

引用格式: Gai Yiming, Samat Alim, Wang Wei, *et al.* Sample Transferring based Fast Land Cover Updating in Arid Land[J]. Remote Sensing Technology and Application, 2022, 37(2): 333-341. [盖一铭, 阿里木·赛买提, 王伟, 等. 基于样本迁移的干旱区地表覆盖快速更新[J]. 遥感技术与应用, 2022, 37(2): 333-341.]  
DOI: 10.11873/j.issn.1004-0323.2022.2.0333

## 基于样本迁移的干旱区地表覆盖快速更新

盖一铭<sup>1,2,3</sup>, 阿里木·赛买提<sup>1,2,3</sup>, 王伟<sup>1,2,3</sup>, 吉力力·阿不都外力<sup>1,2,3</sup>

(1. 中国科学院新疆生态与地理研究所 荒漠与绿洲生态国家重点实验室, 新疆 乌鲁木齐 830011;  
2. 中国科学院大学, 北京 10049;  
3. 中国科学院中亚生态环境研究中心, 新疆 乌鲁木齐 830011)

**摘要:** 阿姆河三角洲作为典型干旱区, 干旱胁迫和次生的盐胁迫决定了本地区生态环境的复杂性和独特性, 给遥感地表覆盖制图带来一定的困难。在土地利用/覆盖(LULC)遥感图像分类任务中, 数量大、质量高、成本低的样本和速度快、性能稳定的分类器是高效实现高精度分类的关键。在一些偏远地区开展土地利用/地表覆盖遥感图像分类依然面临着标记样本空间上稀疏、时间上不连续甚至是缺失, 人工收集成本高等问题。为此, 结合最优树集成和样本迁移的思想, 构建了一种高效的地表覆盖自动更新的新方法。该方法通过变化检测在历史产品上的同期影像上进行样本标签的标记, 并将过去的地表覆盖类型标签转移到同源目标影像上, 使用最优树集成(Ensemble of optimum trees, OTE)完成地表覆盖自动分类。根据阿姆河三角洲地区地表覆盖分类试验结果, 表明该方法可以提取有效的地表覆盖标签, 并能较高精度地实现土地利用/地表覆盖的自动分类更新。

**关键词:** 样本迁移; 最优树集成; 变化检测; 地表覆盖变化; 遥感图像分类; 干旱区地表覆盖  
**中图分类号:** TP79      **文献标志码:** A      **文章编号:** 1004-0323(2022)02-0333-09

### 1 引言

土地利用和地表覆盖变化是全球变化研究领域的一个重要方向, 其动态也是人类对全球变化的响应核心。地表覆盖作为人类活动与自然要素共同干预与作用的结果, 是陆地表层生态系统的重要影响因素<sup>[1]</sup>。因此, 诸如地表过程模型、生态系统服务功能等研究中, 地表覆盖数据经常作为输入变量, 而地表覆盖分类的精度就成为决定研究结果可靠性的重要因素之一<sup>[2]</sup>。尽管目前已经出现大量开放获取的地表覆盖产品, 但是对于一些偏远地区, 可用的良好精度高分辨率产品缺乏且更新缓慢, 同时这些地区还面临全球产品在本地区精度低于平

均精度的难题<sup>[3]</sup>。因此, 对于单期高质量产品和同源的遥感数据集, 高效便捷的地表覆盖自更新方法亟待开发。

在遥感图像处理与地学应用研究中, 地表覆盖遥感图像分类是一项重要的工作。通常从是否有标记样本参与模型训练角度, 遥感图像分类可以分为监督、半监督、弱监督、非监督等4种类型<sup>[4]</sup>。地表覆盖的监督分类方法发展至今, 在输入的特征维度、基本操作单元以及分类器等方面经历了巨大进步<sup>[5]</sup>。但是, 无论数据源和分类器如何进步, 标记样本的获取仍然是建立健全分类模型的重要步骤。标记样本的提取不仅是分类精度的重要决定因素, 而且对于传统方法而言, 由于样本在空间上的分布

收稿日期: 2021-01-23; 修订日期: 2022-02-27

基金项目: 国家自然科学基金项目“样本与特征迁移的中亚典型城市覆被精细分类方法研究”(42071424), 中国科学院战略性先导专项“威海退缩产生的盐尘及其环境影响”(XDA2006030102), 中国科学院青年创新促进会(2018476)。

作者简介: 盖一铭(1995—), 男, 山东郯城人, 硕士研究生, 主要从事环境演变与气候变化研究。E-mail: gaiyiming18@mails.ucas.ac.cn

通讯作者: 吉力力·阿不都外力(1964—), 男, 新疆拜城人, 研究员, 主要从事干旱区资源环境研究。E-mail: jilili@ms.xjb.ac.cn

具有一定的离散性,外业工作在消耗人力与资金的同时,在一定程度上也影响着工期长短。因此,精确样本的快速获取对于缩短地表覆盖分类时间至关重要。

样本迁移是近年来流行的一种基于迁移学习(Transfer Learning, TL)的样本提取思想,相比传统方法而言,样本迁移避免了大量外业调查工作,有效地缩短了采样时间,减轻了分类操作人员的工作负担,节省了所需经费。近年来,吴田军和林聪等<sup>[6-7]</sup>的研究基于迁移学习的思想,通过往期产品向目标影像进行样本的转移,使得采样的工作量和时间大大减少,且在分类结果上具有良好的表现。刘万军等<sup>[8]</sup>通过迁移学习思想,结合光谱相似度方法,解决了小样本的问题。Pan 等<sup>[9]</sup>在变化向量分析的方法的基础上提出基于对象的扩展变化向量检测方法进行样本的迁移,通过支持向量机(Support Vector Machine, SVM)进行分类,较传统的基于像元的向量变化检测方法分类精度有了明显提升。

