

## 基于 Boruta 和极端随机树方法的森林蓄积量估测

韩 瑞<sup>1 2 3</sup>, 吴达胜<sup>1 2 3</sup>, 方陆明<sup>1 2 3</sup>, 黄宇玲<sup>1 2 3 4</sup>

(1. 浙江农林大学 信息工程学院, 杭州 311300; 2. 林业感知技术与智能装备国家林业和草原局重点实验室, 杭州 311300;  
3. 浙江省林业智能监测与信息技术研究重点实验室, 杭州 311300; 4. 醴陵市陶瓷烟花职业技术学校, 湖南 醴陵 412200)

**摘要:** 森林蓄积量是反映森林资源数量的重要指标之一。本研究应用 Boruta 特征选择方法和极端随机树 (Extremely randomized trees, Extra-trees) 方法, 以小班为研究单元, 估测龙泉市部分区域森林资源的每公顷蓄积量, 为县域尺度森林蓄积量的估测提供新的方法和思路。基于研究区的森林资源二类调查数据、高分二号 (GF-2) 遥感影像数据、数字高程模型数据, 提取多元特征组成原始特征集。通过 Boruta 选择方法对原始特征集进行筛选, 利用 Extra-trees 方法建立森林蓄积量估测模型, 选用十折交叉验证法对模型进行检验, 并与随机森林 (Random Forest, RF) 方法和梯度提升 (Gradient Boosting) 方法进行对比分析。研究结果显示: 1) 经过 Boruta 特征选择方法得出的特征有土层厚度、年龄、郁闭度、海拔、坡度和坡向; 2) 极端随机树方法采用网格搜索调参得到的最优参数组合为: 树的个数为 250, 树的最大深度为 14; 3) 基于 Boruta 和极端随机树方法的森林蓄积量估测模型的测试精度为 84.14%,  $R^2$  为 0.92,  $RMSE$  为  $19.65 \text{ m}^3/\text{hm}^2$ ,  $MAE$  为  $13.95 \text{ m}^3/\text{hm}^2$ , 模型优于随机森林方法和梯度提升方法, 表明 Boruta 特征选择方法结合极端随机树方法估测森林蓄积量可取得更好的效果。

**关键词:** Boruta 特征选择; 极端随机树; 随机森林; 森林蓄积量; 机器学习

中图分类号: TP757 文献标识码: A 文章编号: 1002-6622(2020)04-0127-07

DOI: 10.13466/j.cnki.lyzygl.2020.04.018

## Estimation of Forest Reserves Based on Boruta and Extra-trees Methods

HAN Rui<sup>1 2 3</sup>, WU Dasheng<sup>1 2 3</sup>, FANG Luming<sup>1 2 3</sup>, HUANG Yuling<sup>1 2 3 4</sup>

(1. School of Information Engineering Zhejiang Agriculture and Forestry University Hangzhou Zhejiang 311300, China; 2. Key Laboratory of State Forestry and Grassland Administration on Forestry Sensing Technology and Intelligent Equipment Hangzhou Zhejiang 311300, China; 3. Key Laboratory of Forestry Intelligent Monitoring and Information Technology of Zhejiang Province Hangzhou Zhejiang 311300, China; 4. Ceramics-Fireworks Vocational Technology School of Liling Hunan 412200, China)

**Abstract:** Forest reserve is an important index to show the quantity of forest resources. In this study, the boruta feature selection method and the extremely randomized trees (Extra-trees) method are used to estimate the forest resources per *mu* in some areas of Longquan city. The study takes small classes as the research unit to provide new methods for the estimation of forest reserves at the county level. Based on the secondary survey data of forest resources, GF-2 remote sensing image data and digital elevation model data, multiple features are collected to form the original feature set. Through the boruta selection method, the original feature set is screened, the forest volume estimation model is established by extra-trees method,

收稿日期: 2020-05-03; 修回日期: 2020-06-03

基金项目: 浙江省科技重点研发计划资助项目 (2018C02013)

作者简介: 韩瑞 (1995-) 男, 河南周口人, 在读硕士, 主要从事资源与环境信息系统研究。Email: 804874277@qq.com

通讯作者: 吴达胜 (1972-) 男, 浙江庆元人, 教授, 主要从事农业遥感与信息技术研究。Email: 458752249@qq.com

and the ten fold cross validation method is used to test the model ,which is compared with the random forest( RF) method and gradient boosting method. The results show that: ( 1) the features found by the boruta feature selection method are soil thickness ,age ,canopy density ,altitude and slope; ( 2) the optimal parameter combination obtained by the grid search and parameter adjustment of the Extra-trees method is: the number of tree is 250 ,and the maximum depth of the tree is 14; ( 3) the testing accuracy of the forest volume estimation model based on the boruta and extra-trees method is 84. 14% , $R^2$  is 0. 92 ,RMSE is  $19.65\text{m}^3/\text{hm}^2$  ,and MAE is  $13.95\text{m}^3/\text{hm}^2$ . The model is superior to the random forest method and the gradient lifting method. It shows that the boruta feature selection method combined with the Extra-trees method can achieve better results in estimating the forest reserve.

