物分类不确定性越小; 反之, 其值越大, 则地块作物分类不确定性越大。

1.3 农作物分类不确定性的影响因素解析

为了解析不确定性分布模式, 本研究进一步选取可能造成影响的地理环境因素^[26-27], 并借助随机森林回归方法评估各因素的重要程度。具体而言, 可以归纳为以下 4 类:

1) 种植结构。若农业生产模式中存在复种、套种等复杂种植结构, 会使地块时序曲线产生偏差, 在机器学习分类过程中更易被错分或漏分, 进而导致类型不确定性升高, 因此选择作物类型(x_1)作为影响变量。

2) 地块形态。为了探究地块形态对分类不确定性的影响, 计算每个地块的地块面积(x_2, m^2)、地块周长(x_3, m)、地块形状指数(x_4), 其中形状指数描述了地块形态的光滑程度, 公式为:

$$S = \frac{e}{4\sqrt{A}}, \quad (8)$$

式中: S 为地块形状指数; e 为地块边界长度; A 为地块面积。

3) 样本特征。实际采样过程中进入田块的机会有限, 因此样本的空间分布会受到一定限制, 这种非随机、非分层的抽样可能会在精度评估中引入偏差^[28], 进而影响分类不确定性的空间分布。本研究参考 Yu 等^[9]的处理方式, 使用不同半径的核函数来估计样本密度, 分别以 1 km, 2.5 km 和 5 km 为半径计算样本密度($x_5, x_6, x_7, \text{个}/\text{km}^2$)。

合中的隶属度概念; $H_f(\mathbf{x})$ 为 \mathbf{x} 对应的模糊熵。

综上, 鉴于信息熵和模糊熵对不确定性的侧重点不同, 本研究将两者进行耦合, 形成如下的混合熵计算公式:

$$H_h(R, F) = - \sum_{i=1}^j \{ [(p_i \mu_i(\mathbf{x}) \lg(p_i \mu_i(\mathbf{x})) + p_i(1 - \mu_i(\mathbf{x})) \lg(p_i(1 - \mu_i(\mathbf{x})))] \}。 \quad (6)$$

对式(6)进行进一步分解后可得:

4) 环境与农业管理条件。鉴于环境与农业管理因素会影响灌溉水等生产要素的供应, 影响生物有机质含量, 作用于植被导致某些类型之间的特征混淆, 本研究还计算了灌溉渠网密度($x_8, \text{km}/\text{km}^2$)、交通路网密度($x_9, \text{km}/\text{km}^2$)、居民点密度($x_{10}, \text{km}/\text{km}^2$)、地块与水渠的距离(x_{11}, km)作为分析的辅助变量。

在此基础上, 本研究使用随机森林回归方法评估各类地理因素对农作物分类不确定性的影响^[29], 并得到特征重要性指标。由于树在每个节点选择的变量的随机性, 变量重要性计算过程存在一定程度随机性。因此, 对随机森林模型进行 10 次回归实验, 并取 10 次特征重要性结果平均值。

2 实验及结果分析

2.1 实验区概况及数据介绍

本研究选取宁夏回族自治区引黄灌区作为实验区, 如图 3 所示。该区域地处黄河中上游, 覆盖青铜峡市、银川市等 11 个市区县, 属大陆性气候, 干旱少雨、热量充足。区域内耕地地块广阔、空间形态规整; 播种类型丰富, 主要粮食作物包括水稻、玉米、小麦等, 经济作物包括蔬菜、向日葵、西瓜等。宁夏引黄灌区为我国干旱半干旱地区的粮食主产地之一, 播种制度多样, 主要方式为 1 年水稻 1 年旱作物的两段轮作制。在该区域开展地块级农作物分类及其不确定性的评价分析实验, 具有较好的代表性。

本研究所采用的影像数据主要包括 Google Earth 及 Sentinel-2B 影像, 其中 Google Earth 影像空间分辨率达 1 m, 主要用于开展地块边缘特征提取实验, Sentinel-2B 影像共收集并处理 33 期 225 景, 空间分辨率为 10 m, 时间范围为 2020 年 3 月 7 日—10 月 21 日, 覆盖研究区农作物的播种、生长及成熟期, 主要用于构建时序 NDVI 特征和开展农作物类型



图3 研究区位置及 Google Earth 影像切片
Fig.3 Location and view of Google Earth image of the study area