此外,深度学习(Deep Learning, DL)的方法也被引入到迁移学习的遥感影像分类中,除了良好的分类性能,还提供了强大的泛化能力以及伪标签生成,在一系列研究中取得非常良好的结果。Song 等<sup>[10]</sup>提出了一种基于迁移学习的三维滤波器深度学习框架,通过对训练集进行语义分割完成样本知识的迁移,提高变化检测的精度。Rafeal 等<sup>[11]</sup>比较不同神经网络深度学习模型和不同训练集上的迁移学习,肯定了迁移学习在遥感图像分类上的有效性。但是,深度学习方法也存在明显的缺点,方法的笨重决定其对设备算力的要求较高,运行时间较长以及对于跨领域的应用需求者而言具有较高的学习成本。对于很多地学和环境生态学研究人员而言,地表覆盖分类结果作为应用需求,只是研究的重要参数而不是研究主体。在精度满足应用要求的前提下,耗时、耗算力且学习成本更高的深度学习相比更快速便捷的机器学习方法对于精度的有限提升对用户不具有足够的吸引力。

对于机器学习分类器的优选,运算时间和分类精度都是重要考量因素。随机森林(Random Forest, RF)算法作为机器学习中最为经典的一种算法,大量研究表明,随机森林方法在不同研究区以及数据源的遥感影像分类工作中均取得良好的表现<sup>[12-14]</sup>。此外,极端随机树 ExtraTrees(Extremely Randomized Trees, ExT)算法是另一种决策树集成

算法,与随机森林算法非常相似,区别在于节点分裂方式和生长树的采样方式<sup>[5]</sup>。然而,RF 等方法的总体预测误差与森林中单株树木的强度及其多样性密切相关,因而生成的随机树森林中的树木达到一定数量后,增加树的数量不能提高预测精度<sup>[15]</sup>。而对森林中的决策树生长进行优化,可以减轻计算负担,提高分类速度<sup>[16-17]</sup>。因此,优化决策树节点和优选随机森林的树木是解决随机森林冗余的两大重要方法。极端梯度提升树(Extreme Gradient Boosting, XGBoost)算法对节点进行优化,是一个增强树模型<sup>[18]</sup>。此外,移除冗余树木和整体修剪是另一思路<sup>[19-20]</sup>。Khan 等<sup>[21]</sup>基于 Brier 评分提出了最优树集成方法(Ensemble of Optimal Trees, OTE),对经济、环境、体育、互联网、生命和健康等领域的 35 个数据集进行测试,并将分类和回归的结果同神经网络、决策树、随机森林和支持向量机等传统机器学习方法进行对比,结果表明,在大多数情况下,集合的规模显著减小,并且获得了更好的结果。虽然 OTE 方法在其他数据集上取得了良好的表现,但是作为一种比较年轻的机器学习方法,对于多光谱遥感影像的光谱-空间特征相结合地表覆盖分类这一研究方向,OTE 性能仍然有待探究。

阿姆河三角洲由于迅速膨胀的人口、巨大的农业生产压力以及不合理的土地和水资源管理制度,脆弱的生态环境一度恶化。因此,及时准确的土地利用与覆盖分类,以监测阿姆河三角洲土地退化,是保护该地区生态环境的迫切需求。然而,对于覆盖阿姆河三角洲地区的已有开源地表覆盖分类产品而言,精细空间分辨率产品时间尺度覆盖较短,精细时间分辨率的产品空间分辨率又比较粗糙,不同的分类产品还存在着分类体系不一致的问题。此外,研究区作为典型干旱区,土地利用与地表覆盖分类面临缺乏标记样本问题。本地区部分地表覆盖类型如灌丛和水体等,分布稀疏的同类地物图斑可能具有光谱和形态特征上的异质性,由于自然环境恶劣,人工采样费时费力且难以获取全面的样本。

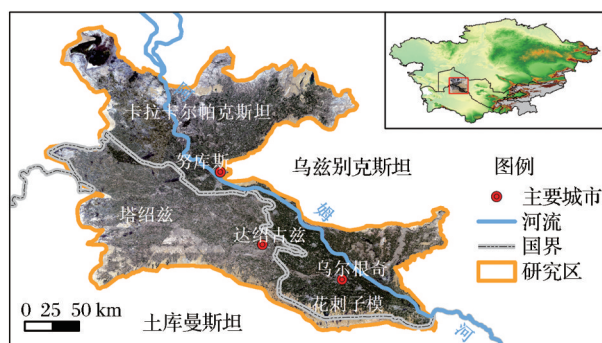
针对上述问题,本研究结合样本迁移思想和最优树集成方法,以阿姆河三角洲为研究区,使用产品迁移的标签,在 Landsat 8 影像上进行地表覆盖产品更新。实验结果证明了本研究提出的样本迁移方法的有效性,同时讨论了最优树集成方法在遥感影像分类上的适用性,为干旱、半干旱地区地表覆

盖产品快速更新提供技术支持。

## 2 研究区和数据

### 2.1 研究区概况

阿姆河三角洲位于阿姆河流域下游地区,由图雅姆雍起至梅杜热臣斯奇地区,地跨乌兹别克斯坦和土库曼斯坦两个国家。该区南起图雅姆雍水库,北至咸海湖盆,东临克孜勒库姆沙漠,西接乌斯泰尔特高原,面积3.6万km<sup>2</sup>。帕米尔高山上的大型永久冰川和积雪区是阿姆河的发源地,是阿姆河的主要水源,该地区东部和南部均与中亚半沙漠地带接壤,其特点是具有低降水和高辐照度的极端大陆性气候<sup>[22]</sup>,年平均降水量为80~120 mm。蒸散量大大超过年平均降水量。由于夏季的大风和高温<sup>[23-24]</sup>,年蒸发量非常高(约1 500 mm)。因此,大多数地区的水文平衡为负,生产生活用水依赖于河流的淡水流入<sup>[25]</sup>。尤其是位于乌兹别克斯坦的花刺子模省和卡拉卡尔帕克斯坦共和国,灌溉农业是区域经济的支柱<sup>[26]</sup>。由于缺水以及不合理灌溉事实上,浅层含盐地下水位广泛存在,该地区大多数农田受到不同程度土壤盐分的影响,进而导致土地退化<sup>[22, 27]</sup>。因此,阿姆河三角洲的特殊性就在于,不仅天然植被(草地、灌丛等)受到水分条件限制而具有复杂的生长策略,作物种植区因为盐渍化程度而具有生长良好、生长不健康乃至休耕等复杂的状态。由此可见,地表覆盖分类对于阿姆河三角洲荒漠化和盐渍化监测而言是必不可少的一项工作。