**Key words:** boruta feature selection ,extremely randomized trees ,random forest ,forest reserves ,machine learning

森林蓄积量( Forest Stock Volume ,FSV) 是指一定面积上的林木树干部分的总材积 ,反映了一个国家或地区森林资源总体规模的大小 ,是全国森林资源调查的一项重要内容<sup>[1]</sup>。森林蓄积量体现了森林生态系统林分信息 ,与森林生物量、生物多样性和碳储量等息息相关 ,是反映森林资源数量的重要指标 ,已经成为林业科学研究中的重点内容<sup>[2]</sup>。基于森林大数据的蓄积量估测 ,通过特征选择可以降低原特征集的维度以及减少冗余信息 ,达到提高预测精度以及减低模型训练耗时的目的<sup>[3]</sup>。刘明艳等<sup>[4]</sup>采用主成分分析方法对实验数据进行降维 ,利用线性回归方法建立森林蓄积量估测模型 ,并取得良好的效果; 周如意<sup>[5]</sup>选用相关性分析对所有自变量因子进行筛选 ,通过多元线性回归、偏最小二乘法以及广义回归神经网络分别建立森林蓄积量估测模型 ,对龙泉市森林蓄积量进行全局预测; 王海宾等<sup>[6]</sup>基于高分一号遥感影像数据 ,使用 K-最邻近方法( K-Nearest Neighbor ,KNN) 建立延庆区森林蓄积量估测模型 ,并与偏最小二乘回归方法进行对比 ,结果显示 KNN 方法优于偏最小二乘回归方法; Wu<sup>[7]</sup>通过 Levenberg-Marquardt( LM) 优化算法( LM-BP) 改进反向传播( BP) 神经网络( NN) 模型估测龙泉市森林资源蓄积量; 汪康宁等<sup>[8]</sup>以黑龙江凉水自然保护区为研究对象 ,使用随机森林算法( Random Forest ,RF) 建立蓄积量反演模型。随着遥感技术、机器学习算法和神经网络技术的发展 ,森林蓄积量的估

测正朝着多源、非线性回归模型的趋势发展。在数据源方面 ,高分遥感数据、雷达数据、数字高程模型数据得到了广泛的运用。在森林蓄积量估测方法方面 ,各种传统的多元线性回归方法得到不断的改善 ,机器学习方法逐渐渗透到森林蓄积量的研究。应用精度更高的多源数据以及合适的回归模型对森林蓄积量进行估测成为了森林蓄积量估测的持续重点和难点。选择合适的森林蓄积量影响因子和森林蓄积量估测算法 ,不仅对精确估算森林蓄积量有着重要的意义 ,也对提高森林资源监测效率有着很大的影响。本研究引入 Boruta 特征选择方法<sup>[9]</sup>结合极端随机树( Extremely Randomized Trees ,Extra-trees) 方法 ,基于高分二号( GF-2) 遥感影像数据、数字高程模型( Digital Elevation Model ,DEM) 数据以及森林资源二类调查数据 ,对龙泉市部分区域小班的每公顷蓄积量进行估测 ,并与随机森林( Random Forest ,RF) 方法<sup>[10]</sup>和梯度提升( Gradient Boosting) 方法<sup>[11]</sup>进行性能对比 ,为快速估测县域尺度森林蓄积量提供新的方法和思路。

## 1 材料与方法

### 1.1 研究区概况

龙泉市位于浙江省西南部的浙闽赣边境(  $27^{\circ}42' \sim 28^{\circ}20'N$  , $118^{\circ}42' \sim 119^{\circ}25'E$ ) ,属于中亚热带季风气候区 ,温暖湿润、四季分明、雨量充沛、气候宜人 ,年日照时数为 1 849. 8h ,年降水量为 1 699. 4 mm ,

年相对湿度为 79% ,无霜期为 263d<sup>[12]</sup>。龙泉市是浙江省的重点林区,素有“浙南林海”之称,森林覆盖率高达 84.3% ,森林面积达到 25.72 万  $\text{hm}^2$  ,蓄积量达到 1 912 万  $\text{m}^3$  ,占丽水市的总森林面积的 17.5% 。

## 1.2 研究数据

本文的研究数据来源于龙泉市 2017 年森林资源二类调查数据(以下简称“二调数据”)、2009 年龙泉市数字高程模型( DEM) 数据和龙泉市 2016 年 8 月 16 日的一景高分二号( GF-2) 遥感影像数据。本文研究的 GF-2 数据只用到其多光谱影像数据。

