

# OBIA 与 RF 结合的龙口市土地利用信息提取方法

王瑗玲, 张校千, 苏晨晨, 于新洋

(山东农业大学资源与环境学院, 泰安 271018)

**摘要:** 为提高中分辨率遥感影像解译精度, 本文提出面向对象影像分析(Object Based Image Analysis, OBIA)与随机森林(Random Forest, RF)结合的土地利用信息提取方法。采用 Landsat 8 OLI 影像, 针对不同地物特点, 阈值分割和多尺度分割结合创建影像对象, 规则集和分类器协同分类, 基于 Relief F 算法分别对光谱特征、纹理特征及所有特征降维筛选特征子集, 并与全部特征一起应用 RF 建模, 对龙口市进行土地利用信息提取与比较。结果表明: OBIA 与 RF 结合提取土地利用信息, 基于 Relief F 算法筛选纹理特征, 保留完整光谱、几何、空间关系特征构建 RF 模型, 建模错分率为 0.0958, 分类总体精度和 Kappa 系数分别为 89.37% 和 0.872, 取得较理想结果。该方法可应用于中分辨率遥感影像土地利用信息提取。

**关键词:** 土地利用信息; 提取方法; 面向对象; Relief F 降维; 随机森林; 龙口市

LUCC (Land Use and Cover Change) 因其对全球社会经济生态的重要影响成为国内外全球变化学术界研究的热点<sup>[1-2]</sup>。而以 RS 和 GIS 为数据源和分析技术的土地利用/土地覆被信息提取方法, 同样得到了学术界广泛的关注, 取得了较多研究成果<sup>[3]</sup>。当今, 经济社会快速发展, 土地利用类型变化复杂, 土地利用变化速度加快, 基于像元的传统遥感影像分类方法已远不能满足社会需求。有学者基于知识与规则<sup>[4]</sup>、面向对象<sup>[5]</sup>和数理统计模型<sup>[6]</sup>等进行高效率、高精度遥感影像分析提取, 取得了较满意的效果。中等分辨率 Landsat TM、ETM+、OLI 影像的免费获取和使用, 使之成为土地利用/土地覆被的重要数据来源。因此, 进一步探索基于中等分辨率遥感影像及多种方法结合的土地利用信息提取方法, 提高分类精度, 成为 LUCC 研究的重要课题, 这对探明土地利用及其动态演变、科学配置和合理利用土地资源具有重要意义。

随机森林<sup>[7]</sup> (Random Forest, RF) 是一种多分类器集成算法, 被创新应用于遥感图像处理。该算法集成了单一分类器的优点, 提高了复杂环境下分类的效率和适宜性, 取得了较好的分类效果。Ghimire 等<sup>[8]</sup>采用 RF 算法进行土地利用/土地覆被分类, 并与支持向量机 (SVM)、决策回归树 (CART)、Bagging、Boosting、最大似然法 (MLC) 等比较。孙雪莲等<sup>[9]</sup>、王书玉等<sup>[10]</sup>基于像元, 应用 RF 算法分别对林地、湿地进行解译。上述研究均表现出 RF 应用于遥感图像解译的适配性。面向对象的图像处理方法综合运用地物光谱、几何、纹理和空间关系特征, 但随着特征维数的增加, 信息冗余、计算量大、模型运行效率低等弊端随之而来, 且可能导致分类精度降低, 因此需要筛选出有效识别分类目标的特征子集<sup>[11]</sup>。Relief F 算法凭借其效率高、数据类型限制少等优点, 广泛应用于

收稿日期: 2018-08-22; 修订日期: 2019-01-17

基金项目: 山东省重点研发计划项目 (2017CXGC0308); 山东省博士后创新项目 (222016)

作者简介: 王瑗玲 (1972-), 女, 山东临沂人, 博士, 教授, 研究方向为土地利用与评价。

E-mail: ailingwang@sda.edu.cn

数据处理中<sup>[12]</sup>。有学者采用Relief F算法筛选特征应用于面向对象土地利用分类中,取得了较好效果<sup>[13-14]</sup>。因Landsat 8 OLI影像受限于空间分辨率大小,在一定信噪比前提下,相较于其他特征,其光谱特征对地物成像较为重要<sup>[15]</sup>,而纹理特征基于像元光谱值计算得来<sup>[16]</sup>,二者冗余度高,易造成“维数灾难”<sup>[19]</sup>。但上述研究均是对所选光谱、几何、纹理和空间关系全部特征子集降维,而分别对光谱、纹理单一类别特征进行降维,结合其他特征应用到随机森林模型中,其解译精度如何,哪类特征子集建模、解译效果最好,是否低于全部特征建模、解译效果,尚未有研究报导。

基于此,本文以山东省龙口市为研究区,Landsat 8 OLI影像为数据源,面向对象和随机森林算法结合提取土地利用信息,基于Relief F算法对影像光谱特征、纹理特征和所选全部特征分别降维,采用OOB(Out of Bag)袋外误差进行参数寻优和建模精度比较,分析特征选取与模型构建对解译精度的影响,评估RF提取土地利用信息的性能与效果,为高效、高精度提取土地利用信息提供技术支持,并为土地合理利用和科学规划提供决策支持。

