



引文格式:郭浩然,张新,郑逸榛.基于样本迁移的无监督植被遥感制图方法[J].武汉大学学报(信息科学版),2025,50(9):1812-1824.DOI:10.13203/j.whugis20230242

Citation: GUO Haoran, ZHANG Xin, ZHENG Yizhen. Unsupervised Vegetation Remote Sensing Mapping Method Based on Sample Migration[J]. Geomatics and Information Science of Wuhan University, 2025, 50(9): 1812-1824. DOI: 10.13203/j.whugis20230242

基于样本迁移的无监督植被遥感制图方法

郭浩然^{1,2} 张新¹ 郑逸榛^{1,2}

1 中国科学院空天信息创新研究院遥感与数字地球全国重点实验室,北京,100101

2 中国科学院大学资源与环境学院,北京,100101

摘要:长时序植被分类与动态制图对于地球表层生态环境变化及规律认知具有重要意义。针对长时序植被分类与动态制图中存在的可靠样本稀缺、效率低、成本高的难题,提出一种基于历史植被分类图数据集自动获取可靠样本的无监督分类方法,可实现对大标签样本数据的局部-全局样本优化迁移与分类制图。首先将历史大标签数据的几何属性与多源遥感数据作为约束,实现大标签样本去噪以获得优化样本;然后通过长时序遥感影像变化检测实现可靠样本迁移;最后使用随机森林方法对植被进行分层分类,获得长时序植被动态制图结果。以内蒙古自治区阿鲁科尔沁旗长时序植被遥感制图为例,实现了2005—2022年多时相遥感分类精度均优于88%,验证了所提方法的科学性和实用性。

关键词:长时序遥感;样本优化;样本迁移;无监督分类;植被制图

中图分类号:P228

文献标识码:A

收稿日期:2024-08-12

DOI:10.13203/j.whugis20230242

文章编号:1671-8860(2025)09-1812-13

Unsupervised Vegetation Remote Sensing Mapping Method Based on Sample Migration

GUO Haoran^{1,2} ZHANG Xin¹ ZHENG Yizhen^{1,2}

1 State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

2 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100101, China

Abstract: Objectives: Vegetation classification and mapping are of great significance to develop ecological environmental protection. Supervised classification is the most widely used method for vegetation classification and mapping because it can generate accurate classification results. However, most of the current vegetation classification and mapping methods rely on single-phase field data, and field sampling often requires a lot of manpower and material resources. So it is difficult to realize long-time sequence dynamic vegetation classification and mapping only by field sample data. **Methods:** This paper proposes an unsupervised classification method for automatically obtaining reliable samples based on historical vegetation classification map datasets. Using the 1 : 1 000 000 Chinese Vegetation Atlas dataset as prior knowledge, a model for optimizing and migrating large-label samples for vegetation types is proposed. Multi-source data of the same or similar temporal phase as the existing vegetation classification maps are used to construct a feature set for local sample clustering optimization and global sample hierarchical Gaussian mixture optimization, which replaces the manual selection of training samples, and then obtains a usable sample set. On this basis, the invariant region sample migration is carried out by combining the results of long-time series Landsat-derived vegetation change analysis. The migrated samples are used for multi-temporal vegetation classification mapping, to quickly obtain the dynamic mapping results of long-time series vegetation classification. **Re-**

基金项目:内蒙古科技重大专项(2021ZD0045);国家重点研发计划(2021YFB3901301)。

第一作者:郭浩然,硕士,研究方向为植被生态遥感。guohaoran21@mails.ucas.ac.cn

通信作者:张新,博士,研究员。zhangxin@radi.ac.cn

sults: The experiment select Arhorqin Banner as the key study area, and complete the regional multi-temporal natural vegetation classification and mapping from 2005 to 2022, with the overall accuracy of the classification better than 88%, and the Kappa coefficient greater than 0.80. The natural vegetation in the study area is dominated by grasslands, forests and scrubs are mainly distributed in the northwestern part of the study area, with temperate tufted grassland dominating the grasslands and the temperate tufted grassland dominated by the extensive needlegrass grassland. From 2005 to 2022, natural vegetation shrinkage increases from north to south, while temperate deciduous broad-leaved forests show little change. The most pronounced changes occur in the southern part of the study area, where temperate graminoid grasslands and temperate deciduous scrub decline steadily, and non-natural vegetation expand year by year. Further analysis of the situation shows that the southern part of the study area is dominated by sandy vegetation, which has been seriously degraded since 2000. Still, the study area has started to implement the artificial forage industry since 2010, so the natural vegetation area has decreased, and the unnatural vegetation area has increased. The unnatural vegetation area is mostly small round patches, which are more concentrated in the southern part of the study area. **Conclusions:** The overall accuracy of the classification meets the needs of long-time series vegetation classification mapping, and the mapping effect is more stable. Therefore, the unsupervised sample migration method based on historical large-label vegetation classification maps can make full use of the existing vegetation classification products, and provide a more convenient way for vegetation classification mapping update. By using the geometric features of the historical vegetation classification products and combining with the multi-source remote sensing data of the same or similar time-phase, we construct the feature attribute set of the patch-by-patch, and hierarchically optimize samples of the large-label data of vegetation classification from local to global, to form the optimized and migratable vegetation classification maps. Optimization of samples is carried out to form a migratable training set of optimized vegetation classification samples. The optimized vegetation classification samples training set is then combined with land use data to carry out multi-temporal natural vegetation classification and mapping, and the results of long-time series natural vegetation classification changes are obtained. This study provides a fast, convenient, lightweight, and reliable mapping method for vegetation remote sensing classification.

Key words: long-time series remote sensing; sample optimization; sample migration; unsupervised classification; vegetation mapping

地球表面上生长的所有植物群落统称为植被,作为反映生态环境变化的重要依据,植被对地球生态系统和人类社会有着极其重要的作用^[1]。植被分类是通过植物物种的生长环境、物种组成和生长状况等特征,建立一套统一的分类准则,准确的植被分类为合理利用和有效保护植被提供科学的依据^[2]。因此,获取植被类别的分布和生长状况并进行植被分类制图,对开展生态环境保护工作有着极其重要的意义。自然界中,植被的类型复杂多样,传统植被分类制图方法主要依靠人工。随着新型遥感数据和分类算法的发展,植被遥感制图的精度和效率不断提高,为植被分类提供一种新方法,并因为其快速和高效被广泛应用^[3-6]。

监督分类因其可以生成准确的分类结果,是目前植被分类制图中应用最为广泛的方法。文献[7]制作土地覆盖分类和植被指数分布图,并

结合植被的分布特征对生态系统空间格局进行相关分析;文献[8]在土地利用/覆被和植被遥感监测等方面开创应用先河;文献[9]利用归一化植被指数(normalized difference vegetation index, NDVI)对1982—2015年哈萨克斯坦的植被信息(农田、草地和灌丛)进行了提取及变化特征分析;文献[10]结合实地采样数据,从遥感图像中提取不同类别草地的光谱、纹理和空间特征,根据一定的规则将图像中的每个像素划分为不同的类别;文献[11-13]使用长时序反映植被生长状况的植被指数,如归一化燃烧指数(normalized burn ratio, NBR)和反映图像同质性的纹理特征,结合原始多光谱图像及地形特征、气象特征等数据进行植被分类,成功改进了植被分类精度。

目前的植被遥感分类制图方法大多依赖于单一时相的实地数据,但实地采样往往需要耗费大量的人力物力,因此仅依靠实地样点数据实现

长时序动态植被分类制图较为困难^[14-15]。对于实现长时序植被分类与动态制图来说,获取较为可靠的历史样本是目前面临的难题。现有的植被分类样本数据集包括全球自然带植被类型分布图及2000年1:1 000 000中国植被图^[16]、2020年更新的1:1 000 000中国植被图^[17]等。尽管现存的植被分类大标签样本数据存在很多噪声,但对于实现长时序植被分类制图来说是丰富而珍贵的历史数据集。因此,文献^[18]提出了一种从历史土地覆盖数据集中获取可靠样本,并利用最近的多光谱图像的时间序列对土地覆盖进行更新的无监督的分类方法。通过充分利用历史植被分类数据集大标签样本,从中挖掘多时相的可利用的植被分类训练样本,在一定程度上能够为解决当前长时序动态快速植被分类制图费时费力的问题提供帮助。兼顾空间和时间分辨率的Landsat、Sentinel-2等系列卫星数据常被用于植被分类及动态监测,尤其是拥有最长连续全球覆盖影像档案(1973年至今)的Landsat系列卫星,为长期环境分析提供了信息丰富的多时相数据^[19]。

