

doi: 10.6046/zrzyg.2021061

引用格式: 吴琳琳, 李晓燕, 毛德华, 等. 基于遥感和多源地理数据的城市土地利用分类[J]. 自然资源遥感, 2022, 34(1): 127–134. (Wu L L, Li X Y, Mao D H, et al. Urban land use classification based on remote sensing and multi-source geographic data[J]. Remote Sensing for Natural Resources, 2022, 34(1): 127–134.)

# 基于遥感和多源地理数据的城市土地利用分类

吴琳琳<sup>1,2</sup>, 李晓燕<sup>1</sup>, 毛德华<sup>2</sup>, 王宗明<sup>2</sup>

(1. 吉林大学地球科学学院, 长春 130012; 2. 中国科学院湿地生态与环境重点实验室, 中国科学院东北地理与农业生态研究所, 长春 130102)

**摘要:** 城市土地利用信息体现了城市的功能及结构, 开展城市土地利用分类研究对于城市可持续发展具有重要指导意义。以哈尔滨市主城区为研究区, 综合应用 Sentinel-2A 遥感影像、OpenStreetMap (OSM) 数据、兴趣点 (point of interest, POI) 数据和珞珈一号夜间灯光数据等多源地理数据, 采用面向对象和随机森林算法对哈尔滨城市土地利用进行分类。结果表明: 一级地类总体精度为 86.0%, Kappa 系数为 0.75; 二级地类总体精度为 73.9%, Kappa 系数为 0.69; POI 数据可以显著提高住宅用地、工矿仓储用地和教育用地分类精度, 夜间灯光数据能有效提高商务办公用地及商业用地分类精度。说明遥感影像与多源地理数据结合对城市土地利用分类有效。

**关键词:** 随机森林; 多尺度分割; 多源数据; 城市土地利用分类; 哈尔滨

**中图法分类号:** TP 79 **文献标志码:** A **文章编号:** 2097-034X(2022)01-0127-08

## 0 引言

城市土地利用信息反映着社会经济功能和活动, 是城市规划和区域管理的重要基础<sup>[1-2]</sup>, 对于解决城市问题、科学合理的发展城市具有重要意义<sup>[3-5]</sup>。因此, 准确、详细的城市土地利用信息对城市规划和发展起到关键性作用<sup>[6-8]</sup>。然而, 城市土地利用类型复杂多样, 人工进行实地调查费时费力, 获取详细的城市土地利用信息较为困难<sup>[9-10]</sup>。国内外学者基于遥感影像获取城市土地利用信息展开了丰富的研究。Gong 等<sup>[11]</sup>利用随机分类方法对中国 332 个地级市进行城市土地利用分类研究; Zong 等<sup>[12]</sup>基于多源数据, 采用随机森林算法对兰州市城市土地利用进行分类并分析不同特征组合对分类结果的影响; Zeferino 等<sup>[13]</sup>利用 Landsat8 卫星影像研究环境数据对土地利用的影响; Saboori 等<sup>[14]</sup>利用高分辨率影像多尺度的纹理特征与神经网络算法相结合, 对城市土地利用信息进行提取。

随着城市化的不断发展, 快速高效获取城市土地利用信息成为研究热点。随机森林是一种新

型的集成算法, 具有训练样本数量少、人工干预少、精度较高、不易过拟合的特点, 相较于其他传统方法具有很好的鲁棒性<sup>[15-16]</sup>。目前, 随机森林已广泛运用到遥感影像分类中。周天宁等<sup>[17]</sup>基于 Landsat TM 影像, 利用随机森林优化算法对昌平区进行土地利用分类, 其结果验证了随机森林在土地利用分类的稳定性; 马玥等<sup>[18]</sup>基于随机森林算法, 对齐齐哈尔市周边的农业耕种区进行土地利用分类研究, 实现了分类效率、精度及适用性之间的平衡, 为农业耕种提供参考; 杨耘等<sup>[19]</sup>利用高分辨率影像提取城市土地利用分类, 通过多个算法对比发现随机森林算法对城市土地利用类型提取效果较好, 可以为城市土地利用分类提供准确的依据。

传统的土地利用调查无法满足城市土地管理的需求。随着通信技术的不断发展, 热力图数据、社交媒体数据、兴趣点数据等具有地理意义的空间大数据获取方便, 多源数据为城市空间结构的研究提供了依据和便利<sup>[20]</sup>。如兴趣点 (point of interest, POI) 数据与传统的土地调查数据相比, POI 数据样本数量多、易获取, 可以记录土地利用类型及位置信息<sup>[21]</sup>, 更能准确详细地展现城市空间布