## 1 研究方法与数据来源

### 1.1 研究区概况

龙口市地处山东东北部、胶东半岛西北部、渤海湾南畔,介于 $120^{\circ}13' \sim 120^{\circ}44'E$ 、 $37^{\circ}27' \sim 37^{\circ}47'N$ 间,地理位置优越。地势东南高、西北低,呈台阶式下降,主要有山地、丘陵、平原三种地貌类型。属暖温带季风性气候,年平均气温 $11.8^{\circ}C$ ,年平均降水量608 mm。研究区下辖8个镇、5个街道办事处、1个省级经济开发区、1个省级高新技术产业园区、1个滨海旅游度假区,总面积 $901.05\text{ km}^2$ ,总人口69.69万人。改革开放以来,龙口市经济发展迅猛,多年入选国家县域经济与县域基本竞争力百强县、山东第一强县。随着经济社会发展以及城市化进程加快,龙口市土地利用变化剧烈,对社会经济可持续发展和生态环境产生重要影响<sup>[17]</sup>。

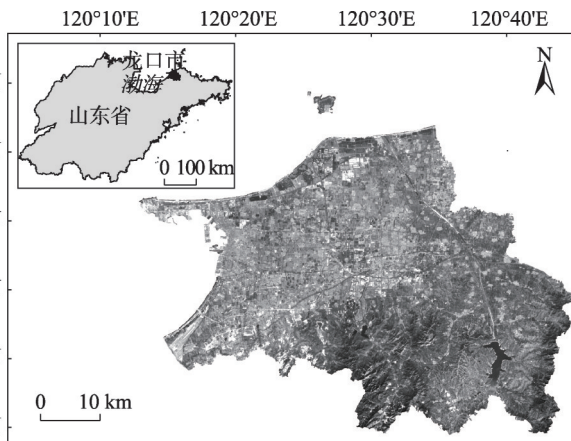


图1 研究区位置图(Landsat 8 OLI红波段)

Fig. 1 Location of the study area (The red band of Landsat 8 OLI)

### 1.2 数据来源及预处理

本文以2014年10月27日Landsat 8 OLI影像为数据源,其行号为120,列号为34,空间分辨率为30 m,通过USGS (<https://glovis.usgs.gov>)下载获得。收集龙口市2014年1:5万土地利用现状图、1:5000正射影像图及Google Earth、天地图等影像资料。

在ArcGIS和ENVI支持下,对影像进行辐射校正和FLAASH大气校正,分别选择高斯—克吕格和西安80为投影和坐标系统,以1:5000正射影像图为准进行几何精校正,均方根误差RMSE为0.32。利用矢量化后的研究区行政边界做掩膜对影像进行不规则裁剪,获得龙口市遥感影像图。

### 1.3 OBIA与RF结合的土地利用信息提取

面向对象影像分析(Object-Based Image Analysis, OBIA)是以分割后产生的同质性像元集合为最小单元进行影像分析的方法。相对于传统的基于像元的影像分析,该方法充分利用光谱特征、几何特征、纹理特征和空间拓扑关系增加不相关的特征空间<sup>[13,18]</sup>,有效减少了“同物异谱”和“异物同谱”对分类结果的影响,一定程度上避免了“椒盐现象”的产生,提高了分类精度<sup>[19]</sup>。本文基于面向对象,采用“规则集+分类器”的分类方法,充分发挥二者优势,对龙口市土地利用信息进行提取。主要步骤有影像对象创建、特征变量选取、随机森林模型构建、随机森林参数寻优和土地利用信息提取,如图2。

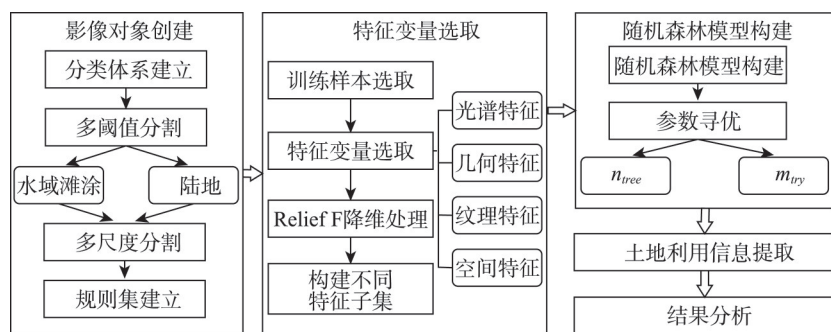


图2 技术路线

Fig. 2 Technical workflow

结合现行土地利用分类体系及研究区土地利用特点,参照土地利用现状图、Google earth、天地图等资料,将研究区土地利用分为11类:耕地、园地、林地、草地、城乡建设用地、交通用地、河流、水库水面、坑塘水面、滩涂、未利用地及其他。

#### 1.3.1 影像对象创建

影像分割是按照一定标准将均质性像元归为一个集合,是面向对象影像分析的首步,也是关键环节,其分割效果的好坏直接关系后续影像分析。在eCognition 9.0支持下,结合影像特点,阈值和多尺度分割结合创建影像对象。首先构建改进后的归一化水体指数(MNDWI),采用阈值分割,区分水域滩涂和陆地,然后对其分别进行多尺度分割,不断调试相应参数,得到最优尺度。MNDWI计算如式(1):

$$MNDWI = \frac{G - SWIR 1}{G + SWIR 1} \quad (1)$$

式中: MNDWI为改进后的归一化水体指数; G为绿波段; SWIR 1为短波红外1。

以水域滩涂为父类进行多尺度分割,经反复调试,设置其分割尺度为50,形状指数为0.6,紧致度指数为0.5,分割波段权重为Coastal: R: G: B: NIR: SWIR 1: SWIR 2: PAN: MNDWI=2: 1: 1: 1: 1: 2: 2: 1: 2: 3,得到最优分割尺度。参照2014年土地利用现状图、天地图中各类水面空间特征,构建规则集依次提取河流、水库、坑塘水面及滩涂。具体规则为: Lenth/Width>6.3为河流, Area>970为水库,其余水域为坑塘水面, MNDWI>0.41为沿海滩涂,其次根据空间特征Rel. Border to提取内陆滩涂。