识别。为分析不确定性空间分布的影响因素,本研究还获取区域内水系、交通、居民点及设施数据,用于计算灌溉渠网密度、交通路网密度、居民点密度等

与地块作物分类相关的环境特征。此外,利用手持全球移动定位系统(global positioning system, GPS)进行农作物定位调查,通过地面采样收集了 465 个耕地地块采样点的空间坐标和种植作物类型信息,并对照 Google Earth 高分影像上随机生成的点位,通过参照地面点的人工目视解译方法,补充了模型训练及验证所需的农作物类型样本,最终得到了 8 大类作物的 2 469 个地块样本,包括水稻 584 个、单种小麦 377 个、复种小麦(即在小麦收割至下一轮播种的间隙,种植蔬菜等其他作物,以提高耕地的利用效率)326 个、玉米 579 个、菜地 249 个、果园 76 个、苜蓿 186 个、其他作物 92 个。

2.2 耕地地块提取与农作物分类结果

本研究基于 Google Earth 高分影像数据,结合耕地地块边界样本,借助 RCF 网络提取了地块边缘强度信息,并经矢量化处理构建地块对象。本次实验共提取研究区内的地块约 149 万个,典型地块的细节如图 4 所示,总体效果较为理想,可为后续农作物类型识别及不确定性分析提供稳定的耕地边界空间约束。

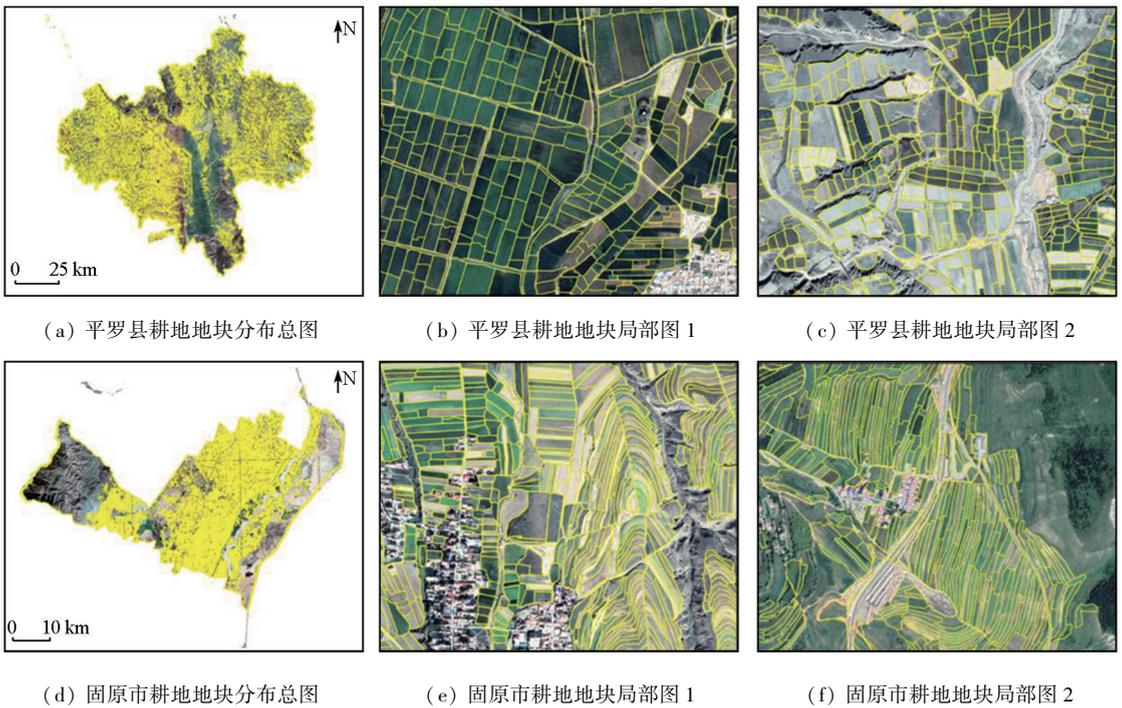


图4 研究区内耕地地块提取效果

Fig.4 Extraction effects of cropland-parcels in the study area

在地块单元构建基础上,本研究基于 Sentinel-2B 卫星时序影像计算了地块内各像元的 NDVI 值,取其均值生成地块尺度的 NDVI 序列,再使用地面采集的样本数据训练 LSTM 网络,预测各地块作物类型。分类结果如图 5 所示,据此统计后结果显示:玉米种植面积最大,占全区耕地面积的 40% 以上;

菜地次之,面积比重超过 16%;水稻和小麦种植面积相近,占比均在 12% 左右,其中单种小麦面积比重较高,为复种小麦的 4 倍左右;灌区果园、苜蓿等面积较小,仅占总耕种面积的 3%;其他作物面积占比约 16%。