收稿日期: 2021-03-05; 修订日期: 2021-08-13

基金项目: 吉林省自然科学基金项目“吉林省中部玉米带城市化进程对耕层有机碳储量的影响”(编号: 20200201048JC)资助。

第一作者: 吴琳琳(1998-), 女, 硕士研究生, 主要研究方向为资源遥感与土地信息系统研究。Email: wull19@mails.jlu.edu.cn。

通信作者: 李晓燕(1975-), 女, 教授, 主要研究方向为资源遥感与土地信息系统研究。Email: lxyan@mails.jlu.edu.cn。

局,满足城市土地利用分类的需要。

哈尔滨市为黑龙江省省会城市,是中国 15 座副省级城市之一。哈尔滨地处中国东北地区,是东北北部的政治、经济、文化中心。哈尔滨市城市用地类型较多,获取困难,研究哈尔滨城市土地利用分类对促进城市发展和解决城市问题具有重要意义。选择哈尔滨城市用地为研究区,分析多源地理空间数据对城市土地利用类型的提取能力,利用多尺度分割和随机森林算法对哈尔滨市城市用地进行分类,为城市规划及管理提供依据。

## 1 研究区概况与数据源

### 1.1 研究区概况

哈尔滨市(图 1)位于东北平原,地处  $E125^{\circ}42' \sim 130^{\circ}10'$ ,  $N44^{\circ}04' \sim 46^{\circ}40'$  之间。哈尔滨市总面积为  $5.31 \times 10^4 \text{ km}^2$ ,其中,建成区面积为  $302.41 \text{ km}^2$ ;截至 2019 年,全市常住人口为 1 076.3 万人,城镇人口 709.3 万人,城镇化率 65.9%。哈尔滨市东部与牡丹江市相连,南部与长春市、吉林市接壤,西部

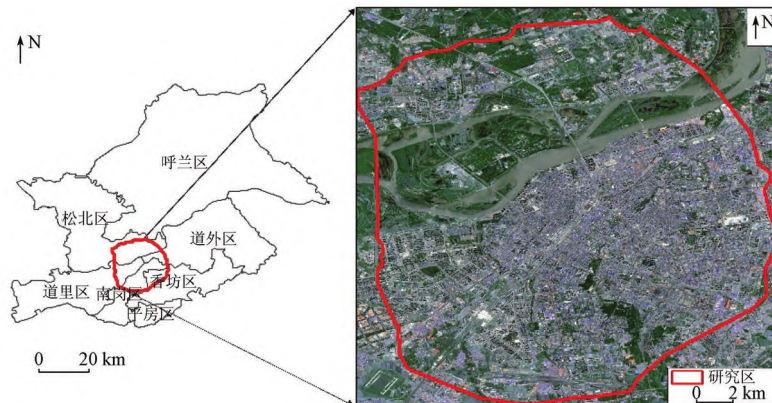


图 1 研究区位置示意图

(Sentinel-2A 真彩色合成影像)

Fig.1 Location of the study area

与绥化市、大庆市相连,北部与伊春市接壤。哈尔滨市不仅是黑龙江省省会城市,同时也是东北亚中心地带。本文选取哈尔滨市环城高速路内为研究区。

### 1.2 数据源

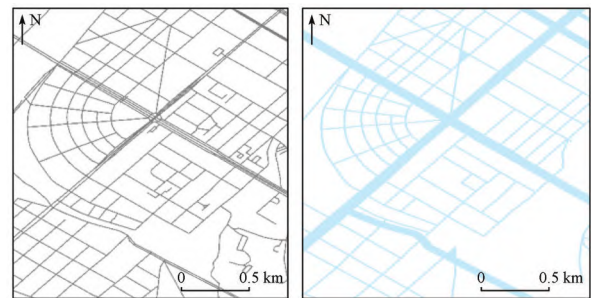
#### 1.2.1 Sentinel-2A 遥感数据

研究所需的影像数据从 Google Earth Engine 平台下载(<https://earthengine.google.com/>),采集的数据为 2019 年 6 月 18 日的 Sentinel-2A 遥感卫星影像。Sentinel-2A 有 13 个波段,幅宽达 290 km,重访周期为 10 d。本研究获取该数据空间分辨率为 10 m 的红光、绿光、蓝光、近红外 4 个波段。影像无云,质量较好,可以满足研究需求。

#### 1.2.2 OSM 路网数据

路网数据在 OpenStreetMap (OSM) 网站上获取(<https://www.openstreetmap.org>),OSM 是目前最大的协作和公开许可的地理空间数据集合,被广泛运用于权威数据的替代或补充<sup>[22]</sup>。研究使用的数据是 2019 年哈尔滨城市道路中心线数据。根据路网数据的属性信息,提取哈尔滨市的高速公路、一级公路、二级公路及居民道路;其次将路网数据根据等级建立缓冲区,根据哈尔滨市车道数量和车道宽度从遥感影像中选取道路样本进行采样测距取得平均值,将缓冲区宽度分别设置为 24 m,42 m,25 m 和

13 m;最后利用缓冲区数据将遥感影像数据进行切割(图 2)。



(a) OSM 路网数据

(b) 路网缓冲区数据

图 2 路网数据生成缓冲区

Fig.2 Convert road network data into buffer

#### 1.2.3 采样地块数据

采样地块数据来源于 Google Earth 和实地调查采样,作为训练样本及验证样本。根据城市基本土地利用制图采样方案<sup>[11]</sup>,将城市土地利用类型分为 5 个一级地类;10 个二级地类(表 1)。在 Google Earth 中选取 318 个训练样本,涵盖了所有二级土地利用类型,选取的地块具有典型性和稳定性,并且土地利用混合性低。哈尔滨市实地选取 66 个样本,包括地理位置、一级类别、二级类别、建筑物名称、混合