基于变化检测的样本迁移方法通过确定多个时相影像之间的不变区域来进行样本迁移,能够降低多时相之间样本传递过程中的噪声,从而具有较好的鲁棒性^[20-21]。基于Landsat时序数据的变化检测方法主要包括模型法、轨迹拟合法和光谱-时间轨迹法3类^[22-23],其中光谱-时间轨迹法通过地表变化过程中光谱-时间轨迹的特点来判定干扰事件是否发生,Landsat干扰和恢复趋势监测(LandTrendr)算法主要检测以年为步长的变化,能更全面地检测出渐变和突变事件,并且能够检测变化趋势和扰动事件^[24]。然而,在基于变化检测的样本迁移的源训练样本往往不够精准,如采用大标签的历史植被分类图数据集作为类别标签,缺乏降低源训练样本中噪声的方法将难以获得理想的长时序动态快速植被分类结果^[25-27]。

综上,针对长时序动态植被分类制图,缺乏一种更为便捷、快速、轻量且较为可靠的方法。直接对大标签植被分类图采用基于变化检测的样本迁移方法缺乏对大标签样本中存在噪声、不确定性的考虑。因此,本文提出一种针对长时序动态植被分类制图方法,基于现有的植被分类数据集,并综合土地覆盖产品,从大标签中自动获取高质量的训练样本集并与基于变化检测的多时相样本迁移方法相结合,从而实现多时相植被

分类制图。本文以内蒙古自治区阿鲁科尔沁旗为试验区,实现利用历史植被分类图进行长时序动态快速植被分类制图,为长时序植被分类分析提供支持。

1 研究区域与数据

1.1 研究区

阿鲁科尔沁旗位于内蒙古自治区中部,赤峰市东北部,地理坐标为119°02'E~121°01'E、43°21'N~45°24'N,研究区范围及地形如图1所示。研究区整体地势为西北向东南倾斜,最高海拔高度1 527 m,最低海拔高度231 m,相对高差1 296 m,平均海拔430 m。总面积为14 277 km²,属于中温带半干旱大陆性季风气候区^[28]。

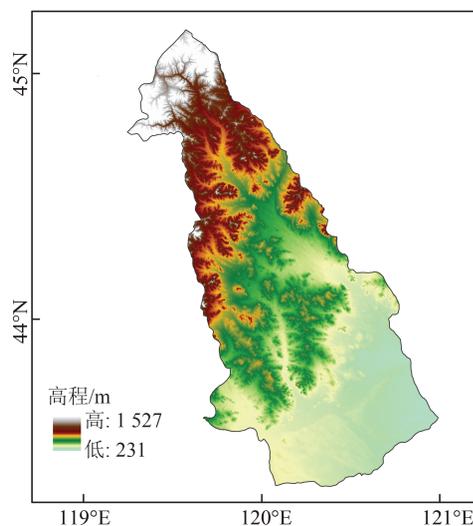


图1 研究区范围及地形

Fig. 1 Scope and Topography of the Study Area

1.2 研究数据

研究选用的植被分类样本数据来自2000年《1:1 000 000中国植被图集》数据集^[16];Landsat数据来自Landsat Collection 2数据集,使用经去云裁剪等预处理的2000年、2005年、2010年、2015年、2020年、2022年Level-2数据;气象数据来自ERA5-Land数据集,在本研究中选取上述年份的一级土壤温度、植被冠层或土壤上薄层含水量、地球表面反射率、地表径流、地表温度5个变量;MODIS数据来自MOD13Q1数据集,本文使用的是2005年、2010年、2015年、2020年、2022年的增强型植被指数(enhanced vegetation index, EVI)时序数据;数字高程模型(digital elevation model, DEM)空间分辨率为30 m,根据研究区范围进行裁剪,并生成海拔、坡度、坡向3个变量。

此外,为消除非自然植被区对分类结果的影响,使用 2005 年、2010 年、2015 年、2020 年中国逐年土地覆盖数据集 (China Land Cover Dataset, CLCD)^[29],对研究区相应年份的非自然植被区域进行掩膜;使用 AI Earth 平台提供的 2022 年中国 10 m 地物分类数据^[30],对研究区 2022 年非自然植被区域进行掩膜。

2 研究方法

2.1 方法原理与流程

制作植被分类图往往需要耗费大量的人力和物力,现有的植被分类图是可利用的存在噪声的历史大标签样本数据。结合相近时相的植被生长期多源数据构建特征集合,从而实现对历史大标签样本数据进行由局部到全局去噪,生成用于植被分类的可靠样本训练集。该框架可在一

定程度上减少或者避免手动标记新样本,提高样本的可重复利用率及分类精度,本文的技术路线如图 2 所示,其中 S-G 滤波表示 Savitzky-Golay 滤波。

植被分类图数据集大标签样本是一种有效的几何约束,可以作为优化原始植被分类样本的一种先验局部空间约束的最小单元。以研究区的《1:1 000 000 中国植被图集》数据集为先验知识,提出一种针对植被类型大标签样本优化与迁移的模型。利用与现有植被分类图同一或相近时相的多源数据构建特征集合进行局部样本聚类优化、全局样本分层高斯混合优化,替代手动选择训练样本,进而获得可利用的样本集。在此基础上,结合长时序 Landsat 植被变化检测结果进行不变区域样本迁移。将迁移出的样本用于多时相植被分类制图,从而快速获取长时序植被分类动态制图结果。

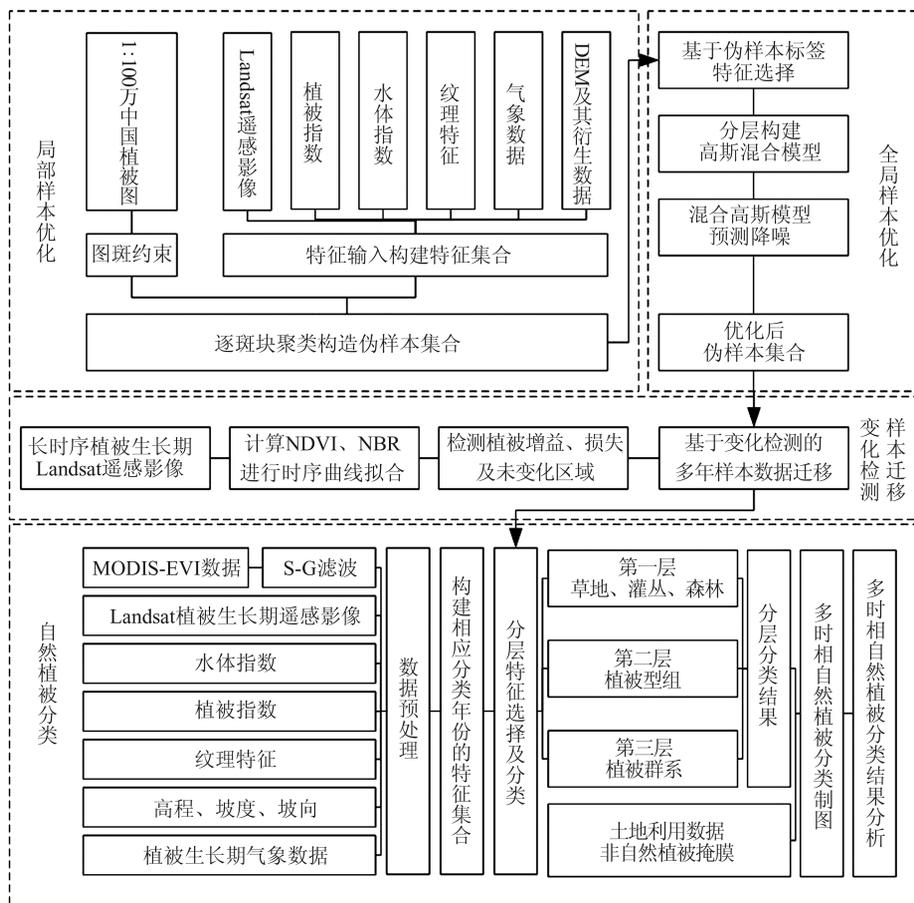


图 2 无监督样本迁移植被遥感制图技术路线

Fig. 2 Unsupervised Sample Migration Workflow for Vegetation Remote Sensing Mapping