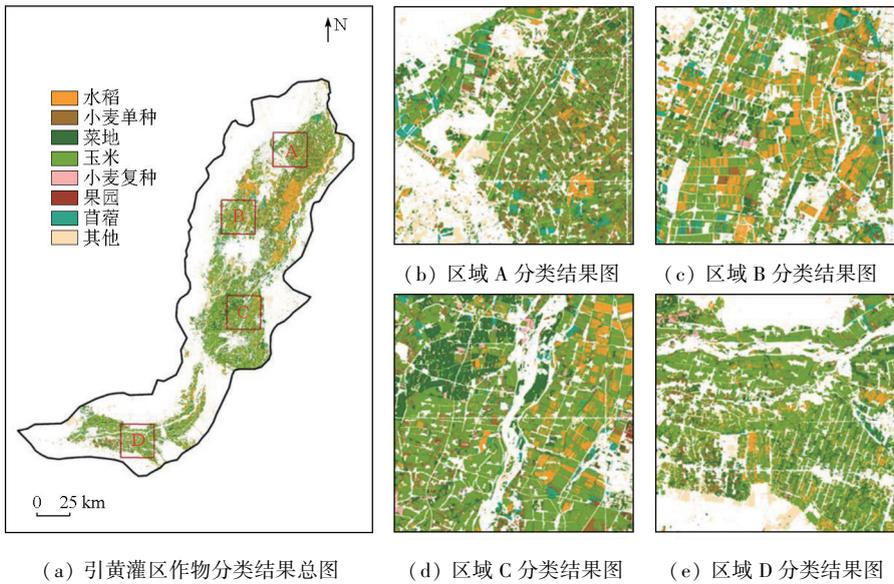


图 5 2020 年宁夏引黄灌区农作物分类结果

Fig.5 Crop classification results of Yellow River Irrigation Area of Ningxia in 2020

2.3 地块农作物分类的不确定性计算结果

进一步,基于 1.2 节的混合熵模型计算得到地块尺度农作物分类的不确定性评价结果。如图 6 所示,在全域上,混合熵取值范围介于 0.000 52 ~

3.449 25 之间,平均值约为 1.130 18; 在局部上,邻近地块的混合熵指标各异、交相分布,显示了不确定性的空间结构较为复杂。

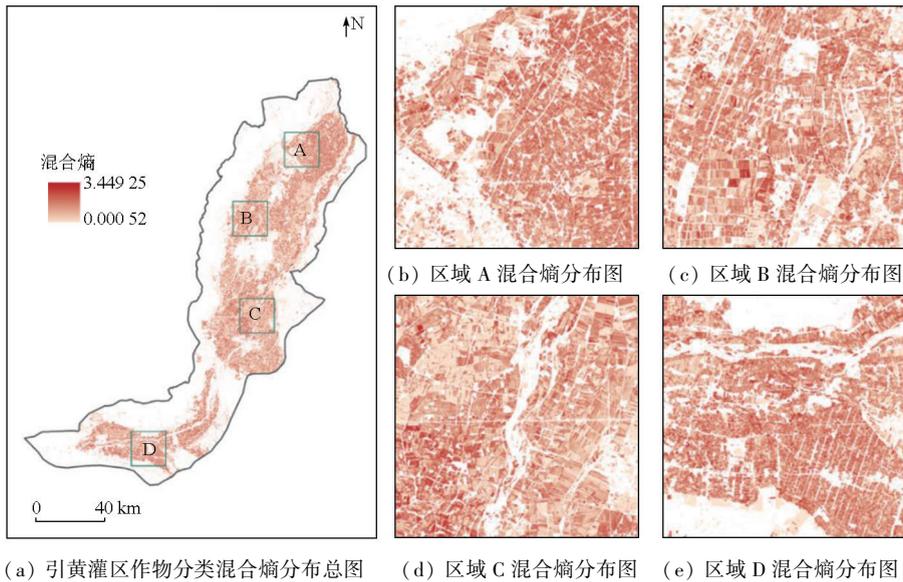


图 6 研究区地块尺度农作物分类的混合熵计算结果

Fig.6 Calculation results of hybrid entropy of crop classification on farmland-parcel scale in the study area

为详细分析不确定性的类型结构,本文统计了各作物类型不同混合熵值区间的地块数量情况,并将综合反映地块光谱特征及类型判定不确定性的混合熵指标(指标 1)、直接反映分类模型判别归类情况的最大隶属度指标(指标 2),以及反映验证集分类召回率与精确率的 F1 指标(指标 3)进行协同分析。其中,由式(7)的定义可知,混合熵由信息熵及模糊熵 2 部分综合而得,整体反映“同物异谱”等地物特征固有随机不确定性,以及由于特征不完备、地

块非均质、模型学习能力不足等引发的类型归属模糊性;该指标与直接反映分类模糊性的最大隶属度有所关联,理论上应遵循最大隶属度越高,信息熵越低的规律。此外,低不确定性意味着地块光谱特征具有较为明确的农作物类型归属倾向,易于准确判定其作物类型,因而混合熵越低,F1 得分应越高。