以陆地为父类进行多尺度分割,结合地物特征,经反复调试后,设置其分割尺度为25,形状指数为0.3,紧致度指数为0.5,加载缨帽变换后亮度(Brightness)、绿度

(Greenness)、湿度(Wetness)参与分割,分割波段权重为Coastal:R:G:B:NIR:SWIR 1:SWIR 2: PAN:MNDWI:NDVI: Brightness: Greenness: Wetness=0:1:1:1:2:2:1:2:3:3:2:2:1,得到最优分割尺度,构建影像对象。该尺度下分割对象形状规整,边界清晰,可明显区分,满足研究需要。

### 1.3.2 特征变量选取

以陆地为父类,参照土地利用现状图选取样本,并选择光谱、几何、纹理、空间特征4类变量构建RF模型。参照土地利用现状图、天地图、Google Earth等高分辨率影像,选取训练样本和验证样本,训练样本用于特征选择和随机森林建模,验证样本用于精度评价。其中,训练样本总像元为192913个,验证样本总像元为93173个,各地类样点所占像元数量见表1。

表1 各地类样点像元数量

Table 1 The number of samples for each land use type

(单位:个)

地类	城乡建设用地	耕地	林地	草地	园地	交通用地	未利用地及其他	总计
训练样本	33118	30717	57932	16826	47798	4089	12567	192913
验证样本	16488	15366	28949	8412	22291	2032	6293	93173

Relief F (Relevant Feature) 算法是过滤式特征选择方法,对其改进后 Relief F 算法可用于处理目标属性为连续值的回归问题。目前 Relief F 算法在特征变量降维处理方面得到了国内外学者的广泛应用<sup>[20]</sup>。其原理是每次随机从样本集中抽取一个样本记为  $R$ , 分别与其同类和异类样本集中找出  $R$  的  $k$  个近邻样本, 分别记为  $H$ 、 $M$ , 通过类间距与类内距不断更新其权重<sup>[6]</sup>。假设对于任意特征  $a$  被随机抽取  $b$  次, 其权重为  $W(a)$ , 权重越高, 其相关性越大。计算公式为:

$$W(a) = W(a) - \sum_{j=1}^k \frac{\text{diff}(a, R, H_j)}{bk} + \sum_{C \neq \text{class}(R)} \left[ \frac{p(C)}{1 - p(\text{Class}(R))} \sum_{j=1}^k \text{diff}(a, R, M_j(C)) \right] / bk \quad (2)$$

$$\text{diff}(a, R, H) = \begin{cases} \frac{|R[a] - H[a]|}{\max(a) - \min(a)} & \text{if } a \text{ is continuous} \\ 0 & \text{if } A \text{ is discrete and } R[a] = H[a] \\ 1 & \text{if } A \text{ is discrete and } R[a] \neq H[a] \end{cases} \quad (3)$$

式中:  $C$  为异类样本类别;  $\text{diff}(a, R, H_j)$  表示样本  $R$ 、 $H_j$  在特征  $a$  上的差值;  $M_j(C)$  为  $C$  类中第  $j$  个最近邻样本;  $p(\text{Class}(R))$  为样本类别与  $R$  相同的概率;  $p(C)$  为异类样本为  $C$  的概率。

初选光谱特征(SPE)、几何特征(GEO)、纹理特征(TEX)、空间特征(SPA)<sup>[13-14]</sup>共94个特征变量,分别对其编号(表2)。

本文采用 Relief F 算法进行降维、筛选特征变量。基于 MATLAB R2016a 软件分别对光谱特征、纹理特征及全部特征进行 Relief F 算法降维、建立特征子集,分别记为 SPE\_RF、TEX\_RF、ALL\_RF,将未做降维处理的全部特征集合记为 ALL。保留与目标类重要性程度较大的特征参与随机森林建模,优选3类特征子集中前8个、9个和32个,即 SPE\_RF (SPE 1、3~5、7、8、12、17、23+GEO 1~8+TEX 1~60+SPA 1~3)、TEX\_RF (SPE 1~23+GEO 1~8+TEX 7、9、29、34、37、39、44、54、55+SPA 1~3)、ALL\_RF



表2 特征变量及编号

Table 2 Feature variables and their number

特征类别	特征变量及编号	合计/个
光谱特征	1~9: 2~6波段均值、 <i>MNDWI</i> 、 <i>NDVI</i> 、光谱最大差分 Max. Diff 10~20: 2~6波段方差、 <i>MNDWI</i> 方差、 <i>NDVI</i> 方差、TC 1~TC 3 方差 21~23: TC 1~TC 3 均值: <i>Brightness</i> 、 <i>Wetness</i> 、 <i>Greenness</i>	23
几何特征	1~8: 面积、边界指数、紧致度、长宽比、对称性、密度、矩形拟合度、形状指数	8
纹理特征	1~40: <i>GLCM</i> Ang. 2nd moment、 <i>GLCM</i> Mean、 <i>GLCM</i> Dissimilarity、 <i>GLCM</i> Contrast、 <i>GLCM</i> Homogeneity、 <i>GLCM</i> Entropy、 <i>GLCM</i> Correlation、 <i>GLCM</i> StdDev (all dir.、0°、45°、90°、135°) 41~60: <i>GLDV</i> Ang. 2nd moment、 <i>GLDV</i> Contrast、 <i>GLDV</i> Entropy、 <i>GLDV</i> Mean (all dir.、0°、45°、90°、135°)	60
空间特征	1~3: 5、6 波段和 <i>NDVI</i> 对象邻域的平均差分 (绝对值)	3

