

基于迁移学习和非监督分类的制种玉米遥感识别方法

常婉秋^{1,2} 姚宇^{1,2} 席晓杰^{1,2} 刘哲^{1,2} 李绍明^{1,2} 张晓东^{1,2} 赵圆圆^{1,2}

(1. 中国农业大学土地科学与技术学院, 北京 100193; 2. 农业农村部农业灾害遥感重点实验室, 北京 100193)

摘要: 作物遥感识别主要基于监督分类方法, 对样本的数量、分布要求较高, 而农作物样本目视解译困难。为提高已采集样本的利用率, 同时降低精细分类中对样本的依赖, 本文将迁移学习与非监督分类方法相结合, 在源域内构建特征工程, 包括: BLUE、GREEN、RED、EDGE1、EDGE2、EDGE3、NIR、SWIR 8 个原始光谱波段, 以及 NDVI、EVI、RVI、GNDVI、TVI、DVI、MSAVI、GCVI、RNDVI、NDRE、RR11、RR12、MSRRE、CLRE、IRECI、LSWI、GCI、SIPI 18 个植被指数, 提取出最能表征制种玉米与大田玉米冠层光谱差异, 且在不同的源域内制种玉米之间差异最小的特征, 将其作为先验知识用于目标域的分类任务中, 再基于 K-means 进行制种玉米识别和制图。结果表明, 在众多特征中, 近红外原始波段表现出最强的优势, 且在制种玉米母本去雄期后表征效果最好。计算此时间段内 NIR 的线性回归斜率作为特征, 相较于直接基于 NIR 原始波段特征分类精度有所提升。利用 K-means 方法对 2019 年、2020 年石河子市和奎屯市的制种玉米分类, 2 个目标域制种玉米 2019 年 F1 值分别为 74.35% 和 64.97%, 2020 年 F1 值分别为 72.50% 和 75.69%。本方法通过提取先验知识, 引入非监督分类器, 有效提高了样本利用率。通过提取波段回归斜率作为特征为原始波段的特征增强提供了思路, 同时也为无样本场景下农作物精细分类绘图提供了方法。

关键词: 制种玉米; 遥感识别; 特征工程; K-means 非监督算法; 作物种间精细分类; 先验知识

中图分类号: S127

文献标识码: A

文章编号: 1000-1298(2024)08-0181-15

OSID:



Remote Sensing Identification for Seed Maize with Integrated Migration Learning and Unsupervised Classification

CHANG Wanqiu^{1,2} YAO Yu^{1,2} XI Xiaojie^{1,2} LIU Zhe^{1,2} LI Shaoming^{1,2}ZHANG Xiaodong^{1,2} ZHAO Yuanyuan^{1,2}

(1. College of Land Science and Technology, China Agricultural University, Beijing 100193, China

2. Key Laboratory of Remote Sensing for Agri-hazards, Ministry of Agriculture and Rural Affairs, Beijing 100193, China)

Abstract: Crop classification studies generally focus on different types of crops, while there are fewer studies on the fine classification of different cropping patterns of the same crop. Research on the spatial distribution of seed production is essential to control the maize market, as private seed production and concealment of acreage are occurring in the maize seed market. Seed maize and common maize are the two modes of maize cultivation. Accurate identification of both and remote sensing mapping of the spatial distribution of seed maize are essential for maize seed industry and food security. Traditional crop remote sensing classification methods require a high number and distribution of samples, while visual interpretation of crop samples is difficult. How to improve the utilization of collected samples and at the same time reduce the dependence on samples in fine classification is a pressing issue nowadays. Based on this, combining migration learning with unsupervised classification methods, firstly, using the idea of transfer learning, Linze and Wuwei were used as the source domain, and the feature engineering in the source domain was constructed, including 8 original spectral bands BLUE, GREEN, RED, EDGE1, EDGE2, EDGE3, NIR, SWIR, and 18 vegetation indices NDVI, EVI, RVI, GNDVI, TVI, DVI, MSAVI, GCVI, RNDVI, NDRE, RR11, RR12, MSRRE, CLRE, IRECI, LSWI, GCI, SIPI. Then, the features that best characterized the differences in canopy spectra between seed maize and common maize

收稿日期: 2023-11-14 修回日期: 2024-02-21

基金项目: 国家重点研发计划项目(2022YFB3903500)

作者简介: 常婉秋(1997—), 女, 博士生, 主要从事农作物种间精细分类研究, E-mail: changwq@cau.edu.cn

通信作者: 刘哲(1985—), 男, 教授, 博士, 主要从事作物品种多尺度遥感与评价研究, E-mail: liuz@cau.edu.cn

and that differed least between seed maize in different source domains were extracted. Finally, it was used as prior knowledge in an unsupervised classification task in the target domain. The results showed that the near-infrared primordial band exhibited the strongest advantage among the many features. By comparing the classification accuracies of three time-series ranges, namely, before the removal of male ears from the seed maize females, after the removal of male ears from the seed maize females, and during the full-life span of maize growth, the NIR bands were best characterized after the removal of male ears from the seed maize females, i. e. , when the DOY was 125 ~ 210. In order to further extract the NIR primary band information and enhance the performance of the features, the slope of the linear regression equation in the NIR band of the seed maize female after removal of male ears was used as a feature, and the K-means unsupervised method was used to classify the seed maize in the target domain of Shihezi and Kuitun. After comparative experiments on two target domains in 2019 and 2020, this method mostly showed different degrees of improvement over the K-means classification accuracies characterized by the near-infrared primitive bands in the same time period. In the two target domains Shihezi and Kuitun seed maize had F1 value of 74.35% and 64.97% in 2019 and 72.50% and 75.69% in 2020, respectively. This method effectively improved the utilization of samples by extracting prior knowledge and introducing an unsupervised classifier. It provided ideas for feature enhancement of the primary band by extracting the slope of the band regression as a feature, and provided a method for fine classification mapping of crops in sample-free scenarios.

Key words: seed maize; remote sensing identification; feature engineering; K-means unsupervised method; fine classification of crop interspecies; prior knowledge

0 引言

种业是国家基础性、战略性产业,优质、足量的种子供应是保障粮食安全的基石^[1]。制种玉米是全国大田玉米生产的种子来源,制种玉米的质量直接决定着全国玉米生产的产量^[2]。准确、快速地获得制种玉米种植信息可以有效预估下一年度全国玉米生产情况,因此对制种玉米空间分布的研究至关重要。传统的制种玉米种植监测采用逐级汇报的方式,这种方式在玉米的生长周期后期才能获得,时效性较差、效率较低。随着全球高空间分辨率、高时间分辨率、高光谱分辨率卫星数量的增加,近几年越来越多开源、稳定的影像数据源为基于遥感的制种玉米监测提供了基础数据源^[3]。

遥感监测大多基于监督分类方法^[4-7],这种方法精度较高,但需要大量的实地样本支撑,同时监督分类的方法受样本所在环境的影响较大,要求样本的覆盖类别充足且代表性较好。对于作物的样本不能直接通过对影像的目视解译获得,需要进行野外实地验证从而确保样本类别的正确性,而野外实地采集增加了样本获取的难度和成本,因此不借助样本的非监督方法在作物分类中逐步被人们研究和应用^[8-9]。

非监督方法没有对样本训练的过程,因而更加依赖选取的特征对不同作物的表征能力。光学遥感特征形式主要包括 2 种:①原始波段光谱值。随着光谱波段的增多,在对作物的表征能力上,不同的波段蕴含着不同的信息。CAI 等^[10]根据 Landsat 提供

的原始光谱构建特征工程,对于可见光、近红外以及短波红外对大豆和玉米的识别效果,实验发现短波光谱的表征能力最好。KANG 等^[11]基于 GF-6 WFV 影像对 2 个红边波段特征进行分析,实验发现 red edge 750 在作物分类中可以提供更多的有效信息。②植被指数值。通过波段间的线性组合构成植被指数,扩大对特定作物的表征能力。DA SILVA 等^[12]对比 NDVI、SAVI、EVI、NDWI 4 个植被指数在土地覆盖分类中的效果,实验结果表明使用单一的植被指数不能代表所有土地覆盖类别。ASGARI 等^[13]分别将 ARVI、CVI、EVI、LAI、GLI 和 NDVI 应用于作物分类中,其中 ARVI、LAI 的识别效果较好。目前对于多光谱和高光谱较多波段的利用还不够充分,较多使用可见光、近红外波段,而对其他波段的利用率较低。对于植被指数的研究也大多集中于常用的几种植被指数,而对其他波段组合以及波段间的多种线性运算对比研究较少。

传统的非监督分类算法是在不施加任何先验知识的情况下,仅通过光谱反射率本身之间的差异规律实现分类。GAIKWAD 等^[14]以 NDVI 植被指数时间序列作为特征,通过 K-means 聚类方法对植被进行分类,基于此对植被覆盖程度的变化进行分析。LAMPERT 等^[15]使用 K-means 的遥感约束时间序列聚类,实验发现与不受约束的聚类相比,向聚类问题添加约束可提高准确性。目前 K-means 方法因其可解释性强且易于实验而在遥感中被广泛使用。通过挖掘样本区的有效特征,并以先验知识的方式加入非监督分类系统中,是提高精度的一个方式。迁移

学习可以将样本区内学习到的知识迁移到无样本区内,实现无样本区的分类优化。HAO 等^[16]通过迁移学习与随机森林结合,实现 3 个实验区的作物分类。目前对迁移学习的利用大多基于监督分类算法,进行特征的迁移^[17-18]或者模型的迁移^[19-20],迁移学习对样本区的学习以及对空间的扩展性同样可以在非监督方法中加以应用。

本文首先通过迁移学习的方法,以临泽县和武威市作为源域,在源域内构建特征工程,再迁移到目标域石河子市和奎屯市,基于非监督算法进行分类,得到 2019 年和 2020 年目标域制种玉米分布图。通过 Sentinel-2 提供的多个光谱波段,构建特征工程,包含 8 个原始光谱波段和 18 个植被指数,并划分 3 个时序范围对比分类效果,筛选出能有效表征