比例并拍摄现场照片作为验证样本。

表 1 采样地块类型及数量  
Tab.1 Number and type of sampling plots

一级地类用地类型	二级地类用地类型	二级地类编码	数量/个	
			训练样本	验证样本
住宅用地	住宅用地	0101	37	7
商服用地	商务办公用地	0201	30	5
	商业用地	0202	32	6
工矿仓储用地	工矿仓储用地	0301	30	7
交通运输用地	交通运输用地	0401	31	6
公共管理与公共服务用地	机关团体用地	0501	30	7
	教育用地	0502	35	8
	医疗卫生用地	0503	32	6
	体育和文化设施用地	0504	30	5
	公园与绿地	0505	31	8
总计			318	66

1.2.4 POI 数据

POI 数据来源于百度地图,通过爬虫获取研究区 2019 年 6 月的所有数据。POI 数据基本涵盖城市地理的全部要素信息,其内容丰富且精度较高。POI 数据包含的土地利用类型较为齐全,共有 13 个类型,18.2 万条数据,为百度地图分类系统。经过对 POI 数据进行清洗和纠偏<sup>[23]</sup>,将 POI 数据依据上述分类系统重新分类(表 2)。

表 2 POI 城市基本土地利用分类体系

POI 类型	二级地类用地类型	数量/个	比例/%
商务住宅	住宅用地	634	13.44
金融服务	商务办公用地	508	10.77
餐饮、购物、住宿服务	商业用地	548	11.62
工厂、机械电子、冶金化工	工矿仓储用地	432	9.16
车站、机场	交通运输用地	385	8.16
政府办公	机关团体用地	459	9.73
科教服务	教育用地	425	9.01
医疗保健用地	医疗卫生用地	473	10.03
体育、文化服务	体育和文化设施用地	385	8.16
风景	公园与绿地	468	9.92

1.2.5 夜间灯光数据

夜间灯光数据来源于珞珈一号(<http://59.175.109.173:8888/app/login.html>)。夜间光照强度与城市用地类型有一定相关性,夜间灯光数据可以用来描述区域人口、城市化和经济发展状况<sup>[24-25]</sup>。人类活动强度越大,夜间光照强度就越大。本研究选择 2018 年 9 月 8 日的哈尔滨市夜间灯光数据作为人类活动指标,空间分辨率为 130 m,是目前可获取的最新数据。由于 2018—2019 年间

城市土地利用变化不大,因此选择此数据。对夜间灯光数据进行特征提取并进行重采样至 10 m,参与哈尔滨市土地利用分类。

2 研究方法

图 3 为研究中使用的主要方法:①用 OSM 获取的路网数据生成城市地块;②对 Sentinel-2A 数据、POI 数据和夜间灯光数据进行特征提取;③通过 Google Earth 和实地测量采集训练样本和验证样本;④进行哈尔滨市制图和分类精度评估,分析不同特征组合对分类结果的影响。

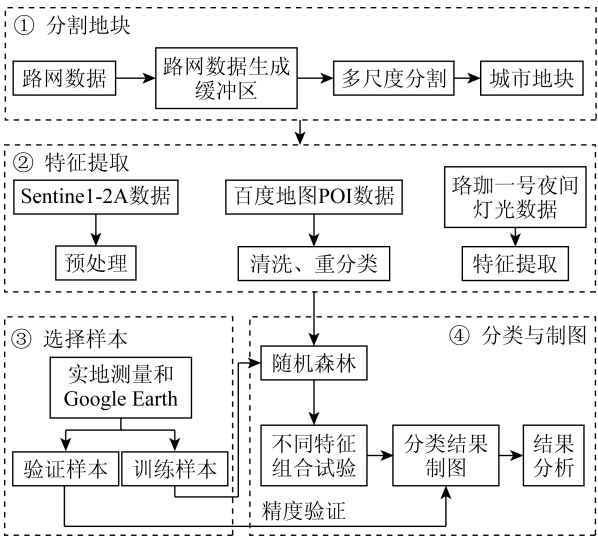


图 3 研究方法流程

Fig. 3 Flowchart of the research methods

2.1 多尺度分割方法

遥感影像多尺度分割的结果直接影响用地分类结果的好坏<sup>[26-27]</sup>。因此,确定最优分割尺度对城市土地利用类型的确定是至关重要的。对于一种特定的土地利用类型,最优分割尺度应能够较好地显示地物的边界,不能出现过于破碎或模糊的现象<sup>[28]</sup>。Woodcock 等<sup>[29]</sup>最先利用布局方差对影像进行多尺度分割,得到的结果精度较好。因此,选择局部方差法对研究区进行最优尺度分割。计算公式为:

$$v_{ar} = \frac{1}{N} \sum_{i=1}^N (x - \bar{x})^2, \quad (1)$$

式中:  $v_{ar}$  为遥感影像的局部方差;  $x$  为影像对象的灰度值;  $\bar{x}$  为影像对象灰度值的平均值。经过计算,最终确定分割尺度为 180,形状因子为 0.5,紧致度因子为 0.8。对影像进行分割,多尺度分割效果如图 4 所示。