审图号: GS(2016)1665

图1 研究区示意图

Fig.1 The study area

### 2.2 数据来源

研究采用 Landsat-8 OLI 影像 (<http://earthexplorer.usgs.gov>), Landsat 数据的空间分辨率为30 m,波段范围为0.45~2.35 μm,具有易于获取和方便预处理等优点<sup>[28]</sup>。最为重要的是,Landsat产品

时间跨度较长,可以为长时序地表覆盖变化监测提供数据支撑<sup>[29-30]</sup>。研究使用的数据级别为L1T1,研究区共覆盖5幅影像,选取时间段为7月15日至9月15日。预处理基于ENVI5.3,对原始影像进行了几何配准、裁剪、镶嵌、辐射定标、大气校正等一系列操作。

为了方便样本的迁移,研究使用的样本标签来自FROM-GLC 2015,数据来源于清华大学地球系统科学系 (<http://data.ess.tsinghua.edu.cn>),空间分辨率为30 m,该系列最早一期的产品基于2010年的Landsat图像生成<sup>[31]</sup>。FROM-GLC系列产品在随后更新到了2015年,最低精度也由63.69%提升至72.43%。

## 3 研究方法

研究所用方法是样本迁移支持下的一系列算法的集合,总体实现如图2所示。本文构架的自动分类算法核心在于,通过变化检测和阈值分割等一系列操作,获取原域影像数据集到目标域影像数据集的变化信息,然后和源域分类知识进行叠加分析,以实现样本自动采集。基于样本迁移思想,共提取训练样本24 932个,验证样本24 932个。为比较OTE方法分类性能,实验设计了支持向量机(Support Vector Machine, SVM)、随机森林、极端梯度提升树和极端随机树方法作为比较。以样本来源的FROM-GLC 2015的最大精度作为阈值,以此为依据判断精度是否合格,在这一过程中除极端梯度提升树之外其他方法均一次性通过阈值判断,完成训练分类出图。本研究的核心是基于样本迁移的标签转移和最优树集成算法,下文给出了方法的阐述和关键环节的说明。

### 3.1 基于样本迁移的标签提取

土地资源作为重要自然资源和社会经济资源,具有整体性、位置固定性和生产性等属性,基于这些特性,一定区域的土地表层覆盖随着社会发展发生一定变化的同时,存在大量土地的地表覆盖仍然维持其原有状态。而遥感影像分类的基础就在于,相同的地物具有共同的光谱或形态特征。实地采样人工解译没有充分利用未改变的地物光谱特征,降低了分类时效性。假设已有地表覆产品对应时相的遥感影像数据集为 $A = \{x_i^t | i=1, \dots, n\}$ ,待分类时相的遥感影像数据集为 $B = \{y_i^t | i=1, \dots, n\}$ , $m$ 为波段数, $n$ 为像素数。则存在 $X$ 的时相到 $Y$ 的时相,未发生变化耕



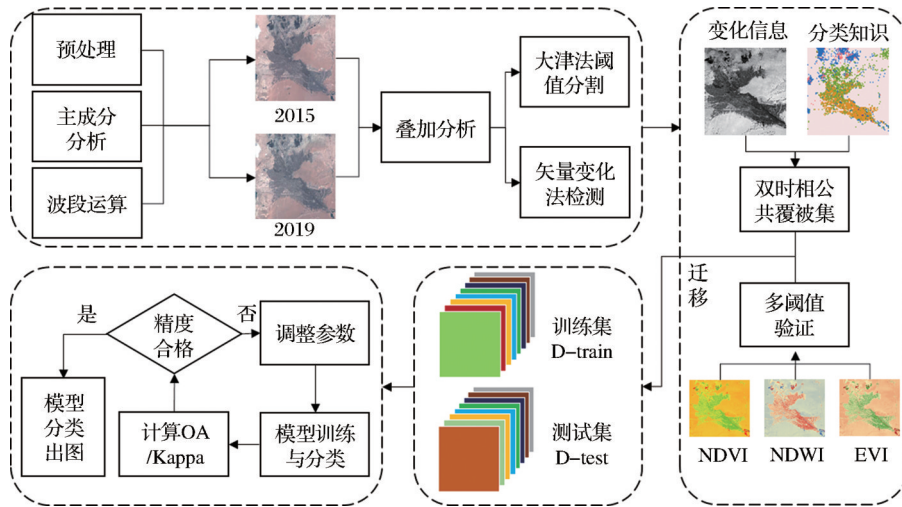


图 2 样本迁移的遥感影像分类流程图

Fig.2 Flowchart of transferred learning based remote sensing image classification

地、建筑、水域等地物,各类地物光谱分布的统计差异具有鲁棒性,其对应  $A^k \sim B^k$  ( $A^k \in A, B^k \in B, k \leq j$ )。样本迁移的工作就是在待分类影像中没有发生改变的地物  $B^k$  上采集样本,用于对影像进行地表覆盖分类。当目标影像中有少量标注样本时,这种迁移的方法被称为归纳迁移学习 (Inductive Transfer Learning, ITL),当目标影像中没有任何标签样本时,则为直推式迁移学习 (Transductive Transfer Learning, TTL)<sup>[6]</sup>。