制种玉米和大田玉米差异的特征指标以及最佳特征时序范围,作为 K-means 方法的先验知识,实现目标域石河子市和奎屯市的制种玉米识别。以期为无样本区的制种玉米识别提供研究思路。

1 研究区与数据源

1.1 研究区概况

以甘肃省临泽县、武威市和新疆维吾尔自治区石河子市和奎屯市为研究区,如图 1 所示。甘肃省和新疆维吾尔自治区位于中国西北部,具有广阔的耕地面积,是我国重要的核心玉米制种基地^[21]。较好的光照条件和灌溉技术,使得甘肃省和新疆维吾尔自治区成为玉米大范围、规模化种植的主产区^[22]。

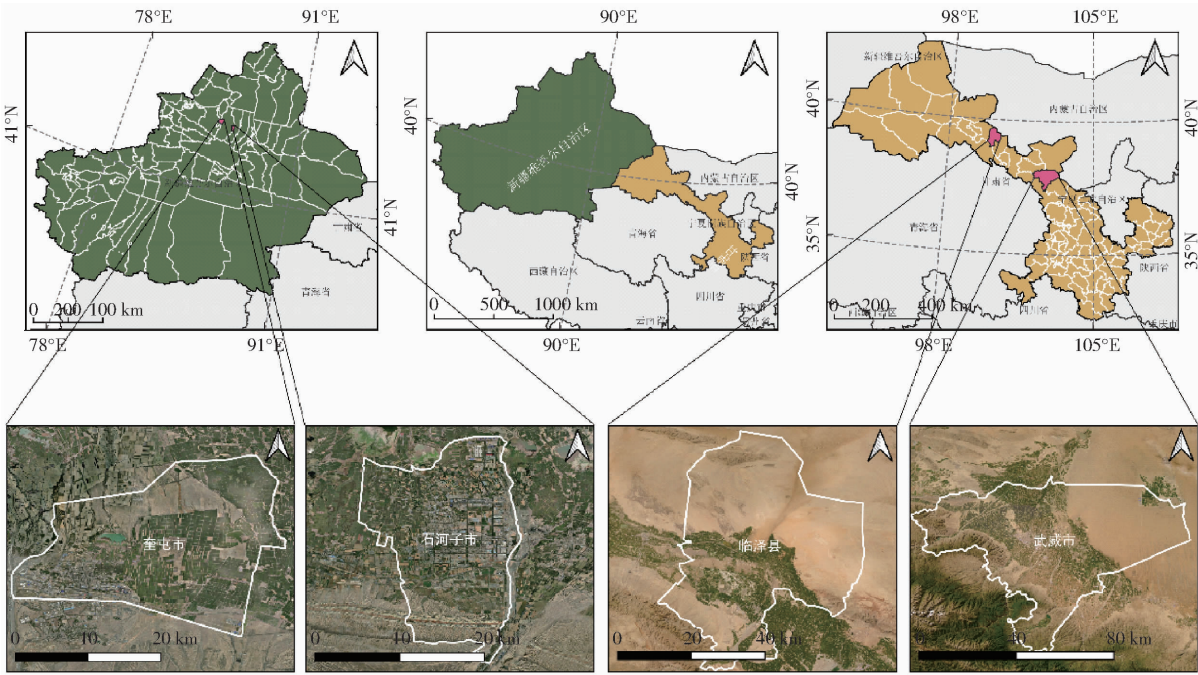


图 1 研究区分布

Fig. 1 Study area distribution

按照地形、气候,新疆维吾尔自治区和甘肃省分为伊犁河谷、天山北麓和河西走廊三大生态区。以天山为界,伊犁河谷位于天山腹地的河谷平原,天山北麓位于天山山脉和准噶尔盆地之间。河西走廊以黄河为界,位于黄河的西部,呈西北—东南走向的狭长走势。选择的研究区包含天山北麓和河西走廊生态区,其中临泽县位于河西走廊中部,武威市位于河西走廊东端,石河子市位于天山北麓中段,奎屯市位于天山北麓东部。本文将临泽县和武威市作为特征训练提取的研究区,石河子市和奎屯市作为迁移学习验证的区域。不同的生态区种植环境具有一定的相似性,4 个研究区所覆盖的耕地面积占比均较大,全年日照相对充足,年均累计日照时数为 3 000 h 左

右。除了光照条件,温度条件也十分适宜种植农作物,年均无霜期为 170 d 左右,全年大于等于 10℃ 的活动积温均在 3 570 ~ 3 729℃ 范围内。

1.2 影像数据

Sentinel-2 包含 Sentinel-2A 和 Sentinel-2B 2 颗卫星,携带 1 个多光谱成像仪 (Multispectral Imager, MSI),是高空间分辨率多光谱成像卫星,可以用于陆地变化监测以及提供植被、土壤、水覆盖等图像。Sentinel-2 卫星飞行高度为 786 km,覆盖 13 个光谱波段,包括 10 m 的蓝、绿、红、近红外波段,20 m 的红边、短波红外波段,60 m 的海岸/气溶胶波段,水蒸气波段,卷云波段;幅宽 290 km,重放周期为 5 d。在光学遥感数据源中,Sentinel-2 卫星影像数据

是唯一 1 颗在红边区域包含 3 个波长的卫星,这对植被信息的监测非常有效,是本文的数据源。欧洲航天局(European Space Agency, ESA)提供的 Sentinel-2 影像数据包括多个产品级别,本文使用的是 Level-2A 提供的经几何矫正和辐射矫正后的地球表面反射率产品。选取云量小于 10% 的影像,将所有波段重采样至 10 m 进行波段合成,并借助团队研发的多尺度格网化影像处理系统(Multilevel raster data cleaning and reconstitution multigrid system, RDCRMG)进行数据存储与计算,按照 RDCRMG 的组织形式,将遥感影像分割存储到 10 km 格网^[23-24]。

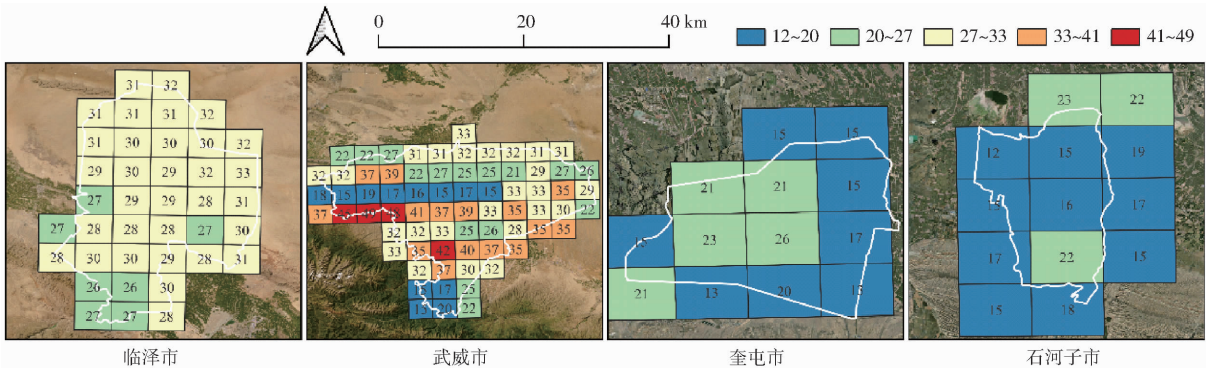


图 2 Sentinel-2 影像数量统计

Fig.2 Sentinel-2 image count statistics

1.3 样本数据

于 2019 年 7 月底和 2020 年 7 月底进行野外实地调查,在临泽县、武威市、石河子市、奎屯市进行样本采集。调研主要包括:种植的作物类型、田块位置及其地理坐标等。对于作物类型,本文综合多个分类系统,将农业部代码细分,最终得到表示作物类型的地信代码^[25],以此作为各类地物和作物类型的唯一标识。

4 个研究区实地采集的主要作物类型有制种玉米、大田玉米、小麦、棉花、辣椒、葡萄、林地等。在样本采集的路线设计中,本文以耕地区域为基准,从南至北、由东到西贯穿均匀分布,实地采集的各样本数量统计如表 1 所示。研究区内主要以种植一年生的作物为主,多年生的作物较少。除跨冬作物外,一般在每年的 3、4 月逐步开始各类作物的播种,在 10 月之前收获,生长周期为 7 个月左右,根据农业农村部统计情况以及野外实地采样情况,研究区主要作物物候历如图 3 所示。

本文重点对玉米的 2 种植场景(制种和大田)进行区分。目前,国内外普通大田玉米种植的均为杂交种。生产杂交种种子的制种田统称为制种玉米,需将杂交种的双亲(即父本、母本,目前均为纯合自交系)混合种植,以便父本的花粉能够授予母本,收获母本上的杂交种种子。玉米杂交种相比

研究区内影像数据量统计分布如图 2 所示,根据各格网影像数量,以研究区内影像数量的最小值和最大值作为划分区域,依据自然断点法的计算规则,即通过迭代比较每种分组方式,将研究对象分成组内质性相似即组内平方差之和最小,计算出自然间断点为 20、27、33、41,将所有影像覆盖数目据此分为 5 个等级并可视化。越偏蓝色表示影像数量越少,越偏红色表示影像数量越多。石河子市的高质量影像覆盖数较少,临泽县整体影像数量较石河子市多一些,且影像数量较为平均,武威市影像数量最多,各格网影像数量差距较其他地区更大。

自交系,其生物量、株高等优势高达 40%。在玉米制种田中,父本、母本一般单独成行,按照 1 行父本比 4~8 行母本混合种植。除此以外,制种玉米的田间操作也会有所不同,在玉米生育期内制种玉米会在抽穗期间采取去除母本雄穗的农业管理措施,并在授粉结束后采取砍除父本行的操作。同时在每年的 9 月玉米收获后进行冬小麦轮作。