各类作物混合熵分布拟合概率密度曲线及不确定性指标对比分析结果分别如图 7、表 1 所示。由图 7 可知,该区域实验的农作物类间不确定性差异

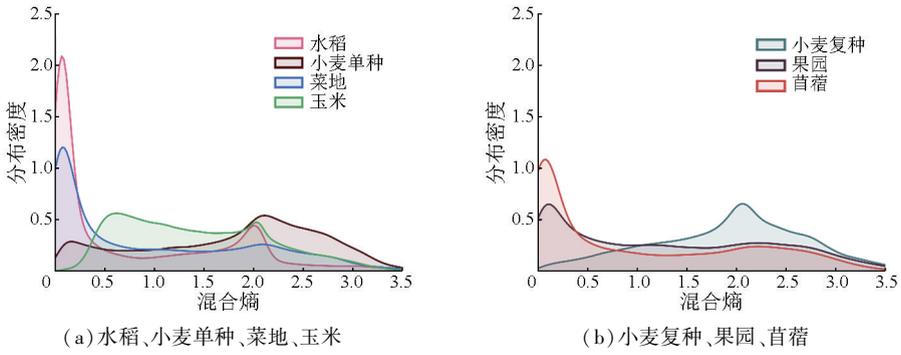


图7 地块尺度各农作物类型的不确定性分布概率密度曲线图

Fig.7 Probability density curve of uncertainty distribution of each crop type at the farmland-parcel scale

表1 地块尺度各农作物类型的不确定性对比分析

Tab.1 Comparative analysis of uncertainty of each crop type at the farmland-parcel scale

农作物类型	混合熵均值		隶属度均值		F1		总体分类精度
	值	由低到高排序	值	由高到低排序	值	由高到低排序	
水稻	1.03	2	0.83	3	0.89	1	0.80
小麦单种	1.93	7	0.67	7	0.68	6	
菜地	0.92	1	0.85	1	0.84	3	
玉米	1.37	5	0.81	4	0.82	4	
小麦复种	1.92	6	0.69	6	0.65	7	
果园	1.29	4	0.79	5	0.86	2	
苜蓿	1.04	3	0.85	2	0.79	5	

显著：水稻、菜地、苜蓿3类作物的混合熵整体较小，均值皆在1.0上下，熵值介于0~0.13之间的地块显著占优，熵值大于2.0的地块数目稀少；果园地块的分布模式与前三者相近，但熵值大于0.13的地块比重有所增加，混合熵均值上升至1.29；玉米地块与果园的混合熵均值相近，但熵值优势分布扩展至0.4~2.0区间；单种及复种小麦的整体不确定性最高，混合熵分布向高值区域偏移，均值达到1.9以上。此外，对比表1中各类型3种评价指标可知，与前述相关分析基本一致，但最大隶属度越高、不确

定性(混合熵)越低的直觉规律并非严格成立，例如，玉米地块的平均归属度高于果园地块，但混合熵均值亦较高，说明分类器的判定倾向性显著，但对于次要类型的判别模糊性较高，或光谱特征的信息熵较高、先验随机性明显。同时，高F1得分并不必然代表低分类不确定性，如水稻地块的F1得分高于菜地，但混合熵亦较高；苜蓿的混合熵低于玉米，但或许由于样本数量差异，F1指标亦较低。这说明不确定性指标与传统精度指标的衡量标准各有侧重，且验证集上的二值化分类精度指标并不能完全反映全局分类效果。

为进一步分析农作物分类不确定性(混合熵)与隶属度的关系，并借此分析不确定性的机理性来源，本文绘制了不确定性与最大隶属度二者之间的散点图(图8)，同时计算验证集样本地块的多维隶属度均值，得到混淆矩阵(图9)。由上述2图可知，地块尺度农作物分类的混合熵与最大隶属度之间呈现近似负相关二次曲线关系，混合熵高处最大隶属度低，且点位密集、高隶属度区域(即农作物类内一致性高、不确定性低)曲线形态特征更加明显。例如，苜蓿地块判识的隶属度较高、混合熵均值较低，且不存在与其他农作物明显的混淆情况，因此散点