注: *GLCM* 为灰度共生矩阵、*GLDV* 为灰度差分矢量、And. 2nd moment 为二阶矩、Mean 为均值、Homogeneity 为同质性、Dissimilarity 为差异性、Entropy 为熵、Correlation 为相关性、Contrast 为对比度、Entropy 为熵、StdDev 为方差。

(SPE 1~23+TEX 6~10、44+SPA 1~3)。三个特征子集的特征个数分别为 80 个、44 个、32 个。

### 1.3.3 随机森林模型构建

随机森林是由多棵决策树构成的分类器集成算法<sup>[7]</sup>, 该方法可高效处理大样本数据而不必担心过度拟合。其原理为通过自助重采样法 (Bootstrap) 从原始数据中重复随机抽取生成  $k$  个样本集, 分别生成  $k$  个分类树组成随机森林, 每个样本集大小约为原始数据集的 2/3, 每次未被抽取的数据称为袋外数据 (Out of Bag, OOB)。以单棵树的每个结点为单位, 从  $M$  个特征中随机挑选  $m$  个特征 ( $m \leq M$ ), 基于节点不纯度最小的原则挑选特征使树充分生长, 集合多棵树分类器结果进行预测, 分类结果依投票打分而定。随机森林的精度取决于用户自定义设置的参数, 即树的数量 ( $n_{tree}$ ) 和变量数量 ( $m_{try}$ ), 不同组合的  $n_{tree}$  和  $m_{try}$  对分类结果影响不同。变量数量  $m_{try}$  通常为所选特征数量的平方根 (SQRT)<sup>[7,10,12]</sup>。OOB 误差可进行泛化误差估计和变量重要性验证, OOB 误差越小, 建模精度越高。

基于 SPM 软件分别对 SPE\_RF、TEX\_RF、ALL\_RF、ALL 4 个特征子集构建随机森林模型, 分别记为 Model 1 (SPE\_RF)、Model 2 (TEX\_RF)、Model 3 (ALL\_RF)、Model 4 (ALL), 根据所选特征数量分别计算相应模型的错分率。

默认  $n_{tree}=200$  时, 计算不同数量的  $m_{try}$  与模型 OOB 错分率间关系。由图 3 可知, 4 类模型 OOB 错分率随  $m_{try}$  数量增加总体呈递减趋势。当  $m_{try}=\text{SQRT}$  时, Model 2 错分率最低, 为 0.0960; 当  $m_{try}=2 \times \text{SQRT}$  时, Model 1、Model 3 和 Model 4 错分率最低, 分别为 0.1066、0.1034 和 0.1037。

### 1.3.4 随机森林参数寻优

采用 SPM 软件对 4 类模型进行参数组合寻优, 计算当  $m_{try}$  分别为  $\text{SQRT} \times 1/8$ 、 $\text{SQRT} \times 1/4$ 、 $\text{SQRT} \times 1/2$ 、 $\text{SQRT}$  和  $\text{SQRT} \times 2$  所对应的错分率, 得到 4 类模型最优参数范围 (表 3)。模型运算结果表明, 当  $m_{try}=\text{SQRT}$  时, Model 2 错分率整体较低; 当  $m_{try}=2 \times \text{SQRT}$  时, Model 1、Model 3 和 Model 4 错分率整体较低, 与上文结论一致 (图 3)。4 类模型优选参数组合 ( $n_{tree}$ ,  $m_{try}$ ) 分别为 Model 1 (140, 18)、Model 2 (148, 7)、Model 3 (144, 11)、Model 4 (150, 19), OOB 错分率分别为 0.1075、0.0958、0.1026、0.1028。

表3 最优参数范围

Table 3 The range of the optimal parameters

参数	Model 1	Model 2	Model 3	Model 4
$n_{tree}$	(1, 2, 4, 9, 18)	(1, 2, 4, 7, 11)	(1, 2, 4, 7, 11)	(1, 2, 4, 10, 19)
$m_{try}$	(135, 140, 145)	(138, 148, 158)	(137, 144, 151)	(140, 150, 160)

1.3.5 土地利用信息提取

以 eCognition 为技术支撑，基于 Relief F 算法筛选出的4个特征子集，应用最优参数组合于 RF 模型中，以陆地为父类执行分类算法，提取龙口市2014年土地利用信息，得到分类结果（图4）。

2 结果分析

2.1 不同特征组合精度分析

根据上述参数组合执行算法，RF算法执行时间、模型 OOB 错分率、分类精度由表4所示。结果表明，只对纹理特征降维 Model 2 的错分率最小、精度最高，其次是所有特征降维 Model 3，再次是全部特征 Model 4，只对光谱降维 Model 1 精度最低。从执行时间看，Model 3 用时最少，其次是 Model 2，Model 4 用时最多。