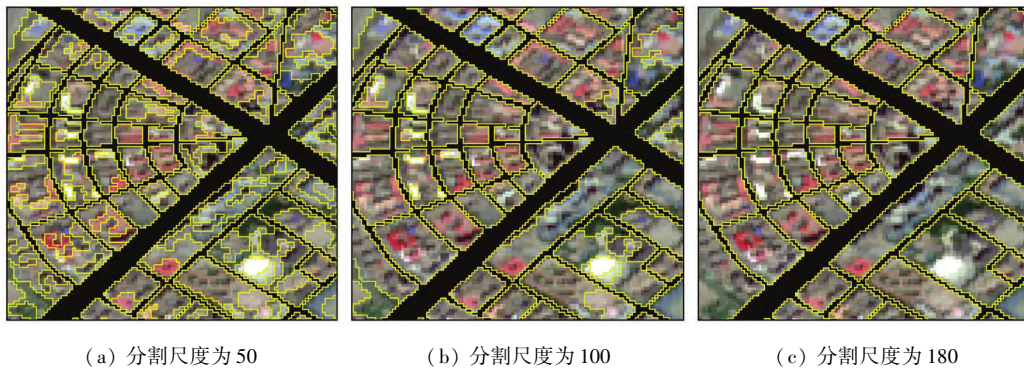


图 4 多尺度分割结果

Fig. 4 Results of multi-segmentation

## 2.2 特征提取

表 3 是对数据提取的特征,分别对光谱、纹理、POI 数据、珞珈一号夜间灯光数据进行特征提取。根据 Sentinel-2A 波段和典型光谱指数计算光谱特征;通过灰度共生矩阵计算纹理特征;计算 POI 总数、二级地类 POI 个数及比例;计算夜间灯光数据的像元亮度 DN 平均值和 DN 总和。

表 3 地块特征汇总

Tab. 3 Summary of features

特征信息	特征系数
光谱	红光、绿光、蓝光、近红外波段的平均值、标准差、标准偏差、均方差、纹理平均值、NDVI 和 NDBI
纹理	红光、绿光、蓝光、近红外波段的均值、标准差、同质性、异质性、熵
POI 数据	POI 总数、二级地类 POI 个数、不同二级地类 POI 比例
夜间灯光数据	DN 平均值和 DN 总和

## 2.3 随机森林分类方法

随机森林是 2001 年由 Breiman 提出的一种机器学习算法,结合 Bagging 集成学习理论和随机子空间的方法,由多颗决策树分类器构成<sup>[30]</sup>。随机森林的训练流程如图 5 所示。首先,通过随机重采样的方式从原始训练样本中有放回的随机抽取  $N$  个训练集,每个训练集数量约为原始数据集的  $2/3$ ;其次,将每个训练集建立分类决策树,产生  $N$  棵 CART 决策树组成的森林。根据基尼系数最小原则选出最优内部节点分支;最后,采用集成学习理论将  $N$  棵

决策树的预测结果进行统计,通过投票结果判定未分类样本的类别,分类结果由多数决策树投票结果决定。在抽取样本时约有  $1/3$  的数据没有抽中,利用这  $1/3$  的数据进行内部误差估计,产生袋外(out-of-bag)数据误差。

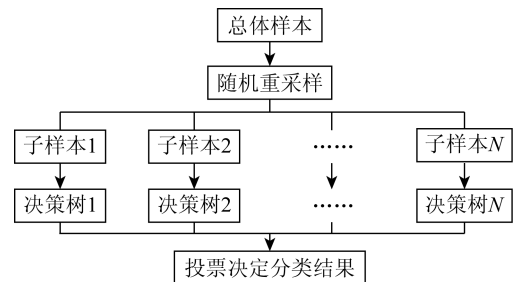


图 5 随机森林原理流程

Fig. 5 Random forest principle flow chart

本研究以地块为分类单元,随机森林算法通过 eCognition Developer 9 平台实现。通过实验得出,当  $N \geq 100$  时,各个地物类型的袋外误差趋于稳定。

## 2.4 精度评价指标

利用验证样本对随机森林分类精度进行验证,采用混淆矩阵对城市土地利用分类精度进行评价。评价指标包括生产者精度、用户精度、总体精度及 Kappa 系数,表达式分别为:

$$PA = P_{jj} / P_{+j}, \quad (2)$$

$$UA = P_{ii} / P_{i+}, \quad (3)$$

$$OA = (P_{k1} + P_{k2} + \dots + P_{kn}) / P, \quad (4)$$

$$Kappa = P \sum_{i=1}^n P_{ii} - \sum_{i=1}^n (P_i + P_{+i}) / P^2 - \sum_{i=1}^n (P_i + P_{+i}), \quad (5)$$