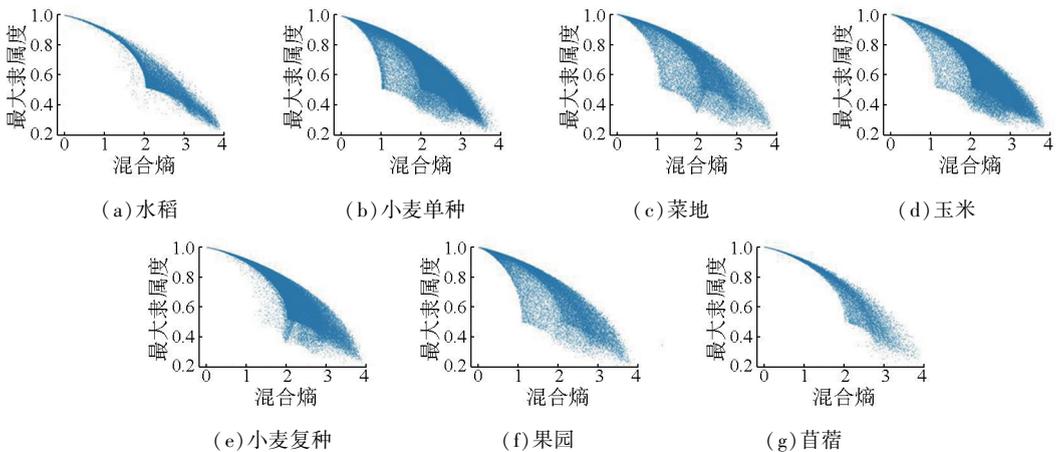


图8 研究区主要农作物分类不确定性(混合熵)与最大隶属度的散点图

Fig.8 Scatterplot of uncertainty and maximum membership of major crop classifications in the study area

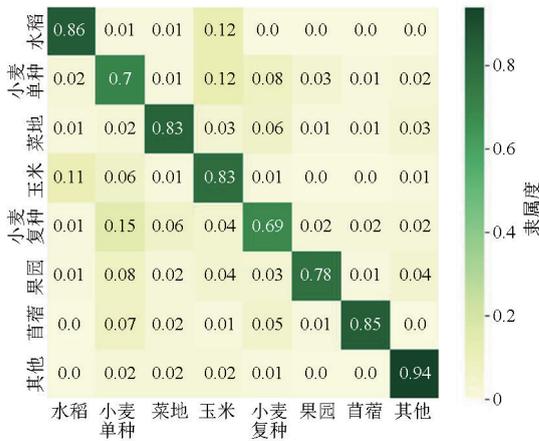


图 9 研究区内验证样本地块对各农作物类型的隶属度值分布

Fig.9 Distribution of membership values of the validation sample land-parcels to each crop type in the study area

图呈现出较为紧凑的抛物线形状; 玉米地块混合熵均值较高, 且与水稻存在误分现象; 小麦单种地块同理, 混合熵均值较高、隶属度低, 且与小麦复种、玉米混淆明显, 散点图分散。另外, 对比各类农作物的 NDVI 时序曲线可知, 类型混淆及高不确定性与作物 NDVI 曲线特征相关: 水稻及玉米的 NDVI 曲线形态相近, 小麦单种、复种作物的部分物候特征完全一致, 导致地块间可区分性下降、类型随机性增加, 因而不确定性较高; 菜地、苜蓿的曲线形态独特, 波峰波谷特征与其他粮食作物差异显著, 故而分类不确定性较低。

2.4 地块农作物分类的不确定性影响因素解析结果

基于 1.3 节所列的种植结构、地块形态、样本特征、环境与农业管理条件等特征因子以及上文以混合熵度量得到的农作物分类不确定性, 本研究借助随机森林回归模型进一步分析了在地块尺度农作物分类不确定性的相关影响因素。为减少空间相关性的影响, 本文随机抽取 268 972 个地块, 并以 9:1 分为训练集与测试集, 采样均方误差 (mean square error, MSE) 作为衡量标准, 并输出特征重要性及其排名。计算结果如表 2 所示, MSE 值为 0.28, R² 值为 0.68, 总体表明随机森林回归模型对地块农作物分类不确定性具有较为好的建模能力。从表 2 可知, 回归模型中最重要的 5 个变量依次为: 作物类型、水渠距离、地块形状指数、灌溉渠网密度、交通路网密度。这表明地块农作物分类不确定性主要受制于种植结构、环境与农业管理条件以及地块形态特征, 具体解析如下:

首先, 重要性最高的影响变量为农作物类型, 这与 2.3 中的相关讨论一致, 不同作物的光谱分布特征不同, 类内一致性、类间差异度有所区别, 直接影

表 2 随机森林回归模型对农作物分类不确定性进行建模生成的特征重要性

Tab.2 Feature importance of random forest regression model in modeling uncertainty of crop classification

类型	变量	特征重要性	排名
种植结构	作物类型	0.27	1
	地块面积	0.10	7
地块形态	地块周长	0.08	8
	地块形状指数	0.14	3
样本特征	样本密度 1 km	0.04	11
	样本密度 2.5 km	0.06	9
	样本密度 5 km	0.13	6
	灌溉渠网密度	0.14	4
环境与农业管理条件	交通路网密度	0.14	5
	居民点密度	0.04	10
	与水渠的距离	0.16	2
MSE	0.28	R ²	0.68