构建 RF 模型时， $m_{try}$  为单棵树随机选取的特征数量，在一定程度上与模型效率相关。由表4可看出，Model 2 中  $m_{try}$  为7，仅筛选与目标类别相关性较高的纹理特征，相较于 Model 1 和 Model 4 分别少了11和12，虽保留完整光谱特征，但 RF 因本身具有抗噪、降维的优越性<sup>[14]</sup>，可部分消除信息冗余的影响，故建模精度和效率在四类模型中最高，执行时间为39 m 49 s，用时也相对较少；由于光谱信息是地物成像的基础，Model 1 损失部分光谱特征，但完整几何、纹理和空间特征的保留易造成冗余，因此其建模和解译精度在四类中最低。可见，计算量大、相关性高的冗余特征造成复杂运算和精度下降；与 Model 3 相比， $m_{try}$  少了4个，且保留了完整几何特征，由分类结果可看出，园地和林地的误分得到改善，道路与建设用地混淆状况得到减轻，这表明多尺度分割形成的破碎

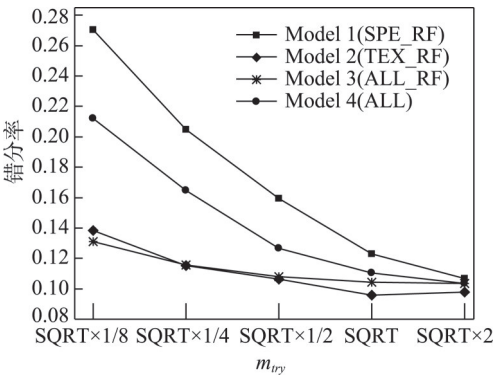


图3 特征数量与错分率关系

Fig. 3 The relationship between the number of features and the OOB misclassification

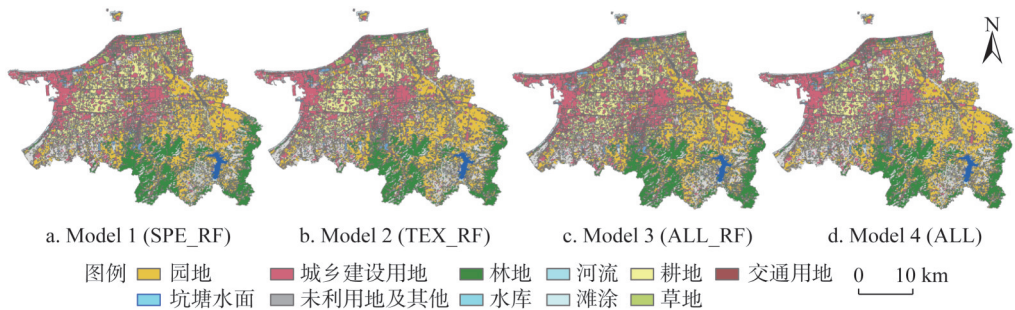


图4 不同模型分类结果

Fig. 4 The classification result of different models

表4 不同模型执行时间、OOB 错分率和分类精度

Table 4 The execution time, OOB misclassification and classification accuracy of different models

分类模型	算法执行时间	OOB 错分率	总体精度/%	Kappa 系数
Model 1(SPE_RF)	42 m 34 s	0.1075	84.93	0.818
Model 2(TEX_RF)	39 m 49 s	0.0958	89.37	0.872
Model 3(ALL_RF)	38 m 32 s	0.1026	87.52	0.849
Model 4(ALL)	47 m 21 s	0.1028	87.16	0.845

斑块，其几何特征与地物类别间相关性虽得到下降，但对于耕地、园地、林地以及线状地物如道路等形状规则、排列整齐的图斑等起到一定约束作用。

上述表明并非选取特征数量越多，其分类效果越好，故在构建和应用模型时，祛除冗余特征对于提高运算效率和精度来说是必要的。

2.2 分类精度分析

参照研究区土地利用现状图，辅以天地图、Google earth 等高分辨率影像，针对以水域滩涂为父类的各地物随机选取均匀分布的20个样区，其中坑塘水面、水库水面和河流三类共10个样区，滩涂10个样区，与上述选取的验证样点（表1）共同建立误差矩阵，计算四类模型分类精度和Kappa系数（表4）。

由表4可看出，只对纹理特征降维的Model 2解译的总体精度和Kappa系数分别为89.37%和0.872，在4类模型中分类精度最高，明显优于其他3个模型，取得较好分类效果。其次是全部特征降维的Model 3，分类总体精度和Kappa系数分别为87.52%和0.849，比Model 2分别低了2.11%和2.71%。再次是所有特征建模的Model 4，分类总体精度和Kappa系数较Model 2低了2.54%和3.20%。全部特征降维建模的Model 1提取精度最低。

2.3 提取结果分析

建模精度和分类精度比较分析表明，Model 2提取土地利用信息效果最好。汇总其分类混淆矩阵（表5）和分类结果（表6）。

表5表明，园地、林地、城乡建设用地、交通用地、河流、水库、坑塘水面7类地物生产者精度和用户精度较高，均达85%以上。滩涂、草地和未利用地生产者精度偏低，草地精度偏低主要由于部分草地易与园地和林地混淆，且草地数量相对较少，多与居民点镶嵌分布，不易精确提取。未利用地精度偏低主要由于研究区西南部未利用地、园地和耕地交错分布，加之山地及水域周边的裸地、工矿废弃地等，导致图斑相对破碎，易与其他地类混淆。滩涂精度偏低主要由于近海滩涂光谱特征与水体相近和龙口市近年来“填海造陆”的行为，沿岸滩涂易与未利用地混淆。从目视效果来看，面向对象规则集与分类器共同作用，将水体分为坑塘水面、水库、河流三类，相较于基于像元分类，更接近地表真实情况；且紧致度、长宽比等特征对具有明显形状特征的斑块起到约束作用，相对减轻了“椒盐现象”；纹理特征的引入和筛选，不仅可较好区分纹理特征差异较大的园地、林地等绿色植被，而且提高了运算效率和分类效果。