表 1 野外实地调研样本数量统计

Tab.1 Statistics on spatial distribution of field research sample

年份	研究区	制种玉米	大田玉米	其他作物
2019	临泽县	107	52	29
	武威市	124	136	123
	石河子市	12	17	115
	奎屯市	4	12	118
	临泽县	360	15	352
2020	武威市	115	188	167
	石河子市	3	10	111
	奎屯市	6	21	207

2 研究方法

本文总体工作流程如图 4 所示,主要包括 2 部分:源域的特征工程构建;目标域的模型构建与精度评价。首先,将采集的样本和影像进行预处理,并存入团队开发的 RDCRMG 格网系统中;其次,以临泽

主要作物	2月			3月			4月			5月			6月			7月			8月			9月			10月			11月					
	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下			
制种玉米							播种			出苗			拔节			抽穗			成熟			收获											
大田玉米							播种			出苗			拔节			抽穗			成熟			收获											
冬小麦	越冬			返青			拔节			抽穗			成熟			收获									播种			出苗			越冬		
春小麦							播种			出苗			拔节			抽穗			成熟			收获											
棉花							播种			出苗			蕾期			花铃期			吐絮成熟			收获											
辣椒	播种			出苗			定植			开花坐果			结果			收获																	
葡萄							开花			生长			成熟			收获																	

图 3 研究区主要作物物候历
Fig.3 Phenology of main crops in study area

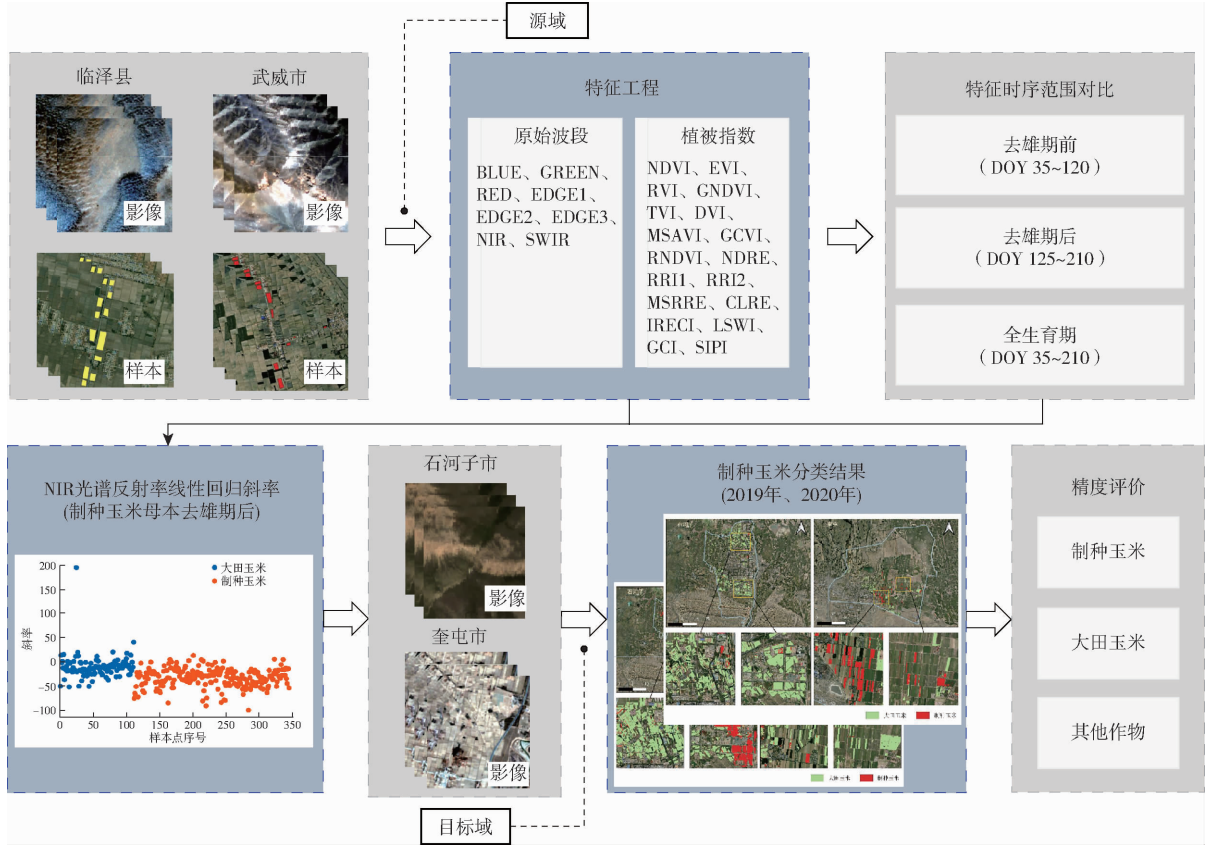


图 4 工作流程图
Fig.4 Workflow of this study

县和武威市作为源域,在源域内提取制种玉米和大田玉米的原始波段和植被指数时序,构建特征工程;然后对最能表征制种玉米的波段时序,根据农作知识设置 3 组不同时序范围实验,提取制种玉米母本去雄期后的近红外波段冠层反射值计算线性回归方程的斜率作为最终的分类特征;最后根据源域内建立的先验知识,通过 K-means 方法实现目标域石河子市和奎屯市无样本情况下的非监督分类,并在不同年际间检验此方法的可行性。

2.1 数据预处理

Sentinel-2 遥感影像数据通过空间维度重采样、波段合成等操作,对 L2A 级影像数据进行预处理。对于野外实地采集的样本,本文根据原始影像将样本地块的边界进行精细确定,并对属性表进行统一标识,增加 CropCode 作物类别唯一标识字段,用于分类结果的验证评价。将影像和样本同时剖分

到团队建立的 RDCRMG 格网系统中,提取包括样本位置、CropCode 编号、波段值的时间序列信息,存储到 txt 文档中,在模型输入中进行调用。

2.2 特征工程构建

根据 Sentinel-2 提供的植被指数,综合前人研究结果^[26-29],提取蓝波段 (BLUE)、绿波段 (GREEN)、红波段 (RED)、红边波段 (EDGE1、EDGE2、EDGE3)、近红外波段 (NIR)、短波红外波段 (SWIR) 8 个原始光谱波段,并计算归一化差值植被指数 (NDVI)^[30]、增强型植被指数 (EVI)^[31]、比值植被指数 (RVI)^[30]、绿色归一化差值植被指数 (GNDVI)^[30]、三角植被指数 (TVI)^[30]、差值植被指数 (DVI)^[30]、修改型土壤调节植被指数 (MSAVI)^[30]、叶绿素指数 (GCVI)^[32]、红边归一化差值植被指数 (RNDVI)^[33]、归一化差分红边指数 (NDRE)^[34]、红边比值植被指数 1 (RRI1)^[35]、红边

比值植被指数 2 (RRI2)^[35]、改良红边土壤调整植被指数 (MSRRE)^[35]、红边叶绿素指数 (CLRE)^[36]、倒置红边叶绿素指数 (IRECI)^[37]、地表水体指数 (LSWI)^[38]、绿色叶绿素指数 (GCI)^[39]、结构不敏感色素指数 (SIPI)^[40] 18 个植被指数,共计 26 种特征作为特征数据集。本文通过迁移学习,在临泽县和武威市的源域内构建特征工程,在石河子市和奎屯市的目标域内利用特征进行制种玉米识别。对研究区源域内每个样本像元提取其时间序列,获得其 26 种特征,对每个像元的特征计算取平均值,并绘制标准曲线作为特征筛选的可视化依据。在临泽县和武威市 2 个源域中相同作物的特征曲线越相似,则该特征用于迁移学习的效果越好;在 2 个源域中不同作物同一特征曲线的差异越大,则该特征对于表征制种玉米和大田玉米差异效果越好。提取优势特征,作为基于迁移学习非监督分类的依据。

2.3 基于迁移学习获取先验知识

迁移学习通过用已有的知识来学习新的知识,其主要目的是解决分类领域中样本缺失以及训练数据不足的问题^[41-42]。迁移学习中对已有的知识称做源域,在本文中可以理解为有样本的区域;对要学习的新知识称做目标域,在本文中可以理解为无样本的区域。以往研究中主要针对大宗作物或地物的非监督分类^[43-45],而对制种玉米的非监督分类研究较少。

在 2 个源域内对已有样本进行学习,分析以上提取的 26 个特征在表征制种玉米和大田玉米的差异且同种植方式在 2 个源域内的相似性上的优势,找到表征效果最好的特征,以此作为非监督分类的特征,进而实现目标域制种玉米识别。

2.4 基于非监督的制种玉米识别迁移学习模型构建

以武威市和临泽县为源域,根据制种玉米种植的农作知识,7 月玉米进入抽穗阶段,在此期间制种玉米会有去除母本雄穗并保留父本雄穗的田间作业。以此为时间节点,在 2019 年进行实验,每年选择去雄期前 DOY 为 35 ~ 120 时的影像、去雄期后 DOY 为 125 ~ 210 时的影像,全生育期 DOY 为 35 ~ 210 时的 Sentinel-2 影像设置对比实验。包括 6 组实验:2019 年临泽县 3 组不同时序范围的 26 个特征的分类精度对比和 2019 年武威市 3 组不同时序范围的 26 个特征的分类精度对比。对于有效特征的提取,根据在源域中学习到的先验知识,通过建立线性回归方程,以线性回归方程的斜率表征制种玉米和大田玉米差异。在通过线性回归提取特征的研究中,斜率被证实是有效衡量不同线性关系的重要特征^[46-48]。线性回归通过线性模型学习同种样本

标记间光谱冠层反射率的值^[49],进而实现预测输出标记值,基于线性回归建立的方程式为

$$f(x_i) = wx_i + b \tag{1}$$

式中 w ——拟合方程斜率

b ——拟合方程截距

x_i ——每个时相对应光谱反射率

在非监督学习中,本文采用 K-means 方法进行制种玉米的分类识别。K-means 属于非监督学习,是聚类算法中轻量且易实现的一种方法。与传统的回归方程、贝叶斯方程、支持向量机等方法相比,传统的方法需要给定类别的标签 y ,而聚类算法则可以在不给定样本标签 y ,仅提供特征向量 x 的情况下实现作物分类。K-means 方法^[50]是将样本聚类成 k 个特征一致的簇,目标是在特征空间中对附近的点进行分组。