为了找到没有发生变化的地物  $B^k$ ,本研究采用基于像元的光谱特征变化检测方法。首先对  $A$  和  $B$  光谱进行直接比较,采用矢量变换分析法 (Change Vector Analysis, CVA),它可以充分利用多光谱图像的所有波段的信息,并产生可靠的变化信息。CVA 的基本理论可以描述为:在同一地理区域内获得的前后时相  $n$  个波段组成的两幅多光谱影像,逐个像元计算各个光谱的差值,

$$\Delta_i^j = y_i^j - x_i^j \quad (1)$$

其中:  $y_i^j - x_i^j$  是  $j$  像元处  $i$  波段两幅影像灰度差值,第  $j$  个像元处的变化幅度被定义为它的欧氏距离:

$$\rho^j = \sqrt{\sum_{i=1}^m (\Delta_i^j)^2} \quad (2)$$

由于不同地物所具有的不同光谱特征,对于  $\rho^j$ ,可以找到某一阈值,将所有像元分成两个部分,即来自没有发生变化地物的像元和来自已经发生变化地物的像元。研究采用大津法<sup>[32]</sup> (Ostu's)对结果进行阈值分割,找到未发生变化的地物斑块,然后进行采样。

考虑到样本标签的来源是全球产品,所以可能

存在错误标签。对于未发生变化的地物斑块,结合 Google Earth 目视解译删除错误斑块。样本数量的增加可以一定程度地减少错误样本对分类精度的影响,但是样本标签的精度依然是决定分类器性能的重要因素。为了提高标签的精度,结合归一化植被指数、归一化水体指数和增强植被指数,分别对各种地物设置阈值,阈值范围之外的采样点被视为不可靠样本进行剔除。

### 3.2 最优树集成

随机森林通过 Bagging 和随机空间子集,引入了树的基本模型以及节点分化的随机性<sup>[33]</sup>。随机森林由于存在大量的冗余树木往往规模很大,为削减冗余树木,在保持森林性能的基础上最小化森林的规模,Khan 等<sup>[21]</sup>提出了最优树集成的方法,该方法利用未被解释的方差所反映的树木多样性和个体精度进行选择,从而细化随机森林。研究采用最优树集成方法进行遥感影像的地表覆盖分类,将训练数据  $L = [X, Y]$  随机划分为两个互不重叠的子集  $L_U = [X_U, Y_U]$  和  $L_V = [X_V, Y_V]$ ;接下来,对第一个子集  $L_U = [X_U, Y_U]$ ,在 T 自助法采集的样本上生长分类树。同时,从整个项目属性集  $d$  中选择  $p < d$  特征的随机样本。这给树木增加了额外的随机性。由于自助方法的使用,样本中会有一些观测值被遗漏,这些被称为袋外 (OOB) 观测值。后者不参加树的训练。它们被用来估计建立在自举样本上的每棵树的无法解释的方差。然后,根据无法解释的差异,按照升序对树进行分类,并选择排名最高的  $M$  棵树。树木的选择和组合如下进行:

(1) 从两个排名靠前的树开始,依次添加连续

排名的树,直到最后的第 $M$ 棵树被测试,以查看它们在独立验证数据 $L_v=[X_v, Y_v]$ 上的表现。

(2)选择树的子集 $L_k(k=1,2,3,\cdots,M)$ ,使树的全集相对于 $L_k$ 的补集满足如下的分类准则:

$$BS^{<k+>} < BS^{<k->} \quad (3)$$

这里 $BS^{<k->}$ 是不含有 $k$ 棵树的分类树集合, $BS^{<k+>}$ 则是包含了 $k$ 棵树的分类树集合。

$$BS = \frac{\sum_{i=1}^{\text{sum}} (y_i - \hat{P}(y_i|\mathbf{X}))^2}{\text{sum}} \quad (4)$$

其中: $y_i$ 是观测 $i$ 在 $(0,1)$ 形式下 $y_i$ 的形态, $\hat{P}(y_i|\mathbf{X})$ 是当前集合能否预测给定特征的二值化反应。这些树便被称为最优树,然后它们被集合起来进行投票以获取最终的分类结果。结果的集合被命名为最优树集成,即OTE。上述的分类算法步骤如下:

(1)采用自助采样法从给定集合中提取训练样本 $L_u=[X_u, Y_u]$ 。

(2)按照随机森林路径在全部自举样本上生长分类树。

(3)根据袋外样本的预测误差升序排列所有回归树,选取个体预测误差最小的前 $M$ 棵分类树。

(4)在 $M$ 棵分类树中,依据Breir评分依次将能在验证数据集上提升分类表现的分类树添加到集合中。

(5)将上述的分类树结合起来进行投票,给出分类结果。

## 4 试验结果

研究采用的分类标签来自FROM-GLC 2015,对应2015年的Landsat影像,待分类的目标影像为2019年的Landsat影像。利用上文提到的样本迁移方法,对两幅影像进行变化检测。为避免阈值分割的误差造成的精度下降,采样前过滤掉面积小的地物斑块(孤立像元无法通过目视解译验证)。FROM-GLC 2015在该地区的地表覆盖类型按照面积大小依次为裸地、耕地、草地、水体、湿地、灌丛、建筑(不透水面)和林地。由于森林面积过小,为采集到足够的样本分别对每个类型随机采样,而不是对整个未变化图斑进行随机采样。为了增强结果的可信度,在前面提到的植被指数和水体指数之外,结合裸地指数和夜间灯光数据,对验证样本初步筛查,筛查过的验证样本通过比对目视解译的图斑,进行验证样本的选择。为对比OTE的分类性能,通过5种分类器分别进行分类,对分类结果进行