综合以往学者研究，面向对象分类技术与基于像元的最大似然法、马氏距离等方法相比具有明显的优越性，整体分类精度得到更高，噪声和破碎图斑等现象也得到减轻<sup>[19]</sup>。在面向对象规则集分类中，选取何种特征与其应用顺序带有较强主观性，特征阈值的设

表5 分类结果混淆矩阵

Table 5 Confusion matrix of classification result (%)

地物类别	坑塘水面	滩涂	草地	园地	耕地	未利用地及其他	林地	城乡建设用地	交通用地	河流	水库水面	未分类	用户精度
坑塘水面	100	6.34	0	0	0	0	0	0	0	0	0	0	2.01
滩涂	0	61.77	0	0	0	0	0	0	0	0	0	0	2.69
草地	0	1.39	76.44	4.33	3.41	0	0	0	0	0	0	0	7.83
园地	0	0	18.40	89.57	3.73	3.83	2.55	0	0	0	0	0	20.74
耕地	0	0	4.33	1.88	86.45	1.16	0	0.31	0	0	0	0	14.30
未利用地及其他	0	22.61	0	0	6.41	75.10	0.76	2.24	0	0	0	0	5.85
林地	0	1.00	0.31	3.62	0	1.43	96.59	0	0	0	0	0	26.93
城乡建设用地	0	1.52	0	0	0	17.64	0	92.72	8.61	0	0	0	15.34
交通用地	0	4.29	0.52	0.60	0	0.84	0.09	4.73	91.39	0	0	0	1.89
河流	0	0	0	0	0	0	0	0	0	100	0	0	0.85
水库水面	0	1.07	0	0	0	0	0	0	0	0	100	0	1.58
未分类	0	0	0	0	0	0	0	0	0	0	0	0	0
生产者精度	100	61.77	76.44	89.57	86.45	75.10	96.59	92.72	91.39	100	100		

表6 2014年龙口市各地类面积统计

Table 6 Land area statistics of Longkou city in 2014

地类	耕地	园地	林地	草地	城乡建设用地	交通用地	滩涂	河流	坑塘水面	水库	未利用地及其他
面积/hm <sup>2</sup>	17311.67	24079.97	14134.58	2568.03	17361.74	3577.91	1074.27	152.49	729.04	576.46	8538.84
比例/%	19.21	26.72	15.69	2.85	19.27	3.97	1.19	0.17	0.81	0.64	9.48

定需要反复试验，依赖工作人员操作经验等。上述弊端均会影响分类精度，而模型因具有一定封装性和稳定性，在一定程度上可克服上述弊端<sup>[2]</sup>。在面向对象分类器分类中，无法灵活运用全要素实现高精度提取，如将水体细分为河流、水库、坑塘水面等，并且无法对于不同地类灵活设置不同分割参数。本文提出的方法，针对陆地和水域滩涂二者不同特点，灵活设置不同分割参数，实现规则集与分类器的有机结合，更接近实际情况，便于操作实现。

由表6可看出，龙口市2014年土地利用结构以园地、城乡建设用地、林地、耕地为主，四种地类占总面积的80.89%。园地面积最多，为24079.97 hm<sup>2</sup>，占总面积26.72%，主要分布在东部和南部低山丘陵区一带，体现了龙口市以水果为主的农业生产特点。城乡建设用地面积为17361.74 hm<sup>2</sup>，占总面积19.27%，为城乡居民点、港口码头用地、旅游度假区等，主要分布在西北部滨海区及中部中心城区，体现了龙口市经济社会发达的经济特点、沿海区域及中心城区的区位优势。耕地面积为17311.67 hm<sup>2</sup>，占总面积19.21%，主要分布在西北部平原区，并与农村居民点、园地镶嵌分布，既便于农民耕种，也体现了根据市场需求耕地与园地相互转换的特征。林地面积为14134.58 hm<sup>2</sup>，占总面积15.69%，主要分布在东南部、南部低山丘陵区、北部旅游度假区一带。交通用地占3.97%，全区域均有分布，促进了区域均衡发展。其余地类面积较少，滩涂多分布于



沿海及内陆水域,未利用地主要分布于山地等。

### 3 结论与讨论

本研究结合Relief F算法和RF进行特征组合优化和影像分类,应用OBIA提取龙口市土地利用信息,取得了较好的模型精度和分类精度,表现出面向对象、Relief F算法和RF模型结合在中分辨率遥感影像土地利用信息提取中具有较高的适宜性。得出如下结论:

(1) 针对不同地物特点,阈值分割和多尺度分割结合创建影像对象,Relief F算法筛选优化特征子集,克服了以往单一规则集和分类器分类的弊端,充分发挥二者结合的优势<sup>[22]</sup>,体现了面向对象的基本思想。