该方法主要包括以下几个步骤:

1) 以《1:1 000 000 中国植被图集》数据集的图斑作为最小几何约束的局部单元,结合植被生长期的光谱属性数据、DEM 数据、气象数据等进

行局部单元内聚类分析,占比最多的簇继承标签,有效降低植被分类制图产品中的不确定性。

2) 结合类别信息全局逐层优化样本标签,考虑不同植被类型在特定特征空间的分布,分

层将不同的植被类型分别构建高斯混合模型(Gaussian mixture model, GMM),通过求解GMM的过程去除错分斑块以及步骤1)带来的错误标签。

3)采用长时序植被生长期Landsat数据进行变化检测分析,确定植被增益、损失、不变区域,结合不变区域迁移样本,构建多时相训练样本集。

4)通过输入的多时相训练样本集及多源遥感数据进行特征筛选,并采用随机森林方法(random forest, RF)分层分类,根据土地利用数据进行掩膜,从而获得多时相的自然植被分类制图。

2.2 局部到全局分层样本迁移与优化

大标签分类结果中存在较大的噪声,因此首先在局部对样本进行迁移优化。以《1:1 000 000中国植被图集》数据集中植被群系图斑的多边形信息作为几何约束,为有效约束斑块内的错误植被分类信息,本文处理研究区与大标签样本数据集相近时相的Landsat植被生长期遥感影像,将影像的光谱特征、植被指数、水体指数、纹理特征、气象数据、DEM及其衍生数据作为特征输入,构建相应图斑的特征集合,作为图斑聚类特征的输入。

将K-means方法作为最基础的聚类方法,对第 j 个植被类型图斑的逐斑块特征数据集实施聚类分析。以第 i 个斑块 v_i 为例,其由 N 个像元组成,对应的属性特征集为 Y_{v_i} 。假设第 j 个植被类型图斑对应的类别标签为 X_j ,为了从中分离出与 X_j 正确关联的子集,K-means方法通过最小化平方误差(sum of squared errors, SSE)完成对簇的聚类划分,计算式为:

$$e_{SSE} = \sum_{k=1}^{K_j} \sum_{y_n \in C_j^k} \|y_n - \mu_k\|_2^2 \quad (1)$$

式中, e_{SSE} 为平方误差; K_j 为预期划分的簇数; y_n 为第 n 个特征向量; C_j^k 为第 k 个簇; μ_k 为 C_j^k 的均值向量。

采用手肘法确定最优聚类簇数,即SSE随着 K_j 的增大而变小。当 K_j 小于真实聚类数时,随着 K_j 值的增大,SSE的下降幅度很大;当 K_j 到达真实聚类数后,随着 K_j 值的增大,SSE的下降幅度骤减,最后趋于平缓。SSE和 K_j 的关系图是一个手肘的形状,而这个肘部对应的 K_j 值就是样本的最优聚类数。

将每个局部单元内的像素集合在几何与多源属性的约束下按最优 K_j 划分,占比最多的簇保

留并继承原始植被分类产品类别标签。但现有植被分类产品不能保证每个图斑内都正确分类,因此当前得到的植被分类伪样本集 S_{pseudo} 中存在一定数量的错误,需要进一步优化迁移。局部聚类分析的总时间复杂度为 $O(K_j Nd) \cdot T$,其中 d 为特征维度, T 为迭代次数。优化方法如图3所示。在图斑内将占比最多的簇保留并继承植被分类标签,得到伪样本。

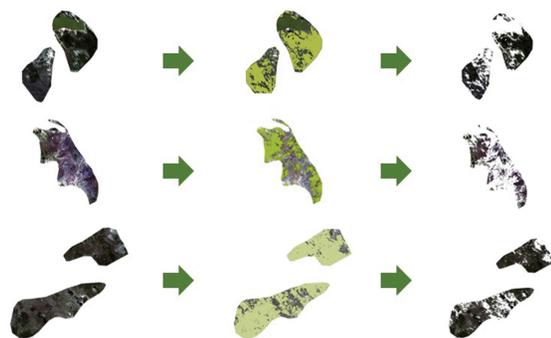


图3 局部聚类优化

Fig. 3 Local Optimization

为了将噪声从伪样本集 S_{pseudo} 中剔除,获得优化后的训练样本集 $S(S \subseteq S_{pseudo})$,本文采用一种基于高斯混合模型的植被分类特征优选的全局分层样本优化方法。从全局多源特征出发,结合植被在不同地域分异中的尺度效应,采用RF方法对特征集合 Y_{v_i} 进行优选,将优选特征集合作为关键特征输入模型,完成对植被分层分类类别的全局优化。采用统计方法对伪样本集中的错误样本进行剔除,分两层对 S_{pseudo} 构建高斯混合模型;将高斯混合模型样本结果与原始标签进行对比,保留正确样本,完成全局样本优化,获得优化后的训练样本集。

对 S_{pseudo} 按照森林、灌丛、草地以及植被型两层分别对应植被类别标签 X_i 进行划分,在不同的分类体系下对伪样本集进行逐类别的伪样本集构建,即 $S_{pseudo} = \{S_{X_1}, S_{X_2}, \dots, S_{X_i}, \dots, S_{X_I}\}$,其中 S_{X_i} 为类别 X_i 对应的伪样本集,其高斯混合计算式为:

$$p_M(y) = \alpha_1 \cdot p(y|u_1, \Sigma_1) + \alpha_2 \cdot p(y|u_2, \Sigma_2) + \dots + \alpha_i \cdot p(y|u_i, \Sigma_i) + \dots + \alpha_I \cdot p(y|u_I, \Sigma_I) \quad (2)$$

式中, $p_M()$ 为概率密度函数; α_i, u_i, Σ_i 分别为第 i 个高斯分布的混合系数、均值向量与协方差向量; I 为类别总数。

计算高斯混合模型的方法采用迭代期望最大化(expectation-maximum, EM)算法,假设训练集 $S_{X_i} = \{s_1^i, s_2^i, \dots, s_{N_i}^i\}$ 由 N_i 个样本组成,令 GM_i

代表生成的单高斯模型,则对于第 n_i ($n_i \leq N_i$) 个样本 $s_{n_i}^x$ 而言,有隶属于 GM_j 的后验概率 $\gamma_{(n_i,j)}$ 。EM 算法计算高斯混合模型分为两个部分,首先执行 E 步骤,通过当前的混合成分参数 $\{\alpha_j, \mu_j, \Sigma_j\}$,计算 $s_{n_i}^x$ 属于 GM_i 的后验概率 $\gamma_{(n_i,j)}$;然后将 E 步骤的后验概率作为权重,采用最大似然法估算新的混合参数 α_j 、均值向量 μ_j 以及协方差向量 Σ_j ,完成算法的 M 步骤。通过 E 步骤和 M 步骤的不断迭代更新,直到 M 步骤的似然函数的变化率小于设定阈值,即可得到优化后的样本集。全局高斯混合模型分析的总时间复杂度为 $O(KNd)$,其中 K 为类别标签数。