式中:  $PA$  为生产者精度;  $P_{jj}$  为类型中第  $j$  类和实测数据类型第  $j$  类所占的组成成分;  $P_{+j}$  为实际观测的第  $j$  类的总和;  $UA$  为用户精度;  $P_{ii}$  为类型中第  $i$  类和实测数据类型第  $i$  类所占的组成成分;  $P_{i+}$  为分类得到的第  $i$  类总和;  $OA$  为总体精度;  $P_{k1}, P_{k2}, \dots, P_{kn}$  为实测数据类型中第  $k$  类所占的组成成分;  $P$  为总样本数;  $P_{+i}$  为实际观测的第  $i$  类的总和。

## 3 结果与分析

### 3.1 精度验证

表 4 为一级地类精度的混淆矩阵,利用总体精度和 Kappa 系数来判定地类精度。一级地类总体精度为 86.0%, Kappa 系数为 0.75; 二级地类总体精

度为 73.9% ,Kappa 系数为 0.69。其中,商业用地用户者精度为 95.1% ,公园与绿地用户者精度为 97.1% ; 机关团体用地和医疗卫生用地分类精度较低,前者用户者精度为 15.7% ,后者用户者精度为 11.1% 。

表 4 一级地类精度混淆矩阵  
Tab.4 Confusion matrix of results of land use accuracy (Level I category)

用地类型	住宅用地	商服用地	工矿仓储用地	交通运输用地	公共管理与公共服务用地	用户精度/%
住宅用地	3 573	157	1 747	0	5 092	33.8
商服用地	0	10 329	0	0	804	92.7
工矿仓储用地	0	0	10 184	0	104	98.9
交通运输用地	468	0	382	669	0	44.0
公共管理与公共服务用地	809	570	976	792	48 446	93.9
生产者精度/%	73.6	93.4	76.6	45.7	60.6	
总体精度/%			86.0			
Kappa 系数			0.75			

3.2 城市土地利用分类结果

通过 OSM 数据对哈尔滨市路网数据进行提取,利用多尺度分割将 Sentinel-2A 遥感卫星影像分割

为地块。对 POI 数据进行筛选和重分类,将夜间灯光数据进行特征提取,利用随机森林分类方法对研究区进行分类,结果如图 6 所示。

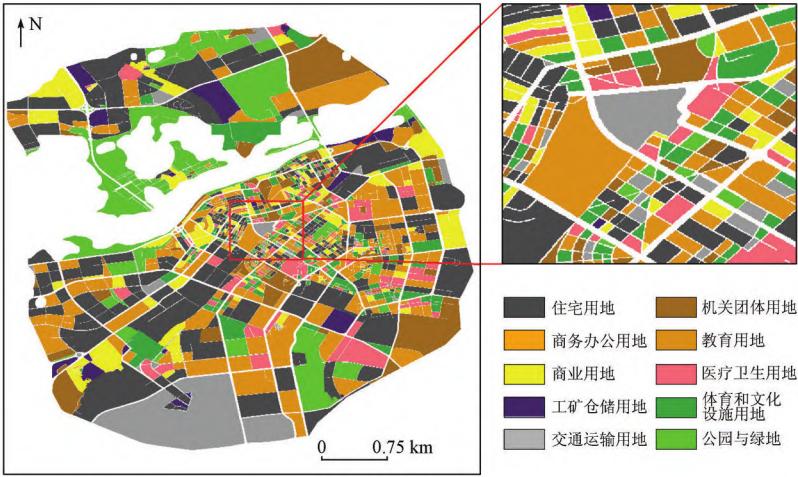


图 6 城市土地利用分类结果  
Fig.6 Urban land use classification results

3.3 不同特征组合的分类精度分析

图 7 为二级地类在不同特征组合下参与土地利用分类的生产者精度及用户精度。可以看出,不同数据组合参与土地利用分类的精度不同。当采样地块参与土地利用分类时,公园与绿地的生产者精度和用户精度最高,分别为 95.5% 和 64.6%。相比加入 POI 数据和夜间灯光数据后精度的提高并不显著。说明对公园与绿地的识别主要依赖于遥感光谱特征和纹理特征。图 7 中,与(a)相比,(b)和(c)中商务办公用地和商业用地精度显著提高,但夜间灯光数据对商服用地精度的提升幅度更大; (b)中教育用地的精度高于(a)中的精度,说明 POI 数据可以提高教育用地的精度,其生产者精度为 71.4% ,用户精度为 62.5% ; (c)中夜间灯光数据对住宅用地的分类精度略有提高,但没有(b)中提高幅度大,表明

POI 数据可以改善对住宅用地分类精度,但程度有限; (a)中工矿仓储用地精度较低,(b)中工矿仓储用地精度有显著提高,说明工矿仓储用地更适用 POI 数据参与分类。在所有试验中,机关团体用地和医疗卫生用地精度均较低,POI 数据和夜间灯光数据对机关团体用地和医疗卫生用地精度提升改善程度有限。同(a)相比,(b)中体育和文化设施用地的精度有所提升,说明 POI 数据可以提高体育和文化设施用地的精度,但是并未达到良好效果。遥感影像数据的光谱特征、纹理特征、POI 数据和夜间灯光数据可以有效提高城市土地利用分类精度。单独使用采样地块数据参与分类时,城市土地利用分类精度较低; 当使用采样地块和 POI 数据时,可以较好地提高住宅用地、工矿仓储用地和教育用地精度; 当使用采样地块和夜间灯光数据时,对商务办公用