(2) 基于Relief F算法优化特征构建4类特征子集,并分别应用于RF模型中。四类模型中,只纹理特征降维的建模精度和分类精度均最高,模型OOB错分率为0.0958,总体分类精度和Kappa系数分别为89.37%和0.872。该类模型滤去了相关性低、计算量大的纹理特征,保留了完整的光谱、几何和空间特征,既减少了运算时间,也提高了提取精度和效率。

(3) 土地利用信息提取结果表明,研究区土地利用以园地、城乡建设用地、林地、耕地为主,占土地总面积的80.89%。园地面积最多,体现龙口市水果为主的农业生产特点。其次是城乡建设用地,主要分布于西北部沿海地区及中部中心城区,体现了龙口经济社会发达及沿海开发和中心城区的区位特点。

受限于Landsat 8数据自身分辨率,经多尺度分割后得到的细小斑块极易融合不同地物。有学者研究指出,融合了空间特征的土地覆被分类,其分类精度虽在一定限度内得以提高,但仍无法较好解决混合像元的问题<sup>[23-24]</sup>。因此本研究基于Relief F算法分别筛选对地物成像影响较大、易造成高度冗余的光谱和纹理特征,并结合传统的全要素和全要素优化子集应用于面向对象土地利用信息提取中。结果表明只纹理特征降维的Model 2,其建模精度和分类精度在4类模型中最高。可见,光谱特征的完整性对地物分类较为重要,RF模型本身的抗噪和降维性能可在后续解译中得到较好体现,几何、纹理、空间特征可对地物分类起到一定约束作用,使得精度得到一定提高。但受到不同研究区时相、地貌、成像条件等的地域性差异,根据实际情况是否融合多源数据协同分类、该类模型在其他地区是否适宜等还需进一步验证,也是下一步需要继续研究探讨的内容。在遥感解译中,数理统计模型的广泛应用在一定程度上降低了主观性和不可控性,如RF模型封装性强,可看作黑箱,只能通过调整参数来完善模型。因此,参数组合最优解的寻找方法对提高分类效率至关重要,关于不同模型的搭配组合对模型和分类精度的影响也是后续需深入研究的内容。

**致谢:** 本文研究过程中,有幸得到中国科学院地理科学与资源研究所王介勇副研究员、王志华老师和南京大学地理与海洋科学学院马磊副研究员的热心指导,在此表示衷心感谢!

### 参考文献(References):

[1] 杜习乐,吕昌河,王海荣.土地利用/覆被变化(LUCC)的环境效应研究进展.土壤,2011,43(3): 350-360. [DU X L,

- LYU C H, WANG H R. Review: Researches on environmental effects of land use/cover change. *Soils*, 2011, 43(3): 350-360.]
- [2] ROUNSEVELL A, PEDROLI G B M, ERB K H, et al. Challenges for land system science. *Land Use Policy*, 2012, 29(4): 899-910.
- [3] 刘纪远, 匡文慧, 张增祥, 等. 20世纪80年代末以来中国土地利用变化的基本特征与空间格局. *地理学报*, 2014, 69(1): 3-14. [LIU J Y, KUANG W H, ZHANG Z X, et al. Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s. *Acta Geographica Sinica*. 2014, 69(1): 3-14.]
- [4] 于新洋, 张安定, 侯西勇. 胶东半岛果园TM影像信息的提取决策树方法. *测绘科学*, 2012, 37(4): 57-60. [YU X Y, ZHANG A D, HOU X Y. Decision tree classification of orchard information extraction from TM imagery in Jiaodong Peninsula of China. *Science of Surveying and Mapping*, 2012, 37(4): 57-60.]
- [5] 王彩艳, 王瑗玲, 王介勇, 等. 基于面向对象的海岸带土地利用信息提取研究. *自然资源学报*, 2014, 29(9): 1589-1597. [WANG C Y, WANG A L, WANG J Y, et al. Coastal zone land use information extraction based on object-oriented classification method. *Journal of Natural Resources*, 2014, 29(9): 1589-1597.]
- [6] 陈元鹏, 罗明, 彭军还, 等. 基于网格搜索随机森林算法的工矿复垦区土地利用分类. *农业工程学报*, 2017, 33(14): 250-257. [CHEN Y P, LUO M, PENG J H, et al. Classification of land use in industrial and mining reclamation area based grid-search and random forest classifier. *Transactions of the CSAE*, 2017, 33(14): 250-257.]
- [7] BREIMAN L. Random forests. *Machine Learning*, 2001, 45(1): 5-32.
- [8] BARDAN G, JOHN R, VICTOR R G, et al. An evaluation of bagging, boosting, and random forests for land-cover classification in cape cod, Massachusetts, USA. *Mapping Sciences & Remote Sensing*, 2012, 49(5): 623-643.
- [9] 孙雪莲, 舒清态, 欧光龙, 等. 基于随机森林回归模型的思茅松人工林生物量遥感估测. *林业资源管理*, 2015, (1): 71-76. [SUN X L, SHU Q T, OU G L et al. Remote sensing estimation of the biomass of artificial simao pine forest based on random forest regression. *Forest Resources Management*, 2015, (1): 71-76.]
- [10] 王书玉, 张羽威, 于振华. 基于随机森林的洪河湿地遥感影像分类研究. *测绘与空间地理信息*, 2014, 37(4): 83-85. [WANG S Y, ZHANG Y W, YU Z H. Classification of Honghe wetland remote sensing image based on random forests. *Geomatics & Spatial Information Technology*, 2014, 37(4): 83-85.]
- [11] 张秀英, 冯学智, 江洪. 面向对象分类的特征空间优化. *遥感学报*, 2009, 13(4): 664-677. [ZHANG X Y, FENG X Z, JIANG H. Object-oriented classification based on feature space optimization. *Journal of Remote Sensing*, 2009, 13(4): 664-677.]
- [12] WANG Y, MAKEDON F. Application of Relief-F feature filtering algorithm to selecting informative genes for cancer classification using microarray data//IEEE computational systems bioinformatics conference. IEEE Computer Society, 2004: 497-498.
- [13] 肖艳, 姜琦刚, 王斌, 等. 基于Relief F和PSO混合特征选择的面向对象土地利用分类. *农业工程学报*, 2016, 32(4): 211-216. [XIAO Y, JIANG Q G, WANG B, et al. Object based land-use classification based on hybrid feature selection method of combining Relief F and PSO. *Transactions of the CSAE*, 2016, 32(4): 211-216.]
- [14] MA L, FU T Y, BLASCHKE T, et al. Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS International Journal of Geo-Information*, 2017, 6(2): 51.
- [15] 洪波. 高光谱遥感图像信噪比估算方法研究. 北京: 中国科学院大学, 2013. [HONG B. Study on methods for SNR estimation of hyperspectral remote sensing images. Beijing: University of Chinese Academy Sciences, 2013.]
- [16] 周廷刚, 郭达志. 灰度矢量多波段遥感影像纹理特征及其描述. *西安科技大学学报*, 2000, 20(4): 336-338. [ZHOU T G, GUO D Z. A study of texture and description for multiband remote sensing image based on the grey vector. *Journal of Xi'an University of Science & Technology*, 2000, 20(4): 336-338.]
- [17] 吴泉源, 侯志华, 于竹洲, 等. 龙口市海岸带土地利用动态变化分析. *地理研究*, 2006, 25(5): 921-929. [WU Q Y, HOU Z H, YU Z Z, et al. Analysis of the dynamic change of land use in Longkou city's coastal zone based on remote sensing technology. *Geographical Research*, 2006, 25(5): 921-929.]
- [18] BENZ U C, HOFMANN P, WILLHAUCK G, et al. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry & Remote Sensing*, 2004, 58: 239-258.
- [19] 常春艳, 赵庚星, 王凌, 等. 滨海光谱混淆区面向对象的土地利用遥感分类. *农业工程学报*, 2012, 28(5): 226-231.