经度验证,计算了分类结果的各类地物的总体精度和Kappa系数,见表1。

表1 5种分类器总体精度

Table 1 Overall accuracy of five classifiers

类别	分类精度/%				
	SVM	RF	XGB	ExT	OTE
耕地	76.50	77.82	76.30	80.24	77.38
森林	93.30	94.04	94.35	96.32	95.76
草地	73.99	78.24	72.58	80.26	79.81
灌丛	87.70	90.28	89.17	90.32	90.82
湿地	75.82	77.33	78.56	82.54	84.12
水体	95.15	96.09	96.23	96.50	95.86
裸地	95.87	96.80	96.52	96.34	97.08
建筑	85.88	85.44	87.06	86.77	89.42
OA	85.46	87.07	86.29	88.78	88.88
Kappa	0.83	0.85	0.84	0.87	0.87

从验证结果来看,最优树集成在总体精度上取得了最好的表现,同时在灌丛、湿地、裸地和不透水地表5个类别中取得了最高的分类精度,而在耕地、森林和草地中,ExT的分类精度更高。此外,RF、ExT和OTE算法在各个地表覆盖类型以及整体的分类表现均优于SVM和XGB算法。从表1中还可以看出,基于RF的两种改进算法—ExT和OTE以89.48%和89.48%的总体精度(OA),超越了RF的87.92%。可以看出,相比于SVM,有3种基于决策树的集成分类方法在本实验中取得了更良好的精度。对于这一现象,研究从特征和样本两个输入变量进行分析。由于SVM和集成学习方法都是高维特征的解决方案,因此本实验中的特征维度不是OTE和ExT等方法精度更高的主要原因。由于本文样本来源于产品,虽然通过光谱指数方法剔除掉了一部分错误的样本,但是难以保证完全样本的精确。从几种机器学习方法的原理的角度可以解释为,由于集成学习通过多个异质的分类器投票决定分类的结果,方差减小的抽样方案更有利于在不完全精确样本中筛选正确信息。因此,研究在基于不完全精确的样本集合的遥感影像分类中,RF、Extra-Trees和OTE等集成学习/多分类器集成方法相对于SVM和XGBoost而言有着更为良好的表现。

如表1所示,具体的地表覆盖类型上,所有5种方法在裸地、水体和森林上的分类都具有良好的分类性能,精度均达到90%以上。其中,OTE在裸地的分类精度最高,达到了97.08%,ExT在水体和森林上的分类精度最高,分别为96.50%和96.32%。



裸地、水体和建筑的光谱特征明显区别于植被因此精度良好,而森林的分类精度明显优于其他各种植被类型,原因如下:首先,森林相比其他地表覆盖类型面积较小,在采集样本量巨大的情况下,训练样本和验证样本来自同一森林图斑的概率高于其他地表覆盖类型;其次,森林的光谱和纹理特征相对其他植被有着一定的显著区别。

但是,5种方法在耕地和草地上的精度较低,由表1可知,ExT在以80.24%的精度优于其他几种方法在耕地上的分类表现,在草地分类,ExT也是表现最好的算法,精度可达80.26%,相较于RF的精度,可以认为算法优化效果显著。几种分类方法在缺点上表现出一致性,即在耕地和草地和湿地的分类精确性明显低于其他地表覆盖类型。耕地和草地的光谱特性、破碎的分布和长势的稀疏导致一定的耕—草误分,但是研究区尤其是位于阿姆河尾间的部分地区耕地又较为依赖水源,较小面积的耕地图斑容易与周围湿地形成边缘误分,本实验获得的

分类效果可以认为是合理的。

为了直观分析制图效果,本研究将标签源域FROM-GLC 2015和OTE分类结果进行对比。由于FROM-GLC 2015将很多弃耕的水田分为水体或是湿地,其分类结果对这两种地类产生了一定的高估,同时,对不透水面的识别有一定的缺陷,比如细长条形的不透水面未能得到良好的识别而本实验的样本标签来自FROM-GLC 2015,实验结果继承了FROM-GLC 2015倾向于本地区高估水体和湿地的缺点,这些问题可以看做样本对分类结果的影响。由此可见,对于样本迁移的机器学习分类而言,样本供体的质量决定着分类结果的精确性,分类产品作为样本供体,其缺点在样本迁移中容易被继承。然而,本研究分类结果在很多方面的表现依然优于源产品,在农田和裸地之间存在的草地过渡区域就被本文方法很好地提取,然而这一区域在源产品中并没有被很好地识别。此外,图3的结果为分类后没有经过任何滤波等分类后处理的生图,因此