基于 Python 自带的 scikit-learn 软件包,调用 sklearn.cluster.KMeans 方法实现 K-means 的非监督分类,通过 init = 'k-means' 设置初始化质心参数。以石河子市和奎屯市作为目标域对所提出的方法进行验证,选取制种玉米母本去雄期后 DOY 为 125 ~ 210 时的影像作为基础数据,并逐像元提取波段光谱时序特征的线性回归方程,计算斜率,通过在源域内训练得到的先验知识,基于 K-means 聚类方法对目标域的制种玉米分布进行制图。

2.5 精度评价

以 2019、2020 年临泽县和武威市的样本作为迁移学习特征工程构建的源域,以 2019、2020 年石河子市和奎屯市样本作为迁移学习目标域的验证样本。其中,验证样本将样本合并为制种玉米、大田玉米、其他作物。2019 年石河子市制种玉米、大田玉米和其他作物的验证样本像素数量分别为 13 073、10 302、83 689;奎屯市制种玉米、大田玉米和其他作物验证样本像素数量分别为 67 609、3 614、2 866。2020 年石河子市制种玉米、大田玉米和其他作物验证样本像素数量分别为 67 609、3 614、2 866;奎屯市制种玉米、大田玉米和其他作物验证样本像素数量分别为 112 681、7 588、2 814。

利用验证样本构建混淆矩阵对分类结果进行评估,可以得到 4 个精度评估指标:总体精度 (Overall accuracy, OA)、制图精度 (Producer accuracy, PA)、用户精度 (User accuracy, UA)、F1 值。总体精度为正确分类样本数占总样本数比例,制图精度为正确分类样本数占此类样本总数比例,用户精度为正确分类样本数占被分为此类的像元总数比例,F1 值为制图精度和用户精度调和均值,F1 值受极端值的影响较小,更适合评价不平衡数据的分类问题^[51]。根

据这 4 个指标得到作物的错分率 (Commission, FR) 和漏分率 (Omission, MR), 即错分成其他类别比例和漏分到这一类别比例。

3 实验结果与分析

3.1 特征工程构建

以临泽县和武威市作为源域构建特征工程, 对

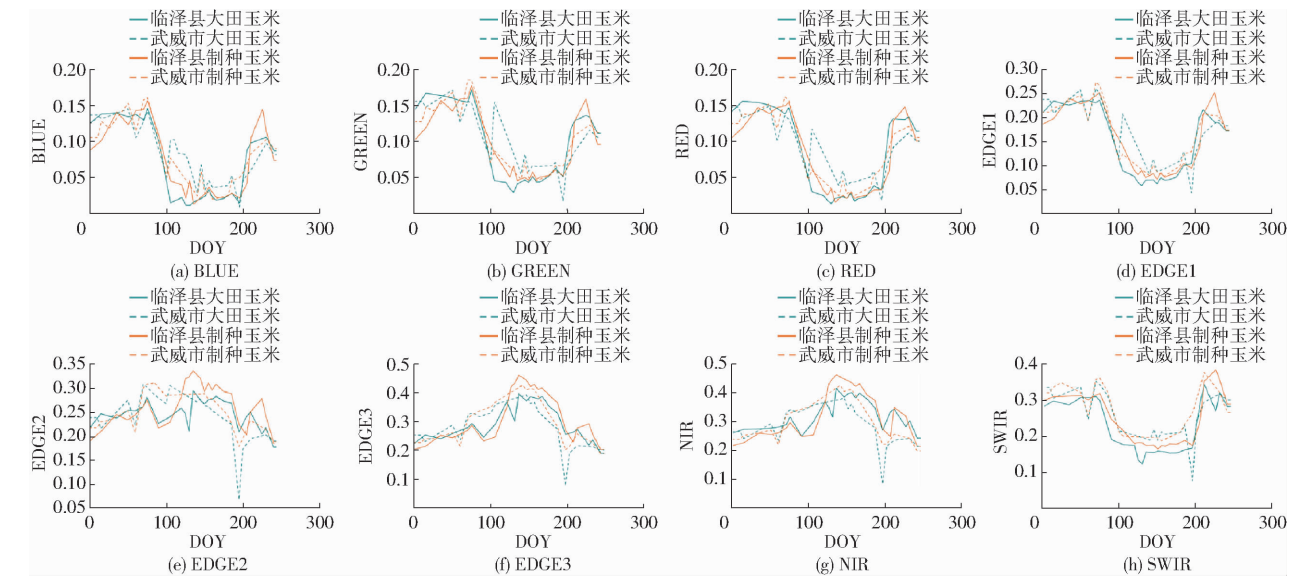


图 5 源域内原始光谱标准曲线对比

Fig. 5 Comparison of original spectral standard curves in the source domain

对每一原始光谱波段在源域内进行对比分析, 如图 5 所示。图 5a~5d 中, 在玉米生长的早期阶段即 DOY 小于 100 时, 武威市和临泽县的制种玉米差异较大, 不利于提取两者之间的共性, 同时在 DOY 处于 130~160 时, 临泽县的制种玉米和大田玉米光谱曲线变化趋势相同, 因此无法基于 BLUE、GREEN、RED、EDGE1 的反射率变化提取出两者之间的差异。图 5e 中, EDGE2 光谱曲线呈现三峰两谷的变化特征, 在武威市和临泽县制种玉米的变化趋势较一致, 但是两地的大田玉米反射率差异较大, 容易出现错分和混分。图 5f、5g 中, EDGE3 与 NIR 2 个波段的光谱曲线变化相似, 但是在 DOY 为 160~185 中, NIR 对不同地区同种植植场景的表征效果更好。图 5h 中, 在武威市出现 DOY 为 100~160 时, 制种玉米和大田玉米变化曲线重叠, 因此当时序时相涉及这些时间时, 无法对制种玉米和大田玉米进行区分。

对每一植被指数在源域内进行对比分析, 如图 6 所示。图 6a 中, 在 DOY 小于 100 时, 2 个源域内制种玉米和大田玉米在 NDVI 上的表征相似, 在之后的玉米生长发育中, 即 DOY 处于 100~200 时, NDVI 在武威市对于区分制种玉米和大田玉米表现出较好的优势, 但是在临泽县则不能成为区分两者

比 BLUE、GREEN、RED、EDGE1、EDGE2、EDGE3、NIR、SWIR 8 个原始光谱波段, 以及 NDVI、EVI、RVI、GNDVI、TVI、DVI、MSAVI、GCVI、RNDVI、NDRE、RR11、RR12、MSRRE、CLRE、IRECI、LSWI、GCI、SIPI 18 个植被指数筛选出适合进行迁移学习的特征, 临泽县和武威市标准特征对比如图 5、6 所示。

的特征, NDVI 对于制种玉米和大田玉米的表征受地理环境位置的影响较大。图 6b 中, EVI 和 NDVI 时序存在相似的问题, 在时相 DOY 小于 100 时, 不同地区以及制种玉米和大田玉米之间的差异均不明显, 当 DOY 处于 150~160 时, 在临泽县制种玉米和大田玉米的 EVI 几乎没有差异, 同样在大喇叭口阶段, 即 DOY 处于 120~150 时, 不同地区制种玉米的差异较大, 因此 EVI 不宜于作为制种玉米迁移学习的特征。图 6c 中, 在拔节抽穗后, 即 DOY 大于 100 时, 植被指数 RVI 仍能保持较大的波动, 对于不同的时间具有较敏感的光谱响应, 但是其变化规则相似, 即在相同的 DOY 时间, 制种玉米和大田玉米变化幅度和转折节点相似, 这不利于提取两者之间的差异。图 6d 中, 对于 GNDVI 时序曲线, 不同地区之间制种玉米在此植被指数上的差异大于同一地区制种玉米和大田玉米之间的差异, 这并不能够有效表征制种玉米生长特征。图 6e、6f、6h 中, 植被指数 TVI、DVI、MSAVI 光谱曲线变化规律相似, 在玉米播种和出苗阶段, 即 DOY 处于 0~80 时, 制种玉米和大田玉米有一定程度的差异, 但差异不明显, 不同地区之间同种玉米种植场景差异较小, 这有利于制种玉米早期识别。但是在玉米抽穗阶段, 即 DOY 处于 120~150, 这 3 种植被指数随时间的增加变化不明