响地块农作物类型的归属判断及不确定性指标。如苜蓿、菜地地块物候特征独特, 不易与其他类型相互混淆, 整体不确定性较低; 小麦、玉米等与其他类型特征相近, 受类内光谱特征差异的影响显著, 不确定性较高。

其次, 地块周边的灌溉条件也是相关因素之一, 水渠距离、灌溉渠网密度的特征重要性分别排名第二、第四。Löw 等^[30]指出, 由灌溉水供应减少引起 NDVI 剖面变平, 一定程度增加了农作物分类的不确定性, 即不理想的生长条件可能会削弱分类模型的稳定性。这意味着田块周围水渠密度高、与黄河及其支流距离近时, 灌溉条件可能更为理想, 不确定性随之降低。以水稻地块为例, 如图 10 所示, 高不确定性耕地地块整体生物量偏低, NDVI 曲线相对平坦; 同时, 曲线缺乏在五月灌水期, 由于水田对红外波段反射能力较弱而应出现的低值特征, 导致光谱特征模糊、分类不确定性增加。

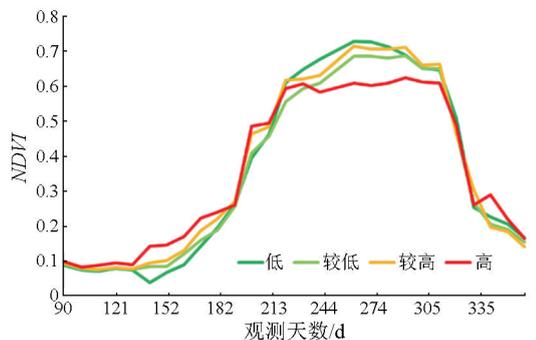


图 10 不同不确定性等级的水稻地块 NDVI 均值曲线
Fig.10 Mean NDVI curves of rice land-parcels with different uncertainty levels

第三, 道路密度是潜在的影响因素, 在基础设施建设完备、人口密度较大的地区, 农作物分类不确定性较低。该相关性是由于较为便利的交通条件保证

了田块管理人员及相关设备的灵活出入,进而为农作物提供了更为理想的生长条件,而地块内植被茂密、NDVI 特征明显、杂草等干扰信号微弱,分类不确定性因之降低。

第四,5 km 半径内的样本密度特征重要性排名靠前,意味着分类器对地块种植农作物进行判识时,受其一定半径内“近邻样本”的影响,拥有更多、更丰富类型的区域,其光谱响应特征将在整体分布中的体现更为显著,进而获得更高质量的分类依据。此外,我们发现半径为 2.5 km 和 1 km 的样本密度特征在建模过程中重要性较低,表明过小半径内的样本密度与不确定性关联有限,可能是由于样本数量以及样本类型丰富度不足,无法显著影响作物类型光谱响应特征的整体分布。

3 结论及展望

3.1 研究结论

针对农作物种植结构精准制图与可靠性评价问题,本文选择宁夏引黄灌区作为研究区,开展地块尺度的农作物分类及不确定性分析,得到如下结论:

1) 研究区内农作物分类的总体不确定性较低,制图结果较可靠,但由于不同农作物物候特征、光谱响应及播种规模等差异,模型区分能力存在一定的类间波动,其中水稻、菜地、苜蓿地块的分类不确定性较小,单种及复种小麦地块易于相互混淆,不确定性较大,玉米的低不确定性及高不确定性地块数量比例均较低。

2) 基于混合熵的不确定性评价指标综合考虑了在基于时序观测、借助深度学习网络进行作物分类过程的不同环节中,由于光谱特征自身混杂缺损、类别稳定性不足造成的先验随机性,以及分类模型最终结果的判别模糊性。实验结果与 F1 得分等传统的精度评价指标倾向性基本一致,但侧重点有所不同,可以视作对分类不确定性来源分解、综合度量的一种尝试,具有一定的科研及实用价值。

3) 较传统的混淆矩阵及像素级精度评价方案,本文构建的不确定性建模过程以地块作为空间“容器”,将不确定性分析结果与地理实体相匹配,直观展示了制图可靠性的空间分布模式,能为后续环境变量相关性分析、地面样本扩充及模型迭代优化提供指示信息。

4) 由随机森林回归模型的特征重要性解析得出,农作物分类的不确定性与种植作物类型、环境与农业管理条件等因素存在相关性:灌溉条件优越、地块形状规则的区域,农作物生物量较高,光谱特征