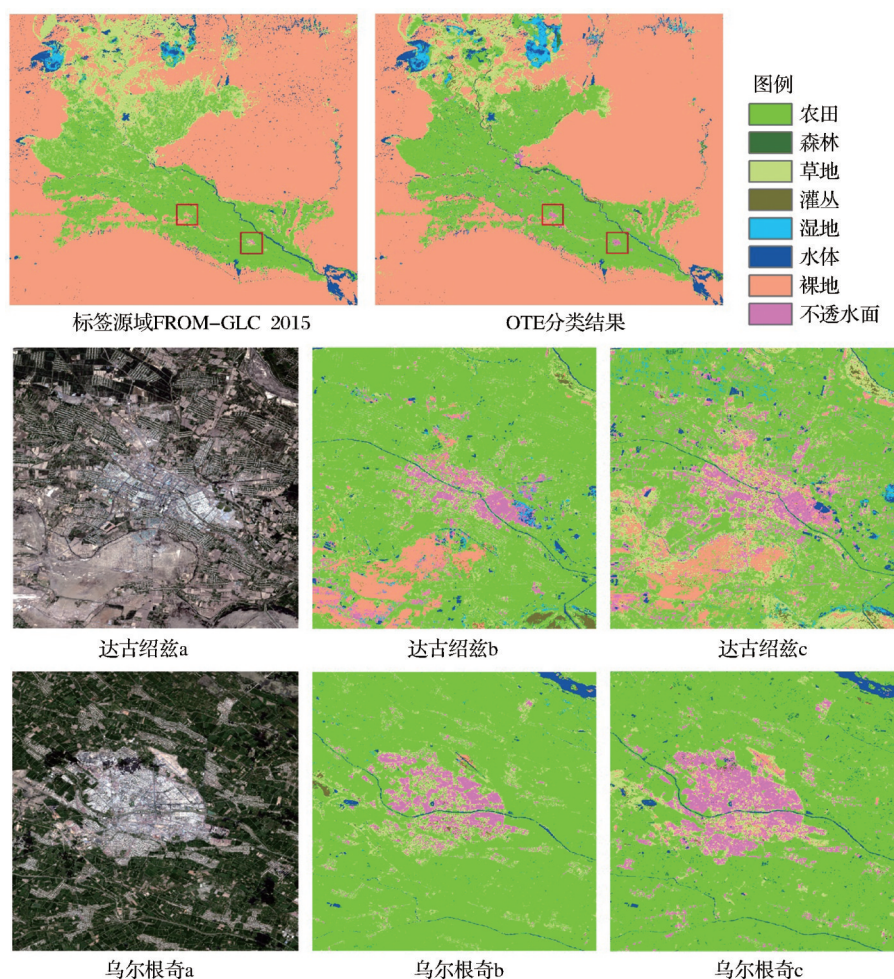


图3 分类结果对比(a为遥感影像,b为产品,c为OTE分类结果)

Fig.3 Comparison of classification results

结果完全服从地物光谱特征。达古绍兹城区不透水面图斑的破碎性和乌尔根奇城区不透水面图斑的完整性很直观地反映出两地的社会经济差异(人口密度和基建水平),然而这一点在产品中就没有得到体现。此外,研究区的盐渍化现象较为严重,因此存在一定量的弃耕土地,由于土壤水分差异,这些土地往往会变为草地和裸地。在达古绍兹的分类结果中,弃耕的土地得到了一定程度的识别。

由于多阈值验证对错误样本的剔除,分类的精度略高于样本来源FROM-GLC 2015,但是从分类图像来看五种方法都存在以下问题:首先,纹理特征的加入未能完全消除椒盐现象;其次,存在裸地和不透水地表相互误分的问题,这种误分在咸海湖盆地区较为明显。这些问题的原因主要是对于多光谱影像而言,存在异物同谱的现象,造成各种相似光谱特征地物的误分。由于经济落后,泥质建筑、裸岩以及咸海湖盆的干涸盐类结皮是裸地—建筑误分的重要来源,这些地物是本研究区的地表覆盖分类的一大难题。此外,混合像元的存在也在一定程度上影像面对像元的分类表现。由于裸地占研究区总面积的65.19%,被误分的裸地像元在裸地总像元中的比例不高,所以对裸地最终分类精度造成的影响较小。

此外,本文方法基于OptiPlex7070工作平台,从训练时间来看,SVM方法用时最长,2 614秒的训练时间明显超出其他几种基于决策树的集成学习方法。RF、ExT和XGBoost训练秒数由高到低分别为975、895和802(图4),体现了基于决策树方法相对于支持向量机的优势,而OTE方法以558秒明显短于其他几种基于决策树的集成学习方法。通过精度和时间的比较,SVM方法工作效率最低而OTE方法工作效率最高。

总体而言,OTE和ExT的精度良好,OTE精度

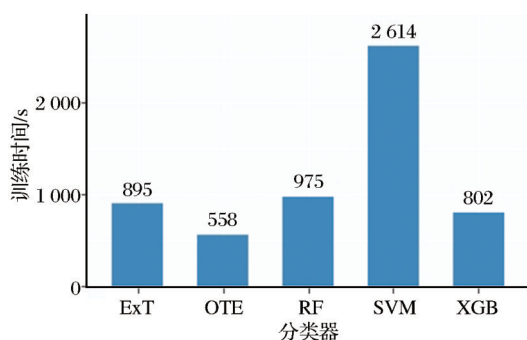


图4 各分类器训练时间对比图

Fig.4 Comparison of training time of each classifier

略高于ExT,ExT在耕地、森林、草地和水体的分类表现最好,而OTE在灌丛、湿地、裸地和不透水地表的分类表现最好。但是由表2可知,本地区森林、灌丛和湿地在本地区分布极为稀少,合计占到研究区总面积的1.17%,在实际应用,尤其是土地利用动态和生态环境变化的监测中,植被分类的精度至关重要。因此,在本研究的5种分类方法中,对于研究区主要植被类型耕地和草地上,一些情况下ExT可以被认为更适合本研究区的地表覆盖分类算法。而OTE精度略高,考虑其在工作效率上的明显优势,OTE方法在本实验中总体表现最佳。

表2 各个地表覆盖类型的面积占比

地类	耕地	森林	草地	灌丛	湿地	水体	裸地	建筑
百分比/%	16.56	0.007	14.13	0.22	0.90	2.80	65.19	0.13

## 5 结 论

本研究基于样本迁移的思想,构建了对地表覆盖产品的快速更新框架。采用FROM-GLC 2015产品作为标签来源,基于矢量变化分析法量化像元级光谱特征变化,提取未变化区域的标签,考虑到样本源域FROM-GLC 2015有提升空间,结合植被指数等信息,对不可靠样本进行了剔除,然后利用迁移的样本标签进行2019地表覆盖更新。除去影像预处理的操作,其余过程基本实现计算机自动化的地表覆盖更新。在方法的优选上,最优树集成方法更新的总体精度可达88.88%,略优于其他几种机器学习算法,这一精度的结果就大部分相关的地学和生态学研究而言,都是满足使用要求的。同时,OTE、ExT和RF的精度表明,集成学习/多分类器集成方法对不完全精确样本集表现更为良好,因此集成学习在样本迁移工作中更值得推荐。从实验的训练用时来看,样本迁移的最优树集成比样本迁移的随机森林的分类速度提升了约42.77%。此外,样本迁移的更新效果对于提供标签的产品FROM-GLC 2015表现出继承性,证明了样本标签的有效迁移。基于样本迁移和最优树集成的方法可以快速且准确地实现干旱区地表覆盖更新,实现高质量产品的扩展应用,为相关用户提供技术参考。