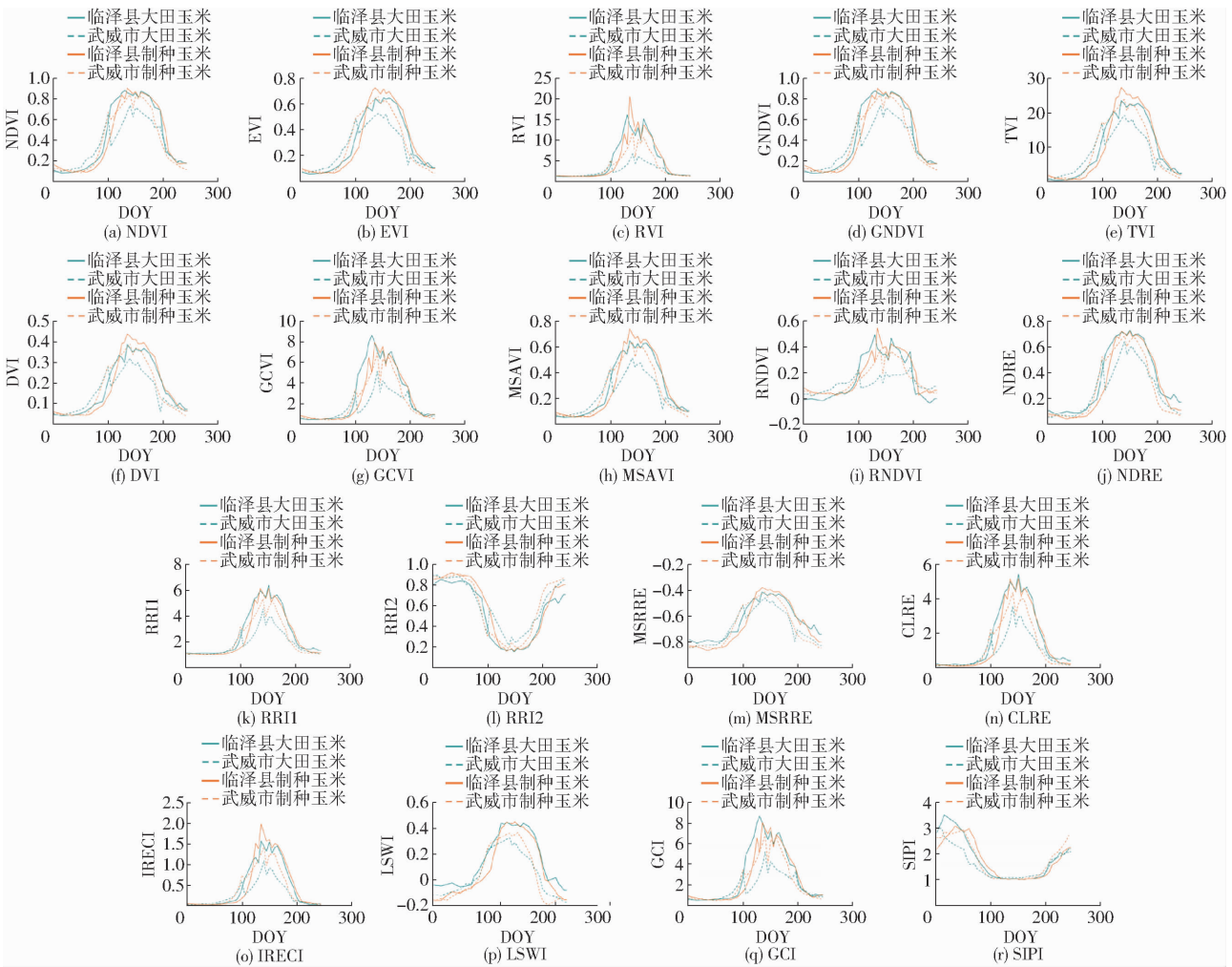


图 6 源域内植被指数标准曲线对比

Fig. 6 Comparison of vegetation index standard curves in source area

显,同时,在临泽县制种玉米和大田玉米差异减小,在武威市制种玉米和大田玉米差异增加,这也导致在不同的研究区玉米的同种种植场景差异增大,不利于制种玉米迁移训练。图 6g、6k、6m、6n、6o 中,植被指数 GCVI、RRI1、MSRRE、CLRE、IRECI 时序曲线在临泽县的制种玉米均表现出一个较高的峰值,在拔节抽穗阶段,即 DOY 处于 90 ~ 150,制种玉米和大田玉米存在错期的现象,但是在武威市和临泽县,同一种种植场景也出现错期的现象,这不利于提取同种植场景的共同特征。图 6i 中,RNDVI 能够很好地表征制种玉米和大田玉米之间的差异,但是对于不同研究区的大田玉米 RNDVI 曲线差异较大,在 DOY 处于 100 ~ 110 时,临泽县大田玉米植被指数曲线表现出增长趋势,而武威市的大田玉米则呈下降趋势。图 6l 中,植被指数 RRI2 曲线具有明显的双峰特征,并且不同地区之间的曲线错峰现象比制种玉米和大田玉米之间的错峰现象更明显,增加了制种玉米和大田玉米差异特征提取的难度。图 6j、6p 中,植被指数 NDRE、LSWI 曲线在拔节抽穗期后,即 DOY 处于 150 ~ 200 之间变化程度较为

平缓,对于分辨细微时序差异的玉米种植场景不适用。图 6q 和图 6r 分别表示 GCI 和 SIPI 2 种色素指数,两者在对玉米的表征上表现相反,以 DOY 为 100 为界,玉米生长发育前期,即 DOY 小于 100 时,GCI 不能区分制种玉米和大田玉米,而 SIPI 则在大田玉米和制种玉米上表现出差异;在玉米生长发育后期,即 DOY 大于 100 时,GCI 表征制种玉米和大田玉米差异在扩大,而 SIPI 则在缩小,因此两者在表征制种玉米和大田玉米差异上均存在一定问题。

图 5 和图 6 中的 DOY 为儒略日,从开始时间(3 月 1 日)以连续的日数计算时间。

除了直观地进行 2 个源域内原始光谱和植被指数标准曲线对比,本文在每个源域研究区内对每个特征进行非监督 K-means 聚类,以源域内实地采集的样本进行精度评价。临泽县验证样本数据集包括大田玉米 2 129 个像元、制种玉米 2 776 个像元,武威市验证样本数据集包括大田玉米 4 023 个像元、制种玉米 4 219 个像元。在每个特征的实验中,根据制种玉米在大喇叭口期会进行母本去雄的农业操作,即 DOY 为 120 ~ 125,基于此确定时间阈值设置

3 组不同时序范围的实验。时序范围分别为:DOY 为 35 ~ 120、125 ~ 210、35 ~ 210,从而对比得到识别

制种玉米最优的时序范围,精度结果如表 2 和表 3 所示。

表 2 临泽县 26 个特征 K-means 聚类精度对比

Tab.2 Comparison of K-means clustering accuracy for 26 features in Linze

DOY	35 ~ 120	波段	BLUE	GREEN	RED	EDGE1	EDGE2	EDGE3	NIR	SWIR
		OA/%	42. 22	42. 57	41. 67	57. 84	46. 56	55. 19	58. 72	43. 18
		植被指数	NDVI	EVI	RVI	GNDVI	TVI	DVI	MSAVI	GCVI
		OA/%	41. 10	40. 88	57. 98	41. 30	59. 14	41. 06	30. 85	36. 11
		植被指数	NDRE	RRI1	RRI2	MSRRE	CLRE	IRECI	LSWI	GCI
		OA/%	40. 22	36. 82	40. 04	48. 07	36. 82	40. 92	41. 47	36. 11
	125 ~ 210	波段	BLUE	GREEN	RED	EDGE1	EDGE2	EDGE3	NIR	SWIR
		OA/%	50. 81	62. 26	50. 89	52. 93	49. 19	43. 77	63. 18	35. 09
		植被指数	NDVI	EVI	RVI	GNDVI	TVI	DVI	MSAVI	GCVI
		OA/%	53. 41	42. 12	61. 41	55. 49	57. 61	58. 23	52. 42	41. 49
		植被指数	NDRE	RRI1	RRI2	MSRRE	CLRE	IRECI	LSWI	GCI
		OA/%	45. 61	41. 37	52. 84	58. 74	41. 37	45. 77	53. 03	41. 49
	35 ~ 210	波段	BLUE	GREEN	RED	EDGE1	EDGE2	EDGE3	NIR	SWIR
		OA/%	43. 45	42. 71	42. 02	45. 18	47. 91	56. 33	59. 20	55. 78
		植被指数	NDVI	EVI	RVI	GNDVI	TVI	DVI	MSAVI	GCVI
		OA/%	44. 85	57. 31	57. 31	41. 57	42. 92	58. 25	56. 29	57. 78
		植被指数	NDRE	RRI1	RRI2	MSRRE	CLRE	IRECI	LSWI	GCI
		OA/%	46. 12	36. 86	45. 30	50. 32	36. 86	46. 06	41. 04	39. 76

表 3 武威市 26 个特征 K-means 聚类精度对比

Tab.3 Comparison of K-means clustering accuracy for 26 features in Wuwei

DOY	35 ~ 120	波段	BLUE	GREEN	RED	EDGE1	EDGE2	EDGE3	NIR	SWIR
		OA/%	45. 91	55. 10	43. 92	42. 83	42. 21	36. 82	34. 20	43. 38
		植被指数	NDVI	EVI	RVI	GNDVI	TVI	DVI	MSAVI	GCVI
		OA/%	45. 37	45. 22	46. 54	45. 34	45. 22	40. 29	42. 17	45. 37
		植被指数	NDRE	RRI1	RRI2	MSRRE	CLRE	IRECI	LSWI	GCI
		OA/%	44. 18	43. 75	43. 25	44. 37	46. 30	40. 29	44. 09	39. 12
	125 ~ 210	波段	BLUE	GREEN	RED	EDGE1	EDGE2	EDGE3	NIR	SWIR
		OA/%	37. 50	42. 31	46. 01	44. 42	53. 34	27. 23	71. 48	44. 56
		植被指数	NDVI	EVI	RVI	GNDVI	TVI	DVI	MSAVI	GCVI
		OA/%	54. 53	44. 81	60. 30	54. 57	43. 31	68. 11	55. 35	55. 08
		植被指数	NDRE	RRI1	RRI2	MSRRE	CLRE	IRECI	LSWI	GCI
		OA/%	43. 52	44. 72	45. 54	36. 70	57. 57	45. 30	61. 74	48. 36
	35 ~ 210	波段	BLUE	GREEN	RED	EDGE1	EDGE2	EDGE3	NIR	SWIR
		OA/%	42. 39	44. 58	55. 54	55. 81	23. 30	27. 38	40. 00	44. 41
		植被指数	NDVI	EVI	RVI	GNDVI	TVI	DVI	MSAVI	GCVI
		OA/%	45. 78	45. 29	39. 74	54. 66	45. 11	44. 94	50. 45	48. 36
		植被指数	NDRE	RRI1	RRI2	MSRRE	CLRE	IRECI	LSWI	GCI
		OA/%	55. 01	42. 43	44. 18	35. 68	42. 30	42. 48	44. 39	44. 92