地和商业用地精度提升较大;将全部特征参与分类时,分类结果的精度最高,即为分类的最佳组合。这表明综合使用多源特征可以改善城市土地利用分类

的精度,光谱特征、纹理特征、POI 数据和夜间灯光数据可以显著提升某些地块类型的准确性,但是综合运用可以有效提升所有地物类型的准确性。

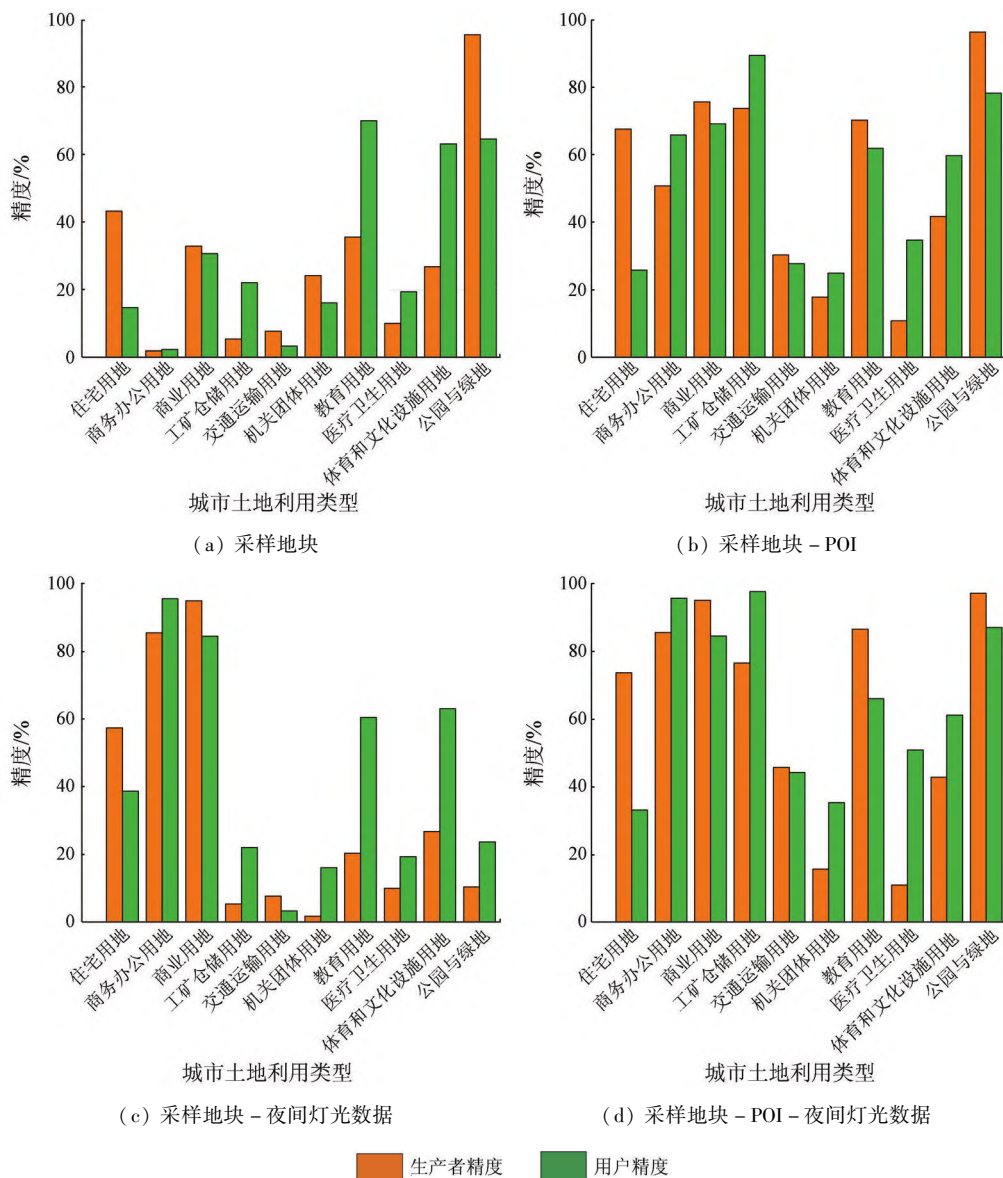


图 7 基于不同特征组合的生产者精度和用户精度(二级地类)

Fig. 7 Producer accuracy and user accuracy of different features combinations (Level II category)

## 4 结论与讨论

本文以哈尔滨市市区为研究区,结合 OSM 数据将地块划分为基本分类单元,基于 Sentinel-2A 遥感影像、POI 数据和夜间灯光数据等,采用随机森林分类方法对研究区进行城市土地信息提取,得到如下结论:

1) 一级地类总体分类精度为 86.0%, Kappa 系数为 0.75; 二级地类总体精度为 73.9%, Kappa 系数为 0.69。

2) 遥感影像、多源地理空间数据及网络数据相

结合能够有效提高分类精度,有利于提高城市土地利用分类效果。

3) POI 数据可以显著提高住宅用地、工矿仓储用地和教育用地分类精度;夜间灯光数据能够有效提高商务办公用地及商业用地分类精度。

相较于 2018 年中国基本城市土地利用分类,本文分类精度有较高的提升。文章选取训练样本的混合度更低,提高了训练样本的纯度和数量;利用更详细的路网数据对地块进行分割,并且选取训练样本数据时尽可能在空间分布上均匀分布。尽管上述方法对于研究区能够得到较好的分类结果,同时也存在局限性。城市土地内部结构复杂多样,实际土



地利用大多为混合地块,无法精确保证每个地块均为单一类型。同时城市的快速发展使城市三维空间利用率增加,一幢建筑中不同高度的建筑用地不是单一用地类型,仅用路网数据很难对其进行精确有效的划分。POI数据和夜间灯光数据可以有效提高分类精度,但是并不适合所有用地类型,造成城市土地利用分类精度差异性。在对城市土地利用类型进行分类过程中,应更进一步对地块进行精准划分,路网数据可以有效对地块进行划分,但是并不能完全将不同类型的地块划分出来。要进一步优化地块分割的过程,尽量减少有多种土地利用类型的地块,根据不同土地利用类型的特征使用不同尺度参数的分割。此外,对于具有多种利用类型的地块,根据不同高度或面积来有针对性地权衡POI数据对分类结果的影响,尽可能将不同土地利用类型进行细化区分。

### 参考文献 (References):

- [1] Ma Q, Gong Z, Kang J, et al. Measuring functional urban shrinkage with multi-source geospatial big data: A case study of the Beijing-Tianjin-Hebei megaregion[J]. Remote Sensing, 2020, 12(16): 2513.
- [2] Mao W, Lu D, Hou L, et al. Comparison of machine-learning methods for urban land-use mapping in Hangzhou City, China[J]. Remote Sensing, 2020, 12(17): 2817.
- [3] Chang S, Wang Z, Mao D, et al. Mapping the essential urban land use in Changchun by applying random forest and multi-source geospatial data[J]. Remote Sensing, 2020, 12(15): 2488.
- [4] Zhang C, Sargent I, Pan X, et al. An object-based convolutional neural network (OCNN) for urban land use classification[J]. Remote Sensing of Environment, 2018, 216: 57-70.
- [5] Zhang C, Sargent I, Pan X, et al. Joint deep learning for land cover and land use classification[J]. Remote Sensing of Environment, 2019, 221: 173-187.
- [6] Hu T, Yang J, Li X et al. Mapping urban land use by using Landsat images and open social data[J]. Remote Sensing, 2016, 8(2): 151.
- [7] Su M, Guo R, Chen B, et al. Sampling strategy for detailed urban land use classification: A systematic analysis in Shenzhen[J]. Remote Sensing, 2020, 12(9): 1497.
- [8] Sun J, Wang H, Song Z, et al. Mapping essential urban land use categories in Nanjing by integrating multi-source big data[J]. Remote Sensing, 2020, 12(15): 2386.
- [9] He S, Chen X, Zhang Z, et al. The exploration of urban material anabolism based on RS and GIS methods: A case study in Jin-chang, China[J]. Remote Sensing, 2020, 12(3): 370.
- [10] Wang R, Hou H, Murayama Y, et al. Spatiotemporal analysis of land use/cover patterns and their relationship with land surface temperature in Nanjing, China[J]. Remote Sensing, 2020, 12(3): 440.
- [11] Gong P, Chen B, Li X. Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018[J]. Science Bulletin, 2020, 65(3): 182-187.
- [12] Zong L, He S, Lian J, et al. Detailed mapping of urban land use based on multi-source data: A case study of Lanzhou[J]. Remote Sensing, 2020, 12(12): 1987.
- [13] Zeferino L B, Souza L F T D, Amaral C H D, et al. Does environmental data increase the accuracy of land use and land cover classification?[J]. International Journal of Applied Earth Observation and Geoinformation, 2020, 91: 102-128.
- [14] Saboori M, Asghar T A, Reza R B H. Combining multi-scale textural features from the panchromatic bands of high spatial resolution images with ANN and MLC classification algorithms to extract urban land uses[J]. International Journal of Remote Sensing, 2019, 40(22): 8608-8634.
- [15] Tu Y, Chen B, Zhang T, et al. Regional mapping of essential urban land use categories in China: A segmentation-based approach[J]. Remote Sensing, 2020, 12(7): 1058.
- [16] Zhang Y, Qin K, Bi Q, et al. Landscape patterns and building functions for urban land-use classification from remote sensing images at the block level: A case study of Wuchang district, Wuhan, China[J]. Remote Sensing, 2020, 12(11): 1831.
- [17] 周天宁, 明冬萍, 赵睿. 参数优化随机森林算法的土地覆盖分类[J]. 测绘科学, 2017, 42(2): 88-94.
- [18] 马玥, 姜琦刚, 孟志国. 基于随机森林算法的农耕地土地利用分类研究[J]. 农业机械学报, 2016, 47(1): 297-303.
- [19] 杨耘, 徐丽, 颜佩丽. 条件随机场框架下基于随机森林的城市土地利用/覆盖遥感分类[J]. 国土资源遥感, 2014, 26(4): 51-55. doi:10.6046/gtzyyg. 2014. 04. 09.
- [20] 薛冰, 肖骁, 李京忠. 基于兴趣点(POI)大数据的东北城市空间结构分析[J]. 地理科学, 2020, 40(5): 691-700.
- [21] 李欣. 基于POI要素空间聚集特征的城市多中心结构识别——以郑州市为例[J]. 北京大学学报(自然科学版), 2020, 56(4): 692-702.
- [22] Brovelli M, Zamboni G. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints[J]. ISPRS International Journal of Geo-Information, 2018, 7(8): 289.
- [23] 郑纪玲. 数据清洗在构建POI数据库中的研究与应用[D].