显著,不确定性相应较低;而对物候变化特征显著的农作物,模型的可分性较强,不确定性同样较低。

3.2 讨论及展望

本文借助混合熵模型,在地块尺度开展了农作物分类不确定性的评价,验证了基于混合熵指标评价农作物分类质量的有效性,展示不确定性的空间分布结构,并对造成不确定性的地理环境要素进行了定量解析。在此基础上,本研究在以下几个方面可进一步深入:

1) 对基于混合熵构建不确定性度量的机理性解释还有待进一步探索。

2) 地块尺度的种植结构制图精度包括形态精度和类型精度,本文仅聚焦于类型判识中的不确定性建模,缺乏针对形态不确定性的解析,后续可尝试引入交并比等形态精度指标。

3) 本研究的不确定性分析仅用于量化评价分类算法的输出结果,尚未将其融入实际制图流程中,对其支撑下的分类算法迭代优化研究还未开展,后续可在不确定性空间分布指示下进行目的性补样及模型强化学习,验证其提升制图精度的可行性和有效性。

参考文献 (References):

- [1] Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: A meta-review [J]. *Remote Sensing of Environment*, 2020, 236: 111402.
- [2] 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.
Wu B F, Zhang M, Zeng H W, et al. Agricultural monitoring and early warning in the era of big data [J]. *Journal of Remote Sensing*, 2016, 20(5): 1027-1037.
- [3] Mitchell P J, Downie A L, Dising M. How good is my map? A tool for semi-automated thematic mapping and spatially explicit confidence assessment [J]. *Environmental Modelling and Software*, 2018, 108: 111-122.
- [4] Olofsson P, Foody G M, Herold M, et al. Good practices for estimating area and assessing accuracy of land change [J]. *Remote Sensing of Environment*, 2014, 148: 42-57.
- [5] 陈焕南, 林红燕, 邢海花. 遥感影像数据分类的不确定性分析及其处理方法综述 [J]. *现代计算机*, 2021, 27(31): 33-39.
Chen H N, Lin H Y, Xing H H. Review of uncertainty analysis and processing methods of remote sensing image data classification [J]. *Modern Computer*, 2021, 27(31): 33-39.
- [6] Steele B M, Winne J C, Redmond R L. Estimation and mapping of misclassification probabilities for thematic land cover maps [J]. *Remote Sensing of Environment*, 1998, 66(2): 192-202.
- [7] Liu J J, Li Z, Qiao Y L, et al. A new method for cross-calibration of two satellite sensors [J]. *International Journal of Remote Sensing*, 2004, 25(23): 5267-5281.
- [8] Comber A, Fisher P, Brunsdon C, et al. Spatial analysis of remote