在临泽县研究区中,如表 2 所示,从整体来看,精度最高的是 DOY 为 125 ~ 210 时的 NIR 波段,为 63. 18%。在基于 NIR 波段的分类中,精度由大到小依次为:DOY 为 125 ~ 210 时、DOY 为 35 ~ 210 时、DOY 为 35 ~ 120 时的精度,说明 NIR 的最优时序范围为制种玉米母本去雄后,即 DOY 大于 125 时,在这一时期 NIR 对 2 种玉米场景的表征能力优

于全生育期,在使用 NIR 波段进行分类时应去除掉玉米生长发育前期时序的影响。从每个时序范围来看,在玉米生长发育早期 DOY 为 35 ~ 120 时,TVI 的 K-means 分类精度最高为 59. 14%,其次是 NIR 的分类精度为 58. 72%;在制种玉米母本去雄后 DOY 为 125 ~ 210 时,NIR 的 K-means 分类精度最高为 63. 18%,其次是 GREEN 的分类精度为 62. 26%;

在玉米全生育期 DOY 为 35 ~ 210 时,NIR 的 K-means 分类精度最高为 59.20%,其次是 DVI 的分类精度为 58.25%。从每个特征的不同时序范围对比来看,图 5 和图 6 中玉米早期生长阶段大田玉米和制种玉米差异不明显的特征 EDGE2、NIR、RVI、DVI、MSAVI、GCVI、RRI1、CLRE、IRECI、GCI,在表 2 的 K-means 聚类精度表现上,DOY 为 125 ~ 210 时的分类精度大于 DOY 为 35 ~ 120 时的精度,说明这些特征在去雄后的阶段更有利于表征制种玉米和大田玉米的差异;TVI 和 SIPI 特征在 DOY 为 35 ~ 120 时的精度要大于 DOY 为 125 ~ 210 时的精度,同时也大于 DOY 为 35 ~ 210 时的精度,与之对应图 5 和图 6 中制种玉米和大田玉米也在 DOY 为 35 ~ 120 时表现出更强的表征能力,说明在玉米生长发育早期,TVI 和 SIPI 对于不同的玉米种植场景表现效果优于制种玉米母本去雄后,并且去雄后这 2 个特征对分类的影响大于玉米生长早期,导致全生育期 DOY 时序的分类精度低于玉米生长发育的早期 DOY 时序阶段;其他原始波段和植被指数在 3 个不同的时序范围内没有表现出非常明显差异,与图 5 和图 6 中特征曲线情况一致。

在武威市研究区中,如表 3 所示,从整体上看在 DOY 为 125 ~ 210 时的 NIR 波段特征具有最优分类效果,精度为 71.48%,在武威市 NIR 全生育期的特征效果优于早期 NIR 波段特征,即 DOY 为 35 ~ 210 的精度大于 DOY 为 35 ~ 120 的精度,但与去雄后的 DOY 为 125 ~ 210 时的精度差距较大,因此 NIR 在 DOY 大于 125 时具有优异的分类特征。从每个时序范围来看,当 DOY 为 35 ~ 120 时,分类效果最好的特征是 SIPI,精度为 47.74%;当 DOY 为 125 ~ 210 时,分类效果最好的特征是 NIR,精度为 71.48%;当 DOY 为 35 ~ 210 时,分类效果最好的特征是 EDGE1,精度为 55.81%。从每个特征的不同时序范围对比来看,武威市同样是当 RVI、MSAVI、GCVI、RRI1、CLRE、IRECI、GCI 特征在制种玉米去雄后,即 DOY 为 125 ~ 210 时表现出更强的制种玉米和大田玉米差异表征能力;在 TVI 和 SIPI 特征的不同时序中,去雄前即 DOY 为 35 ~ 120 时,特征表征效果更好。除此以外,在武威市大部分植被指数在 DOY 为 35 ~ 120 时对大田玉米的识别效果好于制种玉米,即识别正确的大田玉米像元数远大于识别正确的制种玉米。随着时序越接近玉米生育期的后期,大部分植被指数对制种玉米的识别能力迅速增加,即在 DOY 为 125 ~ 210 阶段时,相比 DOY 为 35 ~ 120 阶段,识别正确的制种玉米像元个数逐渐增多,其中 NDVI 增加 4 072 个像元,EVI 增加 99 个

像元,RVI 增加 2 751 个像元,GNDVI 增加 4 013 个像元,TVI 增加 1 728 个像元,DVI 增加 3 673 个像元,MSAVI 增加 472 个像元,GCVI 增加 2 968 个像元,RNDVI 增加 409 个像元,NDRE 增加 401 个像元,RRI1 增加 1 268 个像元,RRI2 增加 72 个像元,MSRRE 增加 66 个像元,CLRE 增加 3 076 个像元,IRECI 减少 35 个像元,LSWI 增加 3 869 个像元,GCI 增加 2 037 个像元,SIPI 减少 475 个像元,由此说明制种玉米去雄是大部分植被指数识别制种玉米的关键节点。

综合图 5、6 和表 2、3 可发现,表征制种玉米和大田玉米冠层光谱差异最明显的阶段为玉米生育期的喇叭口、抽穗、成熟期,尤其是制种玉米母本去雄后,即 7 月下旬至 9 月初,DOY 处于 120 ~ 200。通过对比发现,大部分原始光谱和植被指数无法作为制种玉米识别有效特征,主要包括 2 个问题:①在某些时相内无法表征制种玉米和大田玉米的差异,即两者的特征存在重叠的现象。②在不同的区域之间无法表征相同种植场景的一致性,即在 2 个区域中制种玉米与制种玉米、大田玉米与大田玉米之间的特征差异较大。其中 NIR 波段特征表现出相对较好的优势,在喇叭口期制种玉米母本去雄后对两者的表征效果较好。7 月下旬至 9 月初,临泽县和武威市制种玉米在近红外波段较大田玉米的光谱曲线下下降速度有明显加快的现象,同时在对每一研究区对比中,制种玉米和大田玉米均存在光谱反射率的高低差异,在以此为聚类特征的分类上也表现出最优精度。由此可以初步说明近红外波段 NIR 是本文所构建的特征工程中表征制种玉米和大田玉米差异最大,以及在不同地区之间泛化能力最好的光谱特征。

3.2 制种玉米识别迁移学习分类与精度评价

为了提高近红外波段 NIR 在 DOY 为 125 ~ 210 时对制种玉米和大田玉米差异的表征能力,本文计算此时序范围内 NIR 的时序回归斜率作为 K-means 聚类的最终特征。对源域内制种玉米和大田玉米的样本点计算近红外波段的光谱反射率曲线,并选取 7 月下旬至 9 月初的光谱值,根据式(1)构建源域内制种玉米和大田玉米的线性回归方程,提取出每个样本点的斜率,散点形式可视化如图 7 所示。在 2 个源域内大田玉米的线性回归方程的斜率在 0 附近且相对集中,而制种玉米的斜率则普遍小于 0 且相对分散。因此在母本去雄的田间作业后至 9 月初期间,依据 NIR 波段提取的线性拟合直线斜率特征,在表征制种玉米和大田玉米差异上具有竞争力。

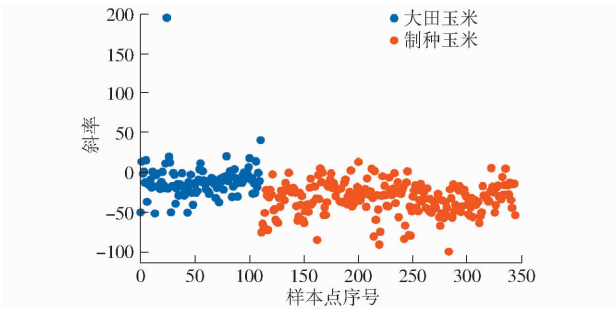


图 7 源域内制种玉米和大田玉米线性回归方程斜率分布
Fig. 7 Distribution of slope of linear regression equation for seed and common maize in source domain

为了探究先验知识的空间可迁移性和时间可迁移性,针对石河子市和奎屯市 2 个目标域分别进行 2019 年和 2020 年的迁移实验。在每年的每个目标域中,以 NIR 波段 DOY 为 125 ~ 210 时的光谱曲线斜率作为分类特征,对制种玉米进行 K-means 非监督分类。将石河子市和奎屯市采集的样本分成 2 个测试样本集,并将样本分为制种玉米、大田玉米、其他作物 3 类,制种玉米识别精度如图 8 所示。

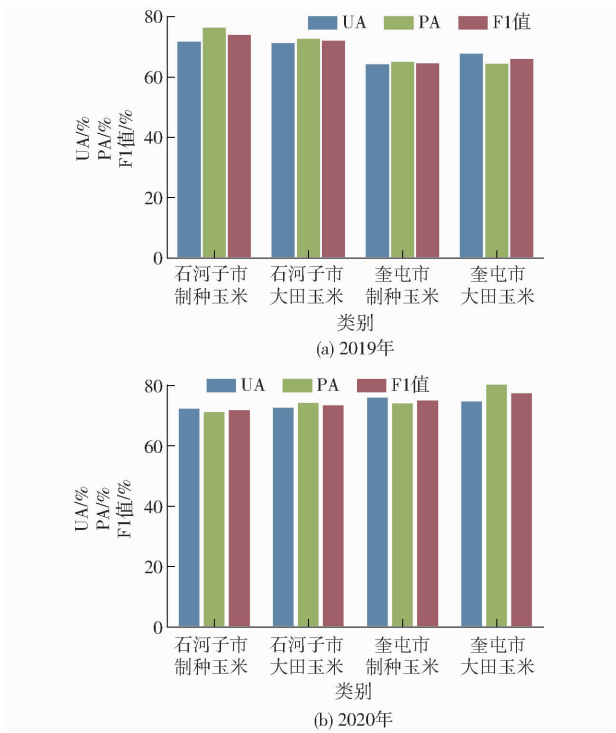


图 8 制种玉米分类精度
Fig. 8 Evaluation of classification accuracy of seeded maize

2019 年石河子市制种玉米用户精度为 76.73%,生产者精度为 72.12%,错分率为 23.27%,漏分率为 27.88%,综合来看,以 F1 值作为制种玉米最终精度评价,为 74.35%,大田玉米 F1 值为 72.34%;奎屯市制种玉米错分率为 34.58%,漏分率为 35.03%,F1 值为 64.97%,大田玉米 F1 值为 66.41%。2020 年石河子市制种玉米错分率为 28.08%,漏分率为 27.50%,F1 值为 72.50%,大田