2.3 多时相植被变化检测与样本迁移

利用 GEE (Google Earth Engine) 中 2000—2022 年的 Landsat TM/ETM/OLI 的遥感影像数据构建时序数据集^[31]。选取每年 6 月 10 日至 8 月 30 日的 Landsat 影像对应的绿 (Green)、红 (Red)、

近红外 (near infrared, NIR) 和短波红外波段 (SWIR1、SWIR2) 5 个波段,构建 22 年逐年最小云量合成图,计算 NDVI 和 NBR,计算式分别为:

$$I_{NDVI} = \frac{R_{NIR} - R_{Red}}{R_{NIR} + R_{Red}} \quad (3)$$

$$I_{NBR} = \frac{R_{NIR} - R_{SWIR2}}{R_{NIR} + R_{SWIR2}} \quad (4)$$

式中, R_{Red} 、 R_{NIR} 和 R_{SWIR2} 分别表示红光波段、近红外和短波红外第二波段的表层反射率。

本文采用 LandTrendr 算法检测研究区内植被的变化,并划分植被变化增益、变化损失以及不变区域。LandTrendr 是一种时间分割算法,能够捕捉研究区 22 年间剧烈的短期变化或者渐进的变化,其核心为构建研究区每个像元的时序谱线,从而监测其变化。每个像素的输入是一个光谱指数或者波段的时间序列^[24]。通过实验分析,设定相关参数来保证研究区变化检测的质量,输入参数如表 1 所示。

表 1 LandTrendr 算法参数

Table 1 Parameters of LandTrendr Algorithm

参数名称	参数描述	参数值
MaxSegments	分割单元的最大数目	4
SpikeThreshold	如果相邻两个时间点值的差异百分比小于该值,那么该值会被认为是异常值,须剔除	0.9
VertexCountOvershoot	在初始阶段的潜在节点回归中可以超过的节点数	3
PreventOneYearRecovery	阻止一年后恢复的情况否	TRUE
RecoveryThreshold	如果某个分割段的恢复率大于该值的倒数,那么这个分割段将会被移除	0.25
PvalThreshold	回归分析中 F 检验的 P 值,超过该值,则认为该像元没有发生变化	0.05
BestModelProportion	简单模型的选择规则,如果超过该值,则被选中	0.75
MinObservationsNeeded	拟合中需要的最少观测数	6

2.4 自然植被分层分类

RF 分类模型是一种常用的机器学习分类算法,其通过构建多个决策树来运行^[32]。与其他机器学习算法相比,RF 算法具有计算精度高、模型训练时间少、能够确定变量在模型中的重要性等优势,因此被广泛应用于分类研究中^[33-34]。在 GEE 中调用 RF 分类模型,该模型参数较少,决策树数量和变量数量是影响其结果的参数。研究采用的植被分类方法对每一层进行分类^[35-36]。分层分类结构如图 4 所示,共 3 层,其中第一层为森林、灌丛、草地分类,第二层为植被型分类,第三层为植被群系分类。在计算可以负荷的情况下,针对每一层对特征重要性进行排序,并进行特征优选,选择精度最高时的决策树数量。研究采用的 RF 决策树数量第 1 层是 200,第 2 层是 150,第 3 层是 200,且变量数量默认为输入的特征变量总

数的平方根。

本文使用混淆矩阵对分类精度进行评价,包括总体精度 (overall accuracy, OA) 和 Kappa 系数。OA 是分类结果正确分类像元数与总像元数的比值,常用于快速评估分类结果的整体精度;Kappa 系数是一种用于统计分类结果准确性和随机性的指标。

3 数据结果处理与分析

针对大标签中国植被分类图样本,以 2000 年 8 月 Landsat 遥感影像为主,计算植被指数、土壤指数、纹理特征,并结合光谱特征、气象数据、DEM 及其衍生数据构建特征集合,采用局部-全局的大标签样本降噪方法构建优化后的样本集合,总计 3 563 个样本,样本空间分辨率为 30 m。

实验中,基于 Python 3.7.3 的局部-全局大标

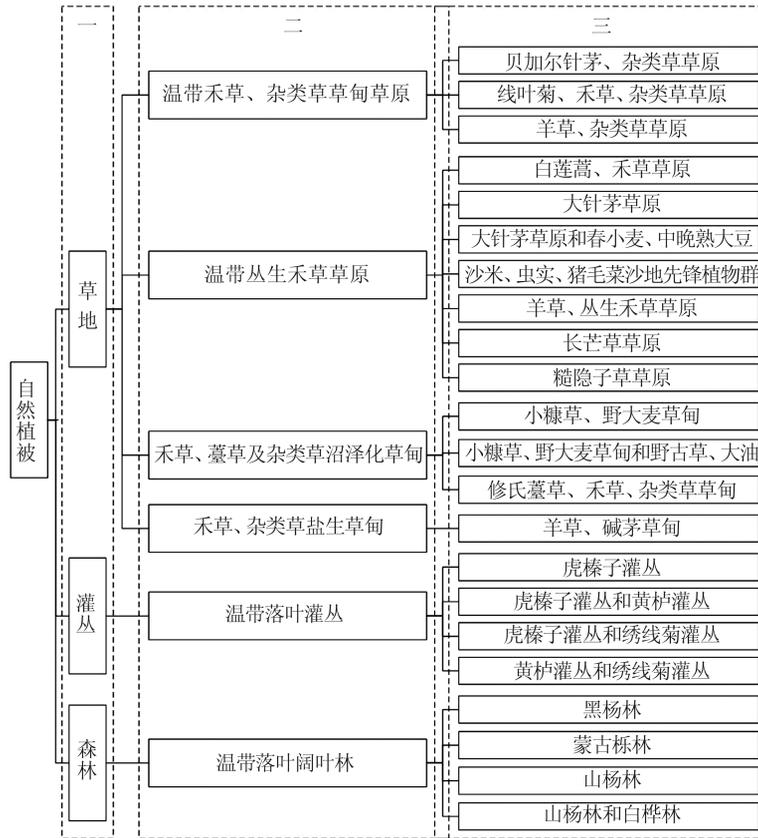


图4 研究区植被分层分类体系

Fig. 4 Hierarchical Classification System of Vegetation in the Study Area

签样本优化方法的局部聚类优化耗时 386.46 s, 全局高斯混合优化耗时 32 656.67 s。同时进行了手动选择样本的对比试验,与手动选择样本相比,局部-全局大标签样本优化方法的工作量更小,速度更快,能更好地结合多源特征进行数据筛选,具有相对快速的优势。

根据 2000—2022 年 Landsat 每年 6 月 10 日到 8 月 30 日的影像数据进行研究区植被变化检测,确定了 2005 年、2010 年、2015 年、2020 年、2022 年植被增益、损失、不变的区域,空间分辨率为 30 m,变化检测结果如图 5 所示。

通过设置去除噪声引起的尖峰识别潜在的断点,返回分段减少与分段增长,同时拟合轨迹和最佳分段数,寻找最佳模型,得到相应年份的变化检测结果。对变化检测结果进一步分析可得,2000—2022 年研究区的变化区域逐年增加,其中损失区域在 2015 年之前逐年上升,在 2015 年呈现峰值,在 2015 年之后有所下降,损失区域在研究区北部分布较多;增益区域在整体上也呈现上升趋势,但 2010 年的增益区域在变化检测的年份中处于最低水平,增益区域在研究区南部分布较多;不变区域逐年减少,但整体分布较为均

匀。结合相关文献、土地利用数据及实际情况进一步分析,研究区 2000 年后存在自然植被退化问题,但北部的自然植被退化问题自 2015 年有所好转;2010 年起,研究区大力发展人工饲草产业,主要集中在南部地区,因此南部植被增益像元数明显增加,多为人工饲草种植。

根据植被不变区域的结果,构建多时相遥感影像与初始影像之间的关系后,将不变区域的样本进行传递,分别得到 2005 年、2010 年、2015 年、2020 年、2022 年 5 个年份的样本,以植被型标签为例,经变化检测迁移的样本统计如表 2 所示。