## 参考文献(References):

- [1] Grimm N B, Faeth S H, Golubiewski N E, *et al.* Global change and the ecology of cities[J]. Science, 2008, 319(5864): 756-60. DOI: 10.1126/science.1150195.

- [2] Fischer M, Bossdorf O, Gockel S, *et al.* Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories[J]. Basic and Applied Ecology, 2010, 11(6): 473-485.
- [3] Stehman S V, Foody G M. Key issues in rigorous accuracy assessment of land cover products[J/OL]. Remote Sensing of Environment, 2019, 23123. <https://DOI.org/10.1016/j.rse.2019.07.009>.
- [4] Du P J, Bai X, Tan K, *et al.* Advances of four machine learning methods for spatial data handling: A review[J/OL]. Journal of Geovisualization and Spatial Analysis, 2020, 4(1): <https://DOI.org/10.1007/s41651-020-00048-5>.
- [5] Du Peijun, Xia Junshi, Xue Zhaohui, *et al.* Review of hyperspectral remote sensing image classification[J]. Journal of Remote Sensing, 2016, 20(2): 236-256.[杜培军, 夏俊士, 薛朝辉. 高光谱遥感影像分类研究进展[J]. 遥感学报, 2016, 20(2): 236-256.]
- [6] Wu Tianjun, Luo Jiancheng, Xia Liegang, *et al.* An automatic sample collection method for object-oriented classification of remotely sensed imageries based on transfer learning[J]. Acta Geodetica et Cartographica Sinica, 2014, 43(9): 908-916. [吴田军, 骆剑承, 夏列钢. 迁移学习支持下的遥感影像对象级分类样本自动选择方法[J]. 测绘学报, 2014, 43(9): 908-916.]
- [7] Lin Cong, Li Erzhu, Du Peijun. An automatic approach for remote sensing classification supported by Sample Transfer[J]. Bulletin of Surveying and Mapping, 2018(4): 57-62.[林聪, 李二珠, 杜培军. 样本迁移支持下的遥感影像自动分类方法[J]. 测绘通报, 2018(4): 57-62.]
- [8] Liu Wanjuan, Li Tianhui, Qu Haicheng. Hyperspectral similar sample classification algorithm based on fisher criterion and TrAdaboost[J]. Remote Sensing for Land & Resources, 2018, 30(4): 41-48.[刘万军, 李天慧, 曲海成. 基于Fisher准则和TrAdaboost的高光谱相似样本分类算法[J]. 国土资源遥感, 2018, 30(4): 41-48.]
- [9] Pan H, Tong X, Xu X, *et al.* Updating of land cover maps and change analysis using GlobeLand30 product: A case study in Shanghai metropolitan area, China[J]. Remote Sensing, 2020, 12(19): <https://DOI.org/10.3390/rs12193147>.
- [10] Song A, Choi J. Fully convolutional networks with multiscale 3D filters and Transfer Learning for change detection in high spatial resolution satellite images[J]. Remote Sensing, 2020, 12(5). <https://DOI.org/10.3390/rs12050799>.
- [11] De Lima R P, Marfurt K. Convolutional neural Network for remote sensing scene classification: Transfer Learning Analysis[J]. Remote Sensing, 2020, 12(1): 20. <https://DOI.org/10.3390/rs12010086>.
- [12] Gu Chunxia, Li Dacheng. Application of spatio-temporal fusion technology in time series classification of regional land cover[J]. Bulletin of Surveying and Mapping, 2021(5): 20-24.[古春霞, 李大成. 时空融合技术在区域地表覆盖时序分类中的应用[J]. 测绘通报, 2021(5): 20-24.]
- [13] Li Binbin, Xie Huan, Tong Xiaohua, *et al.* Land cover classification using ICESat-2 data with random forest[J]. Infrared and Laser Engineering, 2020, 49(11): 115-121. [李彬彬, 谢欢, 童小华, 等. 基于随机森林的ICESat-2卫星数据地表覆盖分类[J]. 红外与激光工程, 2020, 49(11): 115-121.]
- [14] Li Bing, Lu Xiaoping, Li Xinshe, *et al.* Extraction method of wetland surface factors using GF-2 images on optimal features[J]. Geomatics & Spatial Information Technology, 2018, 41(9): 49-52.[李冰, 卢小平, 李新社, 等. 特征优选的GF-2影像湿地地表覆盖要素提取[J]. 测绘与空间地理信息, 2018, 41(9): 49-52.]
- [15] Oshiro T M, Perez P S, Baranauskas J A. How many trees in a random forest[C]// Proceedings of the Machine Learning and Data Mining in Pattern Recognition, F, 2012.
- [16] Bernard S, Heutte L, Adam S. On the selection of decision trees in Random Forests[C]// Proceedings of the International Joint Conference on Neural Network, F, 2009.
- [17] Abdelali Z, Hicham M, Abdelwahed N. An Ensemble of Optimal Trees for Software Development Effort Estimation[C]// Proceedings of the International Conference Advanced Information Technology, Services and Systems, F, 2018.
- [18] Samat A, Li E Z, Wang W, *et al.* Meta-XGBoost for hyperspectral image classification using Extended MSER-Guided morphological profiles[J]. Remote Sensing, 2020, 12(12): 23. DOI: 10.3390/rs12121973.
- [19] Zhang H, Wang M. Search for the smallest random forest[J]. Statistics and Its Interface, 2009, 2(3): 381-388. DOI: 10.1093/biostatistics/kxaa021.
- [20] Adler W, Gefeller O, Gul A, *et al.* Ensemble pruning for glaucoma detection in an unbalanced data set[J]. Methods of Information in Medicine, 2016, 55(6): 557-63. DOI: 10.3414/ME16-01-0055.
- [21] Khan Z, Gul A, Perperoglou A, *et al.* Ensemble of optimal trees, random forest and random projection ensemble classification[J]. Advanced Data Analysis and Classification, 2020, 14(1): 97-116. DOI: 10.1007/s11634-019-00364-9.
- [22] Mo Guifen, Feng Jianzhong, Wang Zhongmei, *et al.* Spatial-temporal evolution characteristics of landscape ecological risk in the transboundary basin of Amu Darya River, Central Asia[J]. Agricultural Research in the Arid Areas, 2022, 40(1): 123-131.[莫贵芬, 冯建中, 王中美, 等. 中亚阿姆河跨境流域景观生态风险时空演变特征分析[J]. 干旱地区农业研究, 2022, 40(1): 123-131.]
- [23] Hu Y, Duan W, Chen Y, *et al.* An integrated assessment of runoff dynamics in the Amu Darya River Basin: Confronting climate change and multiple human activities, 1960-2017[J]. Journal of Hydrology, 2021, 603: 126905. <https://DOI.org/10.1016/j.jhydrol.2021.126905>.
- [24] He H, Hamdi R, Cai P, *et al.* Impacts of historical land use/cover change (1980 - 2015) on summer climate in the Aral Sea Region[J]. Journal of Geophysical Research: Atmo-