玉米 F1 值为 74.17%;奎屯市制种玉米错分率为 25.27%,漏分率为 24.31%,F1 值为 75.69%,大田玉米 F1 值为 78.07%。

直接利用 DOY 为 125 ~ 210 时 NIR 原始波段作为 K-means 聚类的特征,在 2 个源域内的 OA 分别为 63.18% 和 71.48%,平均 OA 为 67.33%。通过进一步对 DOY 为 125 ~ 210 之间的 NIR 波段光谱响应进行分析,提取线性拟合斜率作为 K-means 聚类的特征,在 2019 年的迁移实验中石河子市制种玉米 F1 值提高 7.02%,奎屯市制种玉米 F1 值下降 0.92%。经调查,2019 年奎屯市在 3—5 月的气温比历年略高,玉米的生长并不是完全按照月份和累计时间进行的,而主要受活动积温影响,由于当地当年的温度较高,导致活动积温累计速度较快,较早地达到玉米抽穗拔节的时间,因而在 2019 年奎屯市分类精度没有石河子市高。在 2020 年迁移实验中石河子市制种玉米 F1 值提高 5.17%,奎屯市制种玉米 F1 值提高 8.36%。

将石河子市和奎屯市制种玉米分类结果进行可视化,并通过分类后处理消除椒盐效应,最终石河子市和奎屯市 2019 年和 2020 年制种玉米、大田玉米空间分布如图 9 和图 10 所示,包括制种玉米的整体分布和细节展示。综合 2 个目标域的制种玉米空间分布图来看,制种玉米呈集中分布,且地块规整,边界清晰,符合制种玉米实际种植场景。

4 讨论

4.1 本方法特点

本文对传统的非监督作物分类方法做了一定的改进,具有以下特点:①特征工程构建较为全面。本文提取 Sentinel-2 的 BLUE、GREEN、RED、EDGE1、EDGE2、EDGE3、NIR、SWIR 8 个原始波段,并计算 NDVI、EVI、RVI、GNDVI、TVI、DVI、MSAVI、GCVI、RNDVI、NDRE、RRI1、RRI2、MSRRE、CLRE、IRECI、LSWI、GCI、SIPI 18 个植被指数构建特征工程,从而使得得出的最优特征更具有说服力。②结合农作知识对原始特征表征能力进行增强。将制种玉米特有的农作知识制种玉米母本去雄时间引入特征时序范围筛选中,以此为时间节点建立 3 组时序范围精度对比实验,去雄期前、去雄期后和全生育期。并通过提取最优时序范围的 NIR 线性回归斜率优化 NIR 原始波段特征信息。③为 K-means 聚类引入先验知识。通过对源域内的样本进行分析提取出最优特征,在目标域对每个像元按此特征计算规则进行计算,再基于 K-means 逐像元进行分类。

通过与历年各地统计年鉴中记录的制种玉米种



图 9 2019 年目标域制种玉米识别结果

Fig.9 Results of identification of seed maize in target domain in 2019



图 10 2020 年目标域制种玉米识别结果

Fig.10 Results of identification of seed maize intarget domain in 2020

植面积相对比,从另一个角度评估本文所提方法分类的效果,统计结果如表 4 所示。F1 值越高的年份和地区计算出的制种玉米种植面积,与统计年鉴中的数值越接近。

4.2 与现有方法比较

目前对于作物分类中非监督方法的研究大多在无样本区直接进行分类^[4],但这种方法忽略了对其他地区样本有效信息的利用,同时直接进行非监督

表 4 预测面积与统计年鉴逐年对比
Tab.4 Year-by-year comparison of projected area and statistical yearbook

年份	研究区	预测像元数	预测种植面积/hm ²	统计年鉴面积 ^[53] /hm ²	样本分类 F1 值/%
2019	石河子市	160 225	1 602. 25	1 550	74. 35
	奎屯市	43 495	434. 95	284	64. 97
2020	石河子市	14 295	142. 95	416. 67	72. 50
	奎屯市	71 291	712. 91	640	75. 69

分类所提取的特征信息质量和冗余度无法进行评估^[52],因此本文提出一种在无样本区对作物种间精细分类的新方法,通过迁移学习的思想,首先对研究区的源域建立基于样本的特征工程,提取优势特征及时序,以先验知识的形式加入非监督算法中,再对研究区的目标域进行制种玉米识别。

在特征工程的构建中,与 QIONG 等^[54]对比的 10 种植被指数相比,本文还加入对原始光谱值的研究。REN 等^[55]研究证实,原始光谱波段比常用的植被指数在分类过程中的贡献更大。相较于文献^[55]对比的 8 个传统的植被指数,本文加入了更多 Sentinel-2 提供的植被指数以及原始光谱值进行对比,使得 NIR 波段特征的优越性更具有说服力。相较于传统的时序范围对比实验,与 HAO 等^[16]的常规的时序增量实验不同,本文结合制种玉米母本去雄的农作知识,对玉米种植全生育期时序进行划分对比分类精度,从而得到识别制种玉米最优的时序范围,这样的划分更能结合制种玉米本身的种植特点。

在非监督分类模型构建中,与 GAIKWAD 等^[14]的研究相比,本文不是直接以 NDVI 作为分类特征,而是基于对源域内样本的特征工程的研究,以此作为先验知识,提取对表征制种玉米和大田玉米冠层光谱差异更有效的特征,进而进行非监督分类。通过将迁移学习的思想与非监督分类算法进行结合,既解决了传统的监督分类对样本量依赖过大的问题,同时也提高了对样本的有效利用率。

4.3 未来工作

研究区的选择对作物的种间精细分类非常重要,研究区内作物种植的种类丰富度、栽培管理措施多样性、种植密度等都会对有效特征的提取产生一定的影响。基于光学遥感影像的分类主要是依据作物冠层之间的差异得到,丰富度高的样本可以提高特征的泛化性。除此以外,卫星遥感数据的发展也为多源数据的融合创造了条件,对于以时序为基础的分类问题,时序的质量和长度十分关键,在今后的研究中可以从多源数据融合的角度,利用 GF-1、Sentinel-2、Landsat8 等光学遥感卫星影像,获得更

完整、更准确的时间序列,从而提升分类特征的效果。其次在非监督分类算法上,本文主要是通过优化特征进而提升分类精度,在今后的研究中可以从提高分类算法的角度继续提高分类精度。对于无样本区的制图问题,本研究通过迁移特征的方法实现,样本生成也是目前的热门方法,以有样本区的样本为基准,在无样本区对每种作物生成相似的样本,实现分类,今后的研究可以从这一角度去探究。

5 结论

(1) 基于迁移学习的思想,在研究区内选取临泽县和武威市作为源域,对比 BLUE、GREEN、RED、EDGE1、EDGE2、EDGE3、NIR、SWIR 8 个原始光谱波段,以及 NDVI、EVI、RVI、GNDVI、TVI、DVI、MSAVI、GCVI、RNDVI、NDRE、RRI1、RRI2、MSRRE、CLRE、IRECI、LSWI、GCI、SIPI 18 个植被指数,构建制种玉米和大田玉米的特征工程,筛选出表征制种玉米冠层光谱特点最强的特征,以及最优特征时序的范围。以制种玉米母本去雄后,DOY 为 125~210 时的 NIR 原始光谱时序斜率作为特征,在 2019 年和 2020 年石河子市和奎屯市进行迁移制图。

(2) NIR 原始波段在众多 Sentinel-2 提供的波段和植被指数中,对描述武威市和临泽县制种玉米和大田玉米差异性,以及表征武威市和临泽县之间的制种玉米和制种玉米、大田玉米和大田玉米的相似性上表现出最强的竞争力。

(3) 在基于 NIR 原始波段的 K-means 分类中,精度由大到小顺序为:DOY 为 125~210 时、DOY 为 35~210 时、DOY 为 35~120 时的精度,即最优波段 NIR 在制种玉米和大田玉米的特征差异主要集中在制种玉米母本去雄后至成熟期,平均 OA 为 67.33%。

(4) 以 NIR 时序斜率作为增强特征的 K-means 聚类,2019 年目标域制种玉米平均 F1 值为 69.66%,2020 年目标域制种玉米平均 F1 值为 74.09%。比直接基于 NIR 原始光谱时序的 K-means 聚类精度高,本文所提的方法计算较为简单且基本满足对制种玉米迁移识别的要求。

参 考 文 献

[1] 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.
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. (in Chinese)

[2] 刘哲,曲艺伟,赵祖亮,等. 玉米优良品种推广重心转移及扩散的时空规律[J]. 农业工程学报,2018,34(1):178-185.
LIU Zhe, QU Yiwei, ZHAO Zuliang, et al. Temporal and spatial law of promotion center moving and diffusion of excellent maize varieties[J]. Transactions of the CSAE, 2018, 34(1): 178-185. (in Chinese)

[3] 孙伟伟,杨刚,陈超,等. 中国地球观测遥感卫星发展现状及文献分析[J]. 遥感学报,2020,24(5):479-510.