针对上述年份,采用经 S-G 滤波后的 MODIS-EVI 时序数据,数据来自 MOD13Q1 数据集,时间分辨率为 16 d,因此全年可用数据共计 25 个时相,EVI 时序曲线如图 6 所示。其中图 6(a)~6(c)中的 EVI 和 EVI-S-G 分别表示 S-G 滤波前和滤波后的曲线,图 6(d)为 3 种植被的 EVI 时序曲线,图 6(e)为以植被型作为分类标签的 EVI 时序曲线。由图 6 可见,3 种植被的时序 EVI 曲线表现出不同的幅度特征,且森林最大、草地最小。该现象可归因于各类植被在冠层结构、叶面积及物候强度方面的差异,从而影响 EVI 的振幅。

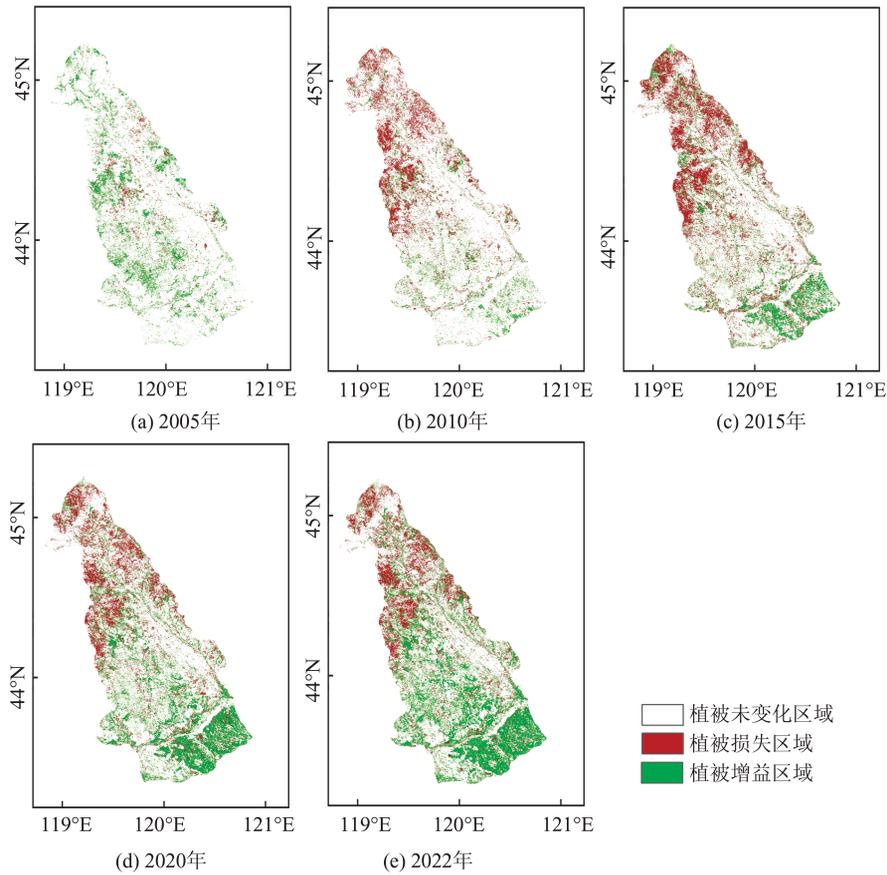


图 5 2000—2022 年研究区变化检测结果

Fig. 5 Change Detection Results in the Study Area from 2000 to 2022

表 2 各时相迁移后的样本数

Table 2 Samples After Migration by Time Phase

年份	温带落叶阔叶林	温带落叶灌丛	温带禾草、杂类草草甸草原	温带丛生禾草草原	禾草、藁草及杂类草沼泽化草甸	禾草、杂类草盐生草甸	总计
2005年	170	806	73	1 573	205	49	2 876
2010年	116	628	79	1 601	197	48	2 669
2015年	98	401	57	1 455	165	42	2 218
2020年	126	341	59	1 344	162	33	2 065
2022年	129	310	55	1 224	145	29	1 892

使用时间序列谐波分析,将EVI时序曲线的正弦系数、余弦系数、相位角、振幅作为关键特征输入^[37-38];采用上述年份植被生长期 Landsat 数据,并由 Landsat 数据计算出植被指数、水体指数以及纹理特征作为分类特征输入;高程值、坡度值、坡向值被视为分类特征。此外,本文也考虑了采用植被生长期的气象数据作为分类特征输入。特征变量集如表 3 所示。

采用 RF 作为分类器,将上述 5 个年份的样本输入分类器,根据特征重要性进行特征选择,如图 7 所示。由图 7 可以看出,气象特征、地形特征、EVI 时序特征及植被指数在分类中起到重要作用;结合分类模型特征重要性排序进行特征优

选,得到上述 5 个年份的自然植被分类结果,并根据对应年份的土地利用数据对非自然植被区域进行掩膜,得到最终分类结果,空间分辨率为 30 m,如图 8 所示。

从整体上来看,研究区自然植被以草地为主,森林和灌丛主要分布在研究区西北部,其中草地以温带丛生禾草草原为主,温带丛生禾草草原以大针茅草原为主。2005—2022 年,自然植被缩减程度由北向南逐渐增加,其中温带落叶阔叶林面积没有发生显著的变化,自然植被变化最显著的区域位于研究区南部,温带禾草丛生草原、温带落叶灌丛(以小叶锦鸡儿灌丛为主)逐年减少,非自然植被区域逐年增加。结合实际情况