- sensing image classification accuracy[J].Remote Sensing of Environment,2012,127:237-246.
- [9] Yu Q,Gong P,Tian Y Q, et al.Factors affecting spatial variation of classification uncertainty in an image object-based vegetation mapping[J].Photogrammetric Engineering and Remote Sensing,2008,74(8):1007-1018.
- [10] 柏延臣,王劲峰.遥感数据专题分类不确定性评价研究:进展、问题与展望[J].地球科学进展,2005,20(11):1218-1225.
Bo Y C,Wang J F.Assessment on uncertainty in remotely sensed data classification:Progresses,problems and prospects[J].Advance in Earth Sciences,2005,20(11):1218-1225.
- [11] 郭逸飞,吴田军,骆剑承,等.基于不确定性迭代优化的山地植被遥感制图[J].地球信息科学学报,2022,24(7):1406-1419.
Guo Y F,Wu T J,Luo J C, et al.Remote sensing mapping of mountain vegetation via uncertainty-based iterative optimization[J].Journal of Geo-Information Science,2022,24(7):1406-1419.
- [12] Lyons M B,Keith D A,Phinn S R, et al.A comparison of resampling methods for remote sensing classification and accuracy assessment[J].Remote Sensing of Environment,2018,208:145-153.
- [13] Goodchild M F,Sun G,Yang S.Development and test of an error model for categorical data[J].International Journal of Geographical Information Systems,1992,6(2):87-103.
- [14] Brown K M,Foody G M,Atkinson P M.Estimating per-pixel thematic uncertainty in remote sensing classifications[J].International Journal of Remote Sensing,2009,30(1):209-229.
- [15] Comber A J.Geographically weighted methods for estimating local surfaces of overall,user and producer accuracies[J].Remote Sensing Letters,2013,4(4):373-380.
- [16] Khatami R,Mountrakis G,Stehman S V.Mapping per-pixel predicted accuracy of classified remote sensing images[J].Remote Sensing of Environment,2017,191:156-167.
- [17] Immitzer M,Vuolo F,Atzberger C.First experience with sentinel-2 data for crop and tree species classifications in central Europe[J].Remote Sensing,2016,8(3):166.
- [18] 孙营伟.复杂成像条件下地块级作物种植结构遥感制图研究[D].北京:中国科学院大学,2020.
Sun Y W.Study on remote sensing mapping of plot-level crop planting structure under complex imaging conditions[D].Beijing:University of Chinese Academy of Sciences,2020.
- [19] Liu Y,Cheng M M,Hu X, et al.Richer convolutional features for edge detection[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,41(8):1939-1946.
- [20] 刘巍.多源遥感协同的西南山地区耕地信息提取方法研究[D].北京:中国科学院大学,2021.
Liu W.Study on extraction method of cultivated land information in southwest mountainous area by multi-source remote sensing collaboration[D].Beijing:University of Chinese Academy of Sciences,2021.
- [21] Hochreiter S,Schmidhuber J.Long short-term memory[J].Neural Computation,1997,9(8):1735-1780.
- [22] 徐慧.遥感影像不确定性对分类结果可靠性影响模式研究[D].武汉:武汉大学,2020.
Xu H.Exploration of the influence model remote sensing image uncertainty on classification[D].Wuhan:Wuhan University,2020.
- [23] Shannon C E.A mathematical theory of communication[J].The Bell System Technical Journal,1948,27(3):379-423.
- [24] 吴浩,程志萍,史文中,等.遥感云分类不确定性的多维混合熵模型评价[J].测绘科学,2016,41(5):50-54,176.
Wu H,Cheng Z P,Shi W Z, et al.Evaluation on uncertainty of RS cloud classification based on multidimensional hybrid entropy model[J].Science of Surveying and Mapping,2016,41(5):50-54,176.
- [25] 刘艳芳,兰泽英,刘洋,等.基于混合熵模型的遥感分类不确定性的多尺度评价方法研究[J].测绘学报,2009,38(1):82-87.
Liu Y F,Lan Z Y,Liu Y, et al.Multi-scale evaluation method for uncertainty of remote sensing classification based on hybrid entropy model[J].Acta Geodaetica et Cartographica Sinica,2009,38(1):82-87.
- [26] Smith J H,Stehman S V,Wickham J D, et al.Effects of landscape characteristics on land-cover class accuracy[J].Remote Sensing of Environment,2003,84(3):342-349.
- [27] Ortiz A,Oliver G.On the use of the overlapping area matrix for image segmentation evaluation:A survey and new performance measures[J].Pattern Recognition Letters,2006,27(16):1916-1926.
- [28] Foody G M.Status of land cover classification accuracy assessment[J].Remote Sensing of Environment,2002,80(1):185-201.
- [29] Grömping U.Variable importance assessment in regression:Linear regression versus random forest[J].The American Statistician,2009,63(4):308-319.
- [30] Löw F,Michel U,Dech S, et al.Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines[J].ISPRS Journal of Photogrammetry and Remote Sensing,2013,85:102-119.

Remote sensing-based classification of crops on a farmland parcel scale and uncertainty analysis

ZHANG Dongyun¹, WU Tianjun², LI Manjia¹, GUO Yifei¹, LUO Jiancheng^{1,3}, DONG Wen¹

(1. State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; 2. School of Land Engineering, Chang'an University, Xi'an 710064, China; 3. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The rapid survey and accurate mapping of the spatial distribution of crops using remote sensing are

fundamental to modern precision agriculture. However, limitations in the acquisition, processing, and analysis of remote sensing images impact the mapping accuracy of traditional crop planting structures. Therefore, there is an urgent need to conduct spatial modeling and feature analysis for the uncertainty in crop classification. Using the Ningxia Yellow River irrigation area as a trial area and farmland parcels as the basic spatial units, this study classified crops on a parcel scale utilizing multi-source remote sensing data and machine learning algorithms. Then, an uncertainty calculation model was constructed based on mixed entropy, yielding the spatial distribution of the uncertainty of crop types in farmland parcels. Afterward, multi-source auxiliary data were employed to build a regression model for the uncertainty, and the potential impacts of related geographical variables on the uncertainty were explored. The experiment results indicate that 1.49 million vector units were constructed for the farmland parcels during the farmland extraction and classification session, yielding an overall crop classification accuracy of 0.80. The mapping results aligned well with the actual agricultural management units, and the classification results proved more better than the traditional pixel-based methods. The uncertainty in the parcel-scale crop classification was generally lower, with significant differences among crop types. The uncertainty was low for rice, vegetable plots, and alfalfa, relatively higher for wheat of single- and double-cropping patterns, and moderate for maize. The uncertainty in parcel-scale crop classification is influenced by various environmental factors such as planting structure and resource conditions, exhibiting the most significant correlations with crop type and water accessibility.

Keywords: remote sensing; crop classification; farmland parcel; uncertainty; machine learning; hybrid entropy

(责任编辑: 张 仙)