- spheres, 2021, 126(6): e2020JD032638.
- [25] Han Q F, Luo G P, Li C F, *et al.* Response of carbon dynamics to climate change varied among different vegetation types in Central Asia[J]. Sustainability, 2018, 10(9): 15. DOI: 10.3390/su10093288.
- [26] Li Qi, Li Fadong, Wang Guoqin, *et al.* Development of irrigated agriculture in Uzbekistan and its impact on ecological environment and economic development[J]. Arid Land Geography, 2021, 44(6): 1810-1820. [李琦, 李发东, 王国勤, 等. 乌兹别克斯坦灌溉农业发展及其对生态环境和经济发展的影响[J]. 干旱区地理, 2021, 44(6): 1810-1820.]
- [27] Ye H, Huang X T, Luo G P, *et al.* Improving remote sensing-based net primary production estimation in the grazed land with defoliation formulation model [J]. Journal of Mountain Science, 2019, 16(2): 323-336. <https://DOI.org/10.1007/s11629-018-5200-2>.
- [28] Roy D P, Wulder M A, Loveland T R, *et al.* Landsat 8: Science and product vision for terrestrial global change research [J]. Remote Sensing of Environment, 2014, 145: 154-172. DOI:10.1016/j.rse.2014.02.001.
- [29] Zhu Z, Woodcock C E. Continuous change detection and classification of land cover using all available Landsat data[J]. Remote Sensing of Environment, 2014, 144: 152-171. DOI: 10.1016/j.rse.2014.01.011.
- [30] Zhang Xiao. Global land-cover classification and mapping at 30 m using quantitative remote sensing technique[D]. Beijing: University of Chinese Academy of Sciences (Aerospace Information Research Institute), 2020. [张肖. 全球 30 m 地表覆盖定量遥感分类与制图研究[D]. 北京: 中国科学院大学 (中国科学院空天信息创新研究院), 2020.]
- [31] Gong P, Wang J, Yu L, *et al.* Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data[J/OL]. International Journal of Remote Sensing, 2013, 34(7): 2607-2654. <https://DOI.org/10.1080/01431161.2012.748992>.
- [32] Otsu N. Threshold selection method from gray-level histogram[J]. IEEE Transactions on Systems Man and Cybernetics, 1979, 9(1): 62-66.
- [33] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32. <https://DOI.org/10.1023/A:1010933404324>.

## Sample Transferring based Fast Land Cover Updating in Arid Land

Gai Yiming<sup>1,2,3</sup>, Samat Alim<sup>1,2,3</sup>, Wang Wei<sup>1,2,3</sup>, Abuduwaili Jilili<sup>1,2,3</sup>

(1. State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China;

2. University of Chinese Academy of Sciences, Beijing 10049, China;

3. Chinese Academy of Sciences Research Center for Ecology and Environment of Central Asia, Urumqi 830011, China)

**Abstract:** Amu river delta, as a typical arid land, was threatened by drought and salination, which contribute to the complexity and specificity of its ecological environment. In the Land Use/Land Cover (LULC) Remote Sensing (RS) image classification tasks, collecting large number of high quality samples at low-cost and a high efficient and robust classifier are always the crucial factors to obtain high-accuracy classification results. However, it was still problems facing RS imageries classification in some remote areas that marked samples were sparsely distributed, timely dissected or even intermittent, and manual tasks for field sampling cost high. In this end, a new frame of automatic land cover classification based on ensemble of optimum trees and sample transfer was promoted in this paper. In this frame, sample labels were marked on the historical image which is same time and source with the product, then these labels were transferred into targeted RS image. Then, OTE method classification was performed. According to the results in this paper, the OTE with sample transferring based method can extract land cover labels efficiently and update LULC map in a fine accuracy.

**Key words:** Sample transfer; Ensemble of optimum trees; Change detection; Land cover change; RS image classification; Arid land cover