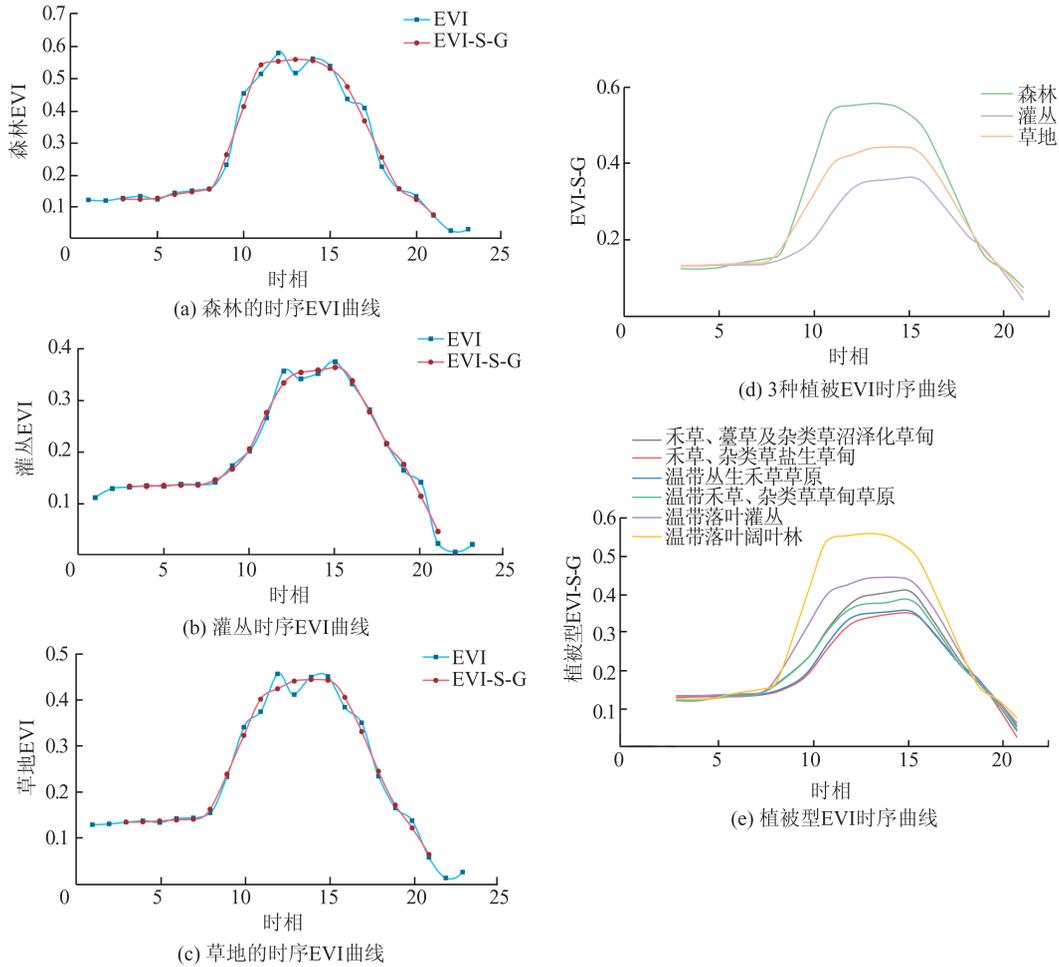


图6 EVI时序曲线

Fig. 6 Time Series Curves of EVI

表3 特征变量集

Table 3 Feature Variable Sets

特征集	特征变量
光谱特征	Blue, Green, Red, NIR, SWIR1, SWIR2
植被指数 ^[39]	NDVI, EVI, 差值环境植被指数(difference vegetation index, DVI)、比值植被指数(ratio vegetation index, RVI)、土壤调节植被指数(soil-adjusted vegetation index, SAVI)
水体指数 ^[40]	水体土壤湿度指数(land surface water index, LSWI)
纹理特征	角二阶矩、对比度、相关性、差分熵、差异、差异方差、熵、逆差矩、corr信息量、惯量、集群突出、求和平均值、总熵、集群阴影、求和方差、方差
地形特征	高程、坡度、坡向
MODIS_EVI 时序特征	正弦系数、余弦系数、相位角、振幅
气象特征	地表径流、植被冠层或土壤上薄层含水量、地表温度、一级土壤温度、地球表面反射率

进一步分析可知,研究区南部以沙地植被为主,

2000年后植被退化严重,但自2010年起研究区开始推行人工饲草产业,因此南部地区自然植被区域减少,非自然植被区域增加,非自然植被区域多为圆形小斑块,在南部较为集中分布。

结合2020年更新的1:1 000 000中国植被图^[17]及2020年CLCD土地利用数据^[29],选取5种自然植被群系类别对研究区2020年植被分类图进行验证,分别是:(1)蒙古栎林,(2)小叶锦鸡儿灌丛,(3)小糠草、野大麦草甸,(4)羊草、碱茅草甸,(5)沙米、虫实、猪毛菜沙地先锋植物群。上述5种自然植被群系类别的整体精度分别为93.08%、96.43%、82.17%、90.79%、86.37%。验证结果表明,本文实验的分类结果较为可靠。

2005—2022年研究区自然植被分类精度结果如表4所示。整体分类精度在88%以上,表明整体分类结果较好,本文所提出的框架能够实现较为可靠的样本迁移,对于实现长时序动态植被分类制图具有巨大应用潜力。