- SUN Weiwei, YANG Gang, CHEN Chao, et al. Development status and literature analysis of China's earth observation remote sensing satellites[J]. *Journal of Remote Sensing*, 2020, 24(5): 479–510. (in Chinese)
- [4] 张超, 童亮, 刘哲, 等. 基于多时相 GF-1 WFV 和高分纹理的制种玉米田识别[J]. *农业机械学报*, 2019, 50(2): 163–168.
- ZHANG Chao, TONG Liang, LIU Zhe, et al. Identification method of seed maize plot based on multi-temporal GF-1 WFV and Komsat-3 texture[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(2): 163–168. (in Chinese)
- [5] 张超, 乔敏, 刘哲, 等. 基于时序光谱和高分纹理分析的制种玉米田遥感识别[J]. *农业机械学报*, 2018, 49(5): 218–225.
- ZHANG Chao, QIAO Min, LIU Zhe, et al. Seed maize field identification based on analysis of remote sensing timing spectrum and high resolution texture[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(5): 218–225. (in Chinese)
- [6] WANG Z, ZHAO Z, YIN C. Fine crop classification based on UAV hyperspectral images and random forest[J]. *ISPRS International Journal of Geo-Information*, 2022, 11(4): 252.
- [7] YANG N, LIU D, FENG Q, et al. Large-scale crop mapping based on machine learning and parallel computation with grids[J]. *Remote Sensing*, 2019, 11(12): 1500.
- [8] MAHESH B. Machine learning algorithms—a review[J]. *International Journal of Science and Research (IJSR)*, 2020, 9: 381–386.
- [9] RIVERA A J, PEREZ-GODOY M D, ELIZONDO D, et al. Analysis of clustering methods for crop type mapping using satellite imagery[J]. *Neurocomputing*, 2022, 492: 91–106.
- [10] CAI Y, GUAN K, JIAN P, et al. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach[J]. *Remote Sensing of Environment*, 2018, 210: 35–47.
- [11] KANG Y, MENG Q, LIU M, et al. Crop classification based on red edge features analysis of GF-6 WFV data[J]. *Sensors*, 2021, 21(13): 4328.
- [12] DA SILVA V S, SALAMI G, DA SILVA M I O, et al. Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification[J]. *Geology, Ecology, and Landscapes*, 2020, 4(2): 159–169.
- [13] ASGARI S, HASANLOU M. A comparative study of machine learning classifiers for crop type mapping using vegetation indices[J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2023, 10: 79–85.
- [14] GAIKWAD S V, VIBHUTE A D, KALE K V, et al. Vegetation cover classification using Sentinel-2 time-series images and K-means clustering[C]//2021 IEEE Bombay Section Signature Conference, 2021: 1–6.
- [15] LAMPERT T, LAFABREGUE B, GANCARSKI P. Constrained distance based k-means clustering for satellite image time-series[C]//IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, 2019: 2419–2422.
- [16] HAO P, DI L, ZHANG C, et al. Transfer learning for crop classification with cropland data layer data (CDL) as training samples[J]. *Science of the Total Environment*, 2020, 733: 138869.
- [17] PAYMODE A S, MALODE V B. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG[J]. *Artificial Intelligence in Agriculture*, 2022, 6: 23–33.
- [18] ZHANG S, YANG J, LENG P, et al. Crop type mapping with temporal sample migration[J/OL]. *International Journal of Remote Sensing*, 2023: 1–19. <https://doi.org/10.1080/01431161.2023.2192881>.
- [19] BOSILJ P, APTOULA E, DUCKETT T, et al. Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture[J]. *Journal of Field Robotics*, 2020, 37(1): 7–19.
- [20] THENMOZHI K, REDDY U S. Crop pest classification based on deep convolutional neural network and transfer learning[J]. *Computers and Electronics in Agriculture*, 2019, 164: 104906.
- [21] 吕小瑞. 加快推进甘肃省国家级玉米制种基地建设的思考[J]. *农业科技与信息*, 2013(6): 28–30.
- LYU Xiaorui. Thoughts on accelerating the construction of national corn seed production base in Gansu[J]. *Agricultural Science-Technology and Information*, 2013(6): 28–30. (in Chinese)
- [22] 付聪林. 甘州区玉米制种面积超 73 万亩[N]. *甘肃经济日报*, (2023-06-01)(3).
- [23] YE S, LIU D, YAO X, et al. RDCRMG: a raster dataset clean & reconstitution multi-grid architecture for remote sensing monitoring of vegetation dryness[J]. *Remote Sensing*, 2018, 10(9): 1376.
- [24] XIONG Q, WANG Y, LIU D, et al. A cloud detection approach based on hybrid multispectral features with dynamic thresholds for GF-1 remote sensing images[J]. *Remote Sensing*, 2020, 12(3): 450.
- [25] 费世民. 农林业系统分类研究综述[J]. *四川林业科技*, 1993, 14(2): 27–32.
- FEI Shimin. Review of research on classification of agroforestry systems[J]. *Sichuan Forestry Science and Technology*, 1993, 14(2): 27–32. (in Chinese)
- [26] SONOBE R, YAMAYA Y, TANI H, et al. Crop classification from Sentinel-2-derived vegetation indices using ensemble learning[J]. *Journal of Applied Remote Sensing*, 2018, 12(2): 026019.
- [27] GAO L, WANG X, JOHNSON B A, et al. Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: a review[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 364–377.
- [28] VISKOVIĆ L, KOSOVIĆ I N, MASTELIĆ T. Crop classification using multi-spectral and multitemporal satellite imagery with machine learning[C]//2019 International Conference on Software, Telecommunications and Computer Networks, 2019: 1–5.

- [29] ZHOU Z, LI S, SHAO Y. Crops classification from Sentinel-2A multi-spectral remote sensing images based on convolutional neural networks[C]//ICARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, 2018: 5300–5303.
- [30] BANNARI A, MORIN D, BONN F, et al. A review of vegetation indices[J]. *Remote Sensing Reviews*, 1995, 13(1–2): 95–120.
- [31] WARDLOW B D, EGBERT S L. A comparison of MODIS 250-m EVI and NDVI data for crop map: a case study for southwest Kansas[J]. *International Journal of Remote Sensing*, 2010, 31(3): 805–830.
- [32] GITELSON A A, VINA A, ARKEBAUER T J, et al. Remote estimation of leaf area index and green leaf biomass in maize canopies[J]. *Geophysical Research Letters*, 2003, 30(5): 1248.
- [33] ANDLEEB T, ALI Z, MAHMOOD Z, et al. Wheat varietal response to relative SPAD index (RSI) and relative normalized difference vegetation index (RNDVI) under variable nitrogen application and terminal heat stress along with yield repercussion [J]. *Agronomy*, 2022, 12(7): 1538.
- [34] SUMNER Z, VARCO J, DHILLOM J, et al. Ground versus aerial canopy reflectance of corn: red-edge and non-red edge vegetation indices[J]. *Agronomy Journal*, 2021, 113(3): 2782–2797.
- [35] SIMS D A, GAMON J A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages[J]. *Remote Sensing of Environment*, 2002, 81(2–3): 337–354.
- [36] WU C, NIU Z, TANG Q, et al. Estimating chlorophyll content from hyperspectral vegetation indices: modeling and validation [J]. *Agricultural and Forest Meteorology*, 2008, 148(8–9): 1230–1241.
- [37] GITELSON A A, VINA A, CIGANDA V, et al. Remote estimation of canopy chlorophyll content in crops[J]. *Geophysical Research Letters*, 2005, 32(8): 8403.
- [38] FRAMPTON W J, DASH J, WATMOUGH G, et al. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013, 82: 83–92.
- [39] HASSAN M A, YANG M, RASHEED A, et al. Time-series multispectral indices from unmanned aerial vehicle imagery reveal senescence rate in bread wheat[J]. *Remote Sensing*, 2018, 10(6): 809.
- [40] PENUELAS J, BARET F, FILELLA I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance[J]. *Photosynthetica*, 1995, 31(2): 221–230.
- [41] PAN S J, YANG Q. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345–1359.
- [42] LONG M, CAO Y, WANG J, et al. Learning transferable features with deep adaptation networks [C] // International Conference on Machine Learning, 2015: 97–105.
- [43] CHOPRA M, CHHIPA P C, MENGI G, et al. Domain adaptable self-supervised representation learning on remote sensing satellite imagery[C]//2023 International Joint Conference on Neural Networks, 2023: 1–8.
- [44] XIE D, YANG R, QIAO Y, et al. Intelligent identification of landslide based on deep semi-supervised learning[C]//2022 5th International Conference on Pattern Recognition and Artificial Intelligence, 2022: 264–269.
- [45] NAVIN M S, AGILANDEESWARI L. Land use land cover change detection using K-means clustering and maximum likelihood classification method in the javadi hills, Tamil Nadu, India [J]. *International Journal of Engineering and Advanced Technology*, 2019, 9: 2249-8958.
- [46] GOYAL N, KUMAR N. On solving leaf classification using linear regression[J]. *Multimedia Tools and Applications*, 2021, 80: 4533–4551.
- [47] KATERJI N, VAN HOORN J W, HAMDY A, et al. Salinity effect on crop development and yield, analysis of salt tolerance according to several classification methods[J]. *Agricultural Water Management*, 2003, 62(1): 37–66.
- [48] BERNALES A M, ANTOLIHAI J A, SAMONTE C, et al. Modelling the relationship between land surface temperature and landscape patterns of land use land cover classification using multi linear regression models[J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, 41: 851–856.
- [49] AZRIEL D, BROWN L D, SKLAR M, et al. Semi-supervised linear regression[J]. *Journal of the American Statistical Association*, 2022, 117(540): 2238–2251.
- [50] YOGESH N. Mathematical approach to representation of locations using K-means clustering algorithm [J]. *International Journal of Mathematics and Its Applications*, 2021, 9(1): 127–136.
- [51] FOODY G M. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification[J]. *Remote Sens. Environ.*, 2020, 239: 111630.
- [52] LI Y. Research and application of deep learning in image recognition[C]//2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications, 2022: 994–999.
- [53] 新疆维吾尔自治区统计局. 新疆统计年鉴[M]. 北京: 中国统计出版社, 2019.
- [54] QIONG H U, WU W, QIAN S, et al. How do temporal and spectral features matter in crop classification in Heilongjiang Province, China? [J]. *Journal of Integrative Agriculture*, 2017, 16(2): 324–336.
- [55] REN T, LIU Z, ZHANG L, et al. Early identification of seed maize and common maize production fields using Sentinel-2 images[J]. *Remote Sensing*, 2020, 12(13): 2140.