- 北京:中国矿业大学(北京),2016.
- Zheng J L. Research and application of data cleaning in the construction of POI data warehouse [D]. Beijing: China University of Mining and Technology (Beijing), 2016.
- [24] 陈颖彪,郑子豪,吴志峰,等. 夜间灯光遥感数据应用综述和展望[J]. 地理科学进展, 2019, 38(2): 205–223.
- Chen Y B, Zheng Z H, Wu Z F, et al. Summary and prospect of the application of nighttime light remote sensing data [J]. Progress in Geography, 2019, 38(2): 205–223.
- [25] 李黔湘,郑 晗,张 过,等. 基于夜间灯光影像的珞珈一号指数分析[J]. 测绘地理信息, 2020, 45(3): 8–15.
- Li Q X, Zheng H, Zhang G, et al. Analysis of LuoJia-1 index based on nighttime light image [J]. Geoinformation of Surveying and Mapping, 2020, 45(3): 8–15.
- [26] 林先成,李永树. 成都平原高分辨率遥感影像分割尺度研究[J]. 国土资源遥感, 2010, 22(2): 7–11. doi: 10. 6046/gtzyyg. 2010. 02. 02.
- Lin X C, Li Y S. Research on the segmentation scale of high-resolution remote sensing image of Chengdu Plain [J]. Remote Sensing for Land and Resources, 2010, 22(2): 7–11. doi: 10. 6046/gtzyyg. 2010. 02. 02.
- [27] 王 露,刘庆元. 高分辨率遥感影像多尺度分割中最优尺度选取方法综述[J]. 测绘与空间地理信息, 2015, 38(3): 166–169.
- Wang L, Liu Q Y. A review of optimal scale selection methods in multi-scale segmentation of high-resolution remote sensing images [J]. Surveying and Spatial Information, 2015, 38(3): 166–169.
- [28] 黄欢欢,黄万里. 基于面向对象分类方法的的城市用地类型提取[J]. 福建师范大学学报(自然科学版), 2014, 30(6): 96–102.
- Huang H H, Huang W L. Urban land type extraction based on object-oriented classification method [J]. Journal of Fujian Normal University (Natural Science Edition), 2014, 30(6): 96–102.
- [29] Woodcock C, Eand Strahler A H. The factor of scale in remote sensing [J]. Remote Sensing of Environment, 1987, 21(3): 311–332.
- [30] 李 治,杨晓梅,孟 樊,等. 物候特征辅助下的随机森林宏观尺度土地覆盖分类方法研究[J]. 遥感信息, 2013, 28(6): 48–55.
- Li Z, Yang X M, Meng F, et al. Random forest macro-scale land cover classification method aided by phenological characteristics [J]. Remote Sensing Information, 2013, 28(6): 48–55.

## Urban land use classification based on remote sensing and multi-source geographic data

WU Linlin<sup>1,2</sup>, LI Xiaoyan<sup>1</sup>, MAO Dehua<sup>2</sup>, WANG Zongming<sup>2</sup>

(1. College of Earth Sciences, Jilin University, Changchun 130012, China; 2. Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China)

**Abstract:** Urban land use (ULU) reflects urban functions and structures, and the study of ULU classification can provide guidance for the sustainable development of cities. This study conducted the ULU classification of the main urban area of Harbin City using the object-oriented and random forest methods by integrating multi-source geospatial data including Sentinel-2A remote sensing images, OpenStreetMap (OSM) data, point of interest (POI) data, and nighttime light data from the LuoJia-1 satellite. The results are as follows. The overall accuracy of the first-level land use type was 86.0%, with a Kappa coefficient of 0.75. The overall accuracy of the second-level land use types was 73.9%, with a Kappa coefficient of 0.69. The introduction of POI data can significantly improve the classification accuracy of residential land, industrial and mining storage land, and educational land. Meanwhile, night light data can effectively improve the classification accuracy of commercial office land and business land. This study shows that the combination of remote sensing images with multi-source geographic data is effective for ULU classification.

**Keywords:** random forest; multi-scale segmentation; multi-source data; urban land use classification; Harbin (责任编辑: 李 瑜)