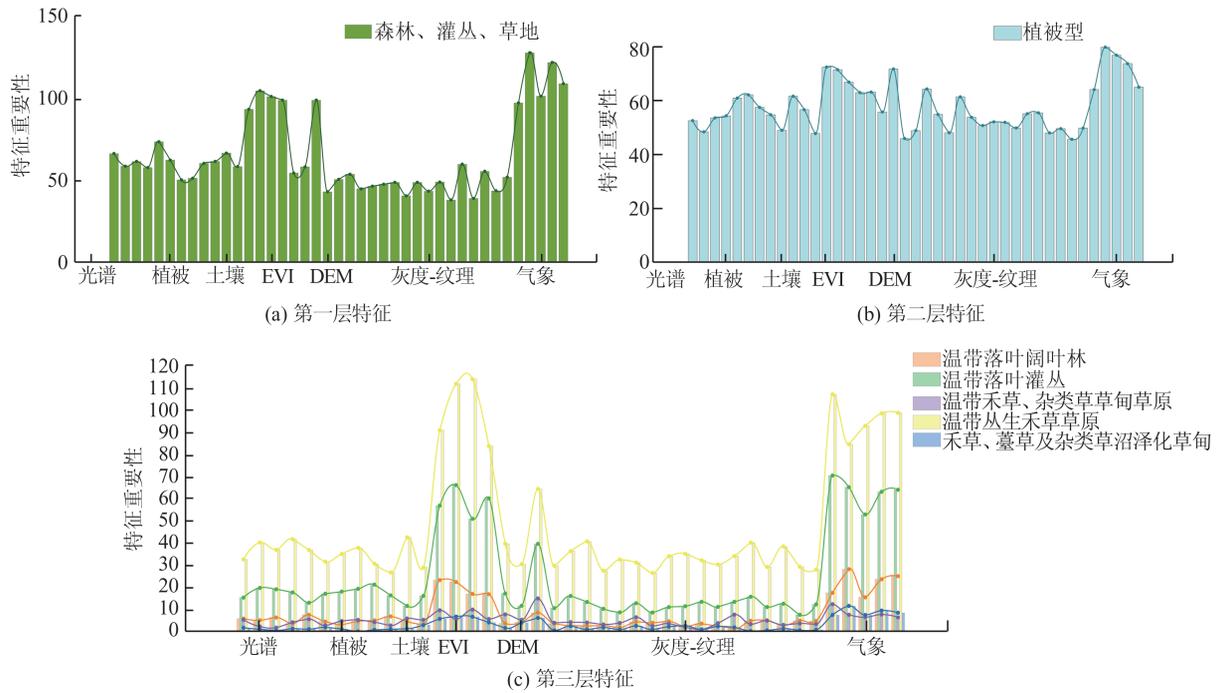


图 7 特征重要性

Fig. 7 Feature Importance

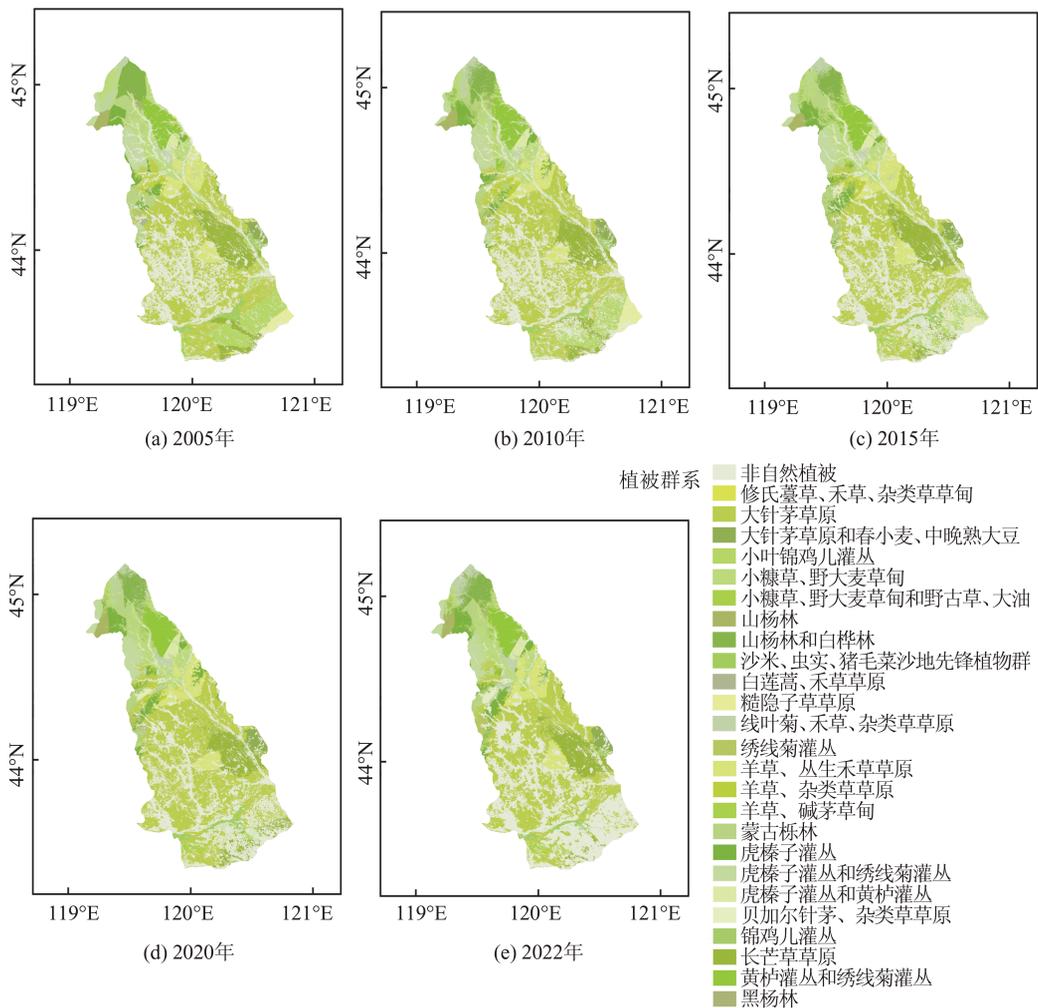


图 8 分层分类结果

Fig. 8 Hierarchical Classification Results

表4 阿鲁科尔沁旗2000–2022年自然植被分类结果分类精度

Table 4 Classification Accuracy of Natural Vegetation Classification Results in Arukorqin Banner from 2000 to 2022

层次	分类	精度	2005年	2010年	2015年	2020年	2022年
第一层	森林、灌丛、草地	OA/%	94.26	95.51	95.62	94.47	95.22
	大类分类	Kappa系数	0.878 6	0.894 0	0.887 4	0.847 9	0.850 9
第二层	草地植被型分类	OA/%	90.36	89.21	91.30	89.61	90.71
		Kappa系数	0.801 6	0.821 9	0.801 8	0.845 1	0.815 0
第三层	温带落叶阔叶林	OA/%	97.39	96.23	95.76	95.83	95.57
	群系分类	Kappa系数	0.954 8	0.943 2	0.951 3	0.935 0	0.934 6
	温带落叶灌丛	OA/%	97.60	96.95	97.01	94.20	91.94
	群系分类	Kappa系数	0.967 8	0.957 6	0.960 7	0.920 8	0.893 6
	温带禾草、杂类草	OA/%	90.91	88.21	88.89	92.31	89.33
	草甸草原群系分类	Kappa系数	0.830 8	0.819 2	0.869 2	0.864 6	0.836 4
	温带丛生禾草	OA/%	96.36	94.58	96.03	94.42	94.26
	草原群系分类	Kappa系数	0.925 7	0.895 8	0.922 1	0.885 3	0.887 2
	禾草、藁草及杂类草	OA/%	96.77	97.13	95.24	96.47	96.34
	沼泽化草甸群系分类	Kappa系数	0.842 2	0.866 5	0.857 1	0.844 6	0.854 0

4 结 语

本文提出了一种基于历史大标签植被分类数据及多源遥感数据的“局部-全局”几何及特征集合约束下的无监督的样本迁移模型及技术框架。实验选取内蒙古自治区阿鲁科尔沁旗为重点研究区,完成了2005—2022年的区域多时相自然植被分类制图,分类总体精度均优于88%,Kappa系数大于0.80,满足长时序植被分类制图的需求,制图效果较为稳定。因此,基于历史大标签植被分类图的无监督样本迁移方法在一定程度上能充分利用已有的植被分类产品,为植被分类制图更新提供了一种更便捷的途径,利用历史植被分类产品的几何特征,结合同一或相近时相的多源遥感数据构建逐图斑的特征属性集,从局部到全局分层对植被分类大标签数据进行样本的优化,从而形成可迁移的优化后的植被分类样本训练集。通过优选特征分层分类的方法结合土地利用数据进行多时相的自然植被分类制图,获得长时序自然植被分类变化结果。本研究为植被遥感分类提供了一种快速、便捷、轻量、较为可靠的快速制图方法。

本研究还存在不足,例如在长时序变化检测样本传递的过程中样本的噪声会增加,后续可结合植被相关知识进一步优化,在样本传递过程中增加适当的实地观测或其他有效观测手段来减少传递过程中产生的噪声,提高快速制图方法的准确性,为长时序植被分类快速动态制图提供一种更为可靠的方法。

参 考 文 献

- [1] DONG S K, SHANG Z H, GAO J X, et al. Enhancing Sustainability of Grassland Ecosystems Through Ecological Restoration and Grazing Management in an Era of Climate Change on Qinghai-Tibetan Plateau[J]. *Agriculture, Ecosystems & Environment*, 2020, 287: 106684.
- [2] 郭庆华, 关宏灿, 胡天宇, 等. 新一代1:50万中国植被图绘制方法探讨[J]. *中国科学: 生命科学*, 2021, 51(3): 229-241.
GUO Qinghua, GUAN Hongcan, HU Tianyu, et al. Remote Sensing-Based Mapping for the New Generation of Vegetation Map of China(1: 500 000) [J]. *Scientia Sinica (Vita)*, 2021, 51(3): 229-241.
- [3] ROUSE J W, HAAS R H, SCHELL J A, et al. Monitoring Vegetation Systems in the Great Plains with ERTS [C]//The 3rd Earth Resources Technology Satellite-1 Symposium, Trento, Italy, 1974.
- [4] XIE Y, SHA Z, YU M. Remote Sensing Imagery in Vegetation Mapping: A Review [J]. *Journal of Plant Ecology*, 2008, 1(1): 9-23.
- [5] DE CÁCERES M, CHYTRÝ M, AGRILLO E, et al. A Comparative Framework for Broad-Scale Plot-Based Vegetation Classification [J]. *Applied Vegetation Science*, 2015, 18(4): 543-560.
- [6] 牛全福, 刘明志, 张曼, 等. 近20年来新疆阿勒泰地区植被动态变化及其对气候与地形的响应[J]. *武汉大学学报(信息科学版)*, 2023, 48(9): 1522-1530.
NIU Quanfu, LIU Mingzhi, ZHANG Man, et al. Vegetation Dynamic Change and Its Response to Cli-