- [CHANG C Y, ZHAO G X, WANG L, et al. Land use classification based on RS object-oriented method in coastal spectral confusion region. Transactions of the CSAE, 2012, 28(5): 226-231.]
- [20] WANG Y, MAKEDON F. Application of Relief F feature filtering algorithm to selecting informative genes for cancer classification using microarray data. IEEE Computational Systems Bioinformatics, 2004: 497-498.
- [21] EISAVI V, HOMAYOUNI S, YAZDI A M, et al. Land cover mapping based on random forest classification of multitemporal spectral and thermal images. Environmental Monitoring & Assessment, 2015, 187(5): 1-14.
- [22] 张贵花, 王瑞燕, 赵庚星, 等. 基于物候参数和面向对象法的滨海生态脆弱区植被遥感提取. 农业工程学报, 2018, 34(4): 209-216. [ZHANG G H, WANG R Y, ZHAO G X, et al. Extraction of vegetation information in coastal ecological vulnerable areas from remote sensing data based on phenology parameters and object-oriented method. Transactions of the CSAE, 2018, 34(4): 209-216.]
- [23] 黄昕, 张良培, 李平湘. 融合形状和光谱的高空间分辨率遥感影像分类. 遥感学报, 2007, 11(2): 193-200. [HUANG X, ZHANG L P, LI P X. Classification of high spatial resolution remotely sensed imagery based on the fusion of spectral and shape features. Journal of Remote Sensing, 2007, 11(2): 193-200.]
- [24] SEGL K, ROESSNER S, HEIDEN U, et al. Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data. ISPRS Journal of Photogrammetry & Remote Sensing, 2003, 58(1): 99-112.

## The extraction approach of land use information combining OBIA with RF in Longkou city

WANG Ai-ling, ZHANG Xiao-qian, SU Chen-chen, YU Xin-yang

(College of Resources and Environment, Shandong Agricultural University, Taian 271018, Shandong, China)

**Abstract:** In order to improve the interpretation precision of the medium resolution satellite image, this paper proposed a new extraction approach of land use information combining Object Based Image Analysis (OBIA) with Random Forest (RF). Using the Landsat 8 OLI image and according to the features of all kinds of ground objects, the image objects were created combined with the multi-threshold and multi-resolution segmentation method, and the rule set and classifier were collaboratively used in the image classification. The Relief F algorithm was used to dimensionally reduce the spectral, texture and all feature variables, and to select 3 feature subsets. Then the RF model was conducted with the 3 feature subsets and all feature subset to build 4 models. The 4 models were applied to extract land use information in Longkou city, and the results were compared. The result indicated that the OOB (Out of Bag) misclassification, classification accuracy and Kappa index were 0.0958, 89.37% and 0.872 respectively with the land use information extraction approach combining OBIA with RF, dimension reduction based on the Relief F algorithm only for texture features. This retained the complete spectral, geometric and spatial features, which has a higher accuracy. The approach can be applied to the extraction of land use information with the medium resolution satellite image.

**Keywords:** land use information; the extraction approach; object-based; Relief F algorithm dimensionally reduced; Random Forest; Longkou city