- mate and Topography in Altay Region of Xinjiang in Recent 20 Years [J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(9): 1522-1530.
- [7] LOBO A, MOLONEY K, CHIC O, et al. Analysis of Fine-Scale Spatial Pattern of a Grassland from Remotely-Sensed Imagery and Field Collected Data [J]. *Landscape Ecology*, 1998, 13(2): 111-131.
- [8] 徐新良, 刘纪远, 庄大方. 国家尺度土地利用/覆被变化遥感监测方法[J]. *安徽农业科学*, 2012, 40(4): 2365-2369.
XU Xinliang, LIU Jiyuan, ZHUANG Dafang. Remote Sensing Monitoring Methods of Land Use/Cover Change in National Scale[J]. *Journal of Anhui Agricultural Sciences*, 2012, 40(4): 2365-2369.
- [9] LUO L, DU W P, YAN H M, et al. Spatio-temporal Patterns of Vegetation Change in Kazakhstan from 1982 to 2015 [J]. *Journal of Resources and Ecology*, 2017, 8(4): 378-384.
- [10] PITKÄNEN T P, KÄYHKÖ N. Reducing Classification Error of Grassland Overgrowth by Combining Low-Density LiDAR Acquisitions and Optical Remote Sensing Data [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017, 130: 150-161.
- [11] HONG G, ZHANG A N, ZHOU F Q, et al. Integration of Optical and Synthetic Aperture Radar (SAR) Images to Differentiate Grassland and Alfalfa in Prairie Area [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 28: 12-19.
- [12] 杨超, 邬国锋, 李清泉, 等. 植被遥感分类方法研究进展[J]. *地理与地理信息科学*, 2018, 34(4): 24-32.
YANG Chao, WU Guofeng, LI Qingquan, et al. Research Progress on Remote Sensing Classification of Vegetation [J]. *Geography and Geo-Information Science*, 2018, 34(4): 24-32.
- [13] 张晓羽, 李凤日, 甄贞, 等. 基于随机森林模型的陆地卫星-8 遥感影像森林植被分类[J]. *东北林业大学学报*, 2016, 44(6): 53-57.
ZHANG Xiaoyu, LI Fengri, ZHEN Zhen, et al. Forest Vegetation Classification of Landsat8 Remote Sensing Image Based on Random Forests Model [J]. *Journal of Northeast Forestry University*, 2016, 44(6): 53-57.
- [14] YU W J, ZHOU W Q, QIAN Y G, et al. A New Approach for Land Cover Classification and Change Analysis: Integrating Backdating and an Object-Based Method [J]. *Remote Sensing of Environment*, 2016, 177: 37-47.
- [15] WU Z J, ZHANG J H, DENG F, et al. Super-pixel-Based Regional-Scale Grassland Community Classification Using Genetic Programming with Sentinel-1 SAR and Sentinel-2 Multispectral Images [J]. *Remote Sensing*, 2021, 13(20): 4067.
- [16] ENQUIST B J, NIKLAS K J. Global Allocation Rules for Patterns of Biomass Partitioning in Seed Plants [J]. *Science*, 2002, 295(5559): 1517-1520.
- [17] SU Y J, GUO Q H, HU T Y, et al. An Updated Vegetation Map of China (1: 1 000 000) [J]. *Science Bulletin*, 2020, 65(13): 1125-1136.
- [18] PARIS C, BRUZZONE L, FERNÁNDEZ-PRIETO D. A Novel Approach to the Unsupervised Update of Land-Cover Maps by Classification of Time Series of Multispectral Images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(7): 4259-4277.
- [19] 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.
- [20] 李贺, 何志杰, 黄翀, 等. 2000—2019年缅甸南部橡胶林时空演变[J]. *资源科学*, 2021, 43(12): 2403-2415.
LI He, HE Zhijie, HUANG Chong, et al. Spatio-temporal Evolution of Rubber Forests in Southern Myanmar During 2000—2019 [J]. *Resources Science*, 2021, 43(12): 2403-2415.
- [21] 张立福, 王飒, 刘华亮, 等. 从光谱到时谱: 遥感时间序列变化检测研究进展 [J]. *武汉大学学报(信息科学版)*, 2021, 46(4): 451-468.
ZHANG Lifu, WANG Sa, LIU Hualiang, et al. From Spectrum to Spectrotemporal: Research on Time Series Change Detection of Remote Sensing [J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(4): 451-468.
- [22] 汤冬梅, 樊辉, 张瑶. Landsat 时序变化检测综述 [J]. *地球信息科学学报*, 2017, 19(8): 1069-1079.
TANG Dongmei, FAN Hui, ZHANG Yao. Review on Landsat Time Series Change Detection Methods [J]. *Journal of Geo-Information Science*, 2017, 19(8): 1069-1079.
- [23] 眭海刚, 冯文卿, 李文卓, 等. 多时相遥感影像变化检测方法综述 [J]. *武汉大学学报(信息科学版)*, 2018, 43(12): 1885-1898.
SUI Haigang, FENG Wenqing, LI Wenzhuo, et al. Review of Change Detection Methods for Multi-

- Temporal Remote Sensing Imagery [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 1885-1898.
- [24] KENNEDY R E, YANG Z Q, COHEN W B. Detecting Trends in Forest Disturbance and Recovery Using Yearly Landsat Time Series; 1. LandTrendr—Temporal Segmentation Algorithms [J]. *Remote Sensing of Environment*, 2010, 114 (12) : 2897-2910.
- [25] 杜培军, 林聪, 陈宇, 等. 多时相遥感影像样本迁移模型与地表覆盖智能分类[J]. 同济大学学报(自然科学版), 2022, 50(7): 955-966.
DU Peijun, LIN Cong, CHEN Yu, et al. Training Sample Transfer Learning from Multi-temporal Remote Sensing Images for Dynamic and Intelligent Land Cover Classification [J]. *Journal of Tongji University (Natural Science)*, 2022, 50 (7) : 955-966.
- [26] 黄亚博, 廖顺宝. 多源数据的土地覆被样本自动提取[J]. 遥感学报, 2017, 21(5): 757-766.
HUANG Yabo, LIAO Shunbao. Automatic Collection for Land Cover Classification Based on Multi-source Datasets [J]. *Journal of Remote Sensing*, 2017, 21(5): 757-766.
- [27] 鲍蕊, 薛朝辉, 张像源, 等. 综合聚类 and 上下文特征的高光谱影像分类[J]. 武汉大学学报(信息科学版), 2017, 42(7): 890-896.
BAO Rui, XUE Zhaohui, ZHANG Xiangyuan, et al. Classification Merged with Clustering and Context for Hyperspectral Imagery [J]. *Geomatics and Information Science of Wuhan University*, 2017, 42 (7): 890-896.
- [28] 滕思翰. 基于 SPOT-5 的阿鲁科尔沁旗土地荒漠化动态变化研究[D]. 呼和浩特: 内蒙古农业大学, 2016.
TENG Sihan. Study on the Dynamic Changes of Land Desertification in Arhorchin Based on SPOT-5 [D]. Hohhot: Inner Mongolia Agricultural University, 2016.
- [29] YANG J, HUANG X. The 30 m Annual Land Cover Dataset and Its Dynamics in China from 1990 to 2019 [J]. *Earth System Science Data*, 2021, 13 (8): 3907-3925.
- [30] AI Earth. DAMO AI Earth [EB/OL]. (2021-06-24) [2023-04-30]. https://engine-aiearth.aliyun.com/#/dataset/DAMO_AIE_CHINA_LC/.
- [31] 殷崎栋, 柳彩霞, 田野. 基于 Landsat 时序影像和 LandTrendr 算法的森林保护区植被扰动研究: 以陕西柴松和太白山保护区为例[J]. 生态学报, 2020, 40(20): 7343-7352.
YIN Qidong, LIU Caixia, TIAN Ye. Detecting Dynamics of Vegetation Disturbance in Forest Natural Reserve Using Landsat Imagery and LandTrendr Algorithm: The Case of Chaisong and Taibaishan Natural Reserves in Shaanxi, China [J]. *Acta Ecologica Sinica*, 2020, 40(20): 7343-7352.
- [32] BREIMAN L. Random Forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [33] 雷震. 随机森林及其在遥感影像处理中应用研究 [D]. 上海: 上海交通大学, 2012.
LEI Zhen. Random Forest and Its Application in Remote Sensing [D]. Shanghai: Shanghai Jiao Tong University, 2012.
- [34] 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.
- [35] FASSNACHT F E, LATIFI H, STERENĆZAK K, et al. Review of Studies on Tree Species Classification from Remotely Sensed Data [J]. *Remote Sensing of Environment*, 2016, 186: 64-87.
- [36] GRIME J P. Vegetation Classification by Reference to Strategies [J]. *Nature*, 1974, 250 (5461) : 26-31.
- [37] EVANS J P, GEERKEN R. Classifying Rangeland Vegetation Type and Coverage Using a Fourier Component Based Similarity Measure [J]. *Remote Sensing of Environment*, 2006, 105(1): 1-8.
- [38] WANG X T, ZHANG S, FENG L L, et al. Mapping Maize Cultivated Area Combining MODIS EVI Time Series and the Spatial Variations of Phenology over Huanghuaihai Plain [J]. *Applied Sciences*, 2020, 10(8): 2667.
- [39] 田庆久, 闵祥军. 植被指数研究进展[J]. 地球科学进展, 1998, 13(4): 327-333.
TIAN Qingjiu, MIN Xiangjun. Advances in Study on Vegetation Indices [J]. *Advances in Earth Science*, 1998, 13(4): 327-333.
- [40] RICHARDSON A J, WIEGAND C L. Distinguishing Vegetation from Soil Background Information [J]. *Photogrammetric Engineering and Remote Sensing*, 1977, 43(12): 1541-1552.