本文的估测指标为单位蓄积量(  $\text{m}^3/\text{hm}^2$  ) ,研究单元为森林资源小班。考虑到森林蓄积量的影响特征以及参考前人的研究<sup>[13-15]</sup> ,本文从二调数据中获取土层厚度( TU\_CENG\_HD) 、腐殖质厚度( FU\_ZHI\_HD) 、年龄( NL) 和郁闭度( YU\_BI\_DU) 等 4 个自变量因子;从 DEM 数据中提取海拔( HAI\_BA) 、坡度( PO\_DU) 和坡向( PO\_XIANG) 等 3 个自变量因子;从 GF-2 数据中提取蓝色波段( BAND\_1) 、绿色波段( BAND\_2) 、红色波段( BAND\_3) 和近红外波段( BAND\_4) 等 4 个波段因子。通过对波段因子的计算,得到归一化植被指数( Normalized Difference Vegetation ,NDVI) 、比值植被指数( Ratio Vegetation Index ,RVI) 、增强型植被指数( Enhanced Vegetation Index ,EVI) 、差值植被指数( Difference Vegetation Index ,DVI) 、土壤调节植被指数( Soil-Adjusted Vegetation Index ,SAVI) 等 5 个植被指数。所有上述因子数据提取后,都集成到森林资源小班记录中,剔除蓄积量为零后的小班数据共 4 002 个,作为本研究的数据集参与后续实验,从中随机抽取 90% 的小班数据共 3 602 个,作为训练集数据,剩余的 10% 的小班数据共 400 个作为测试集数据。

## 1.3 Boruta 特征选择方法

特征选择是从初始特征中按照一定的准则或方法选出部分特征,因此,被选出的特征集合必然是原始特征集的子集,保留原始特征的物理意义。Boruta 特征选择方法的目标是选出所有与因变量相关的特征集合,从而更全面地理解因变量的影响因素。Boruta 特征选择方法的主要步骤如下:

1) 创建阴影特征:对于每个真实特征 R ,随机打乱特征顺序,得到阴影特征矩阵 S ,拼接至真实特征之后,得到新的特征矩阵  $N = [R \ S]$  。

2) 用新的特征矩阵 N 作为输入,训练模型(可选随机森林方法、极端梯度提升方法等),得到真实特征和阴影特征的重要性。

3) 取阴影特征矩阵 S 中重要性的最大值  $S_{\max}$  ,真实特征矩阵 R 中的重要性大于  $S_{\max}$  的特征,记录一次命中。

4) 利用 3) 中的真实特征累计命中,标记特征重要或不重要。

5) 删除不重要的特征,重复步骤 1) —4) ,直到所有特征都被标记。

最终所有的特征变量得到确认或拒绝,或者算法到达模型运行的规定限制时,算法停止<sup>[16]</sup> ,因此,最终的特征集被分为两种类型:被确定和被拒绝。

## 1.4 极端随机树方法

极端随机树( Extremely Randomized Trees ,Extra-trees) 方法是在 2006 年由 Pierre Geurts 等学者经过大量的实验研究之后提出的一种机器学习方法<sup>[17]</sup> 。它是随机森林方法的一个变种,原理基本和随机森林一样,通过集成多个决策树进行打分,根据各个决策树预测值的平均值来进行投票。极端随机树方法的核心如下:

1) 选择样本。极端随机树方法的每个子决策树采用原始数据集训练。

2) 选择特征。极端随机树方法随机选择一个特征值来划分决策树。

3) 构建决策树。在形成决策树的过程中,每一个节点都要按照步骤 2) 进行分裂,直到无法再分裂为止。

4) 极端随机树预测。通过步骤 1) —3) 的迭代执行,建立大量的决策树,进一步构成森林。把测试样本输入到森林中,利用每一株决策树对测试样本进行分类或者回归,得到最终的分类或者回归估计结果。

由于极端随机树方法是基于随机选择特征点进行划分的方法,最终得到的决策树规模大于随机森林生成的决策树,但泛化能力强于随机森林。

### 1.5 模型评价

本研究选取交叉验证 Cross Validation( CV) 的方式来检验模型的精度。交叉验证常用的方法有留一交叉( LOOCV) 和 K 折交叉验证( K-fold CV) 。K-fold CV 是将数据集分为 K 个子集,并在 K 个子集中取一个作为验证数据集,其余的 K-1 个数据集作为训练集,计算所得的 K 个模型的预测精度的平均值作为最终的精度值。本研究采用十折交叉验证法。常用评价指标有决定系数( R-squared , $R^2$ )、均方根误差( Root Mean Square Error ,RMSE)、平均绝对误差( Mean Absolute Error ,MAE)、平均百分比误差( Mean Absolute Percentage Error ,MAPE) 和估测精度( Prediction accuracy ,P) 。计算公式如式( 1) — ( 5) 所示。

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_{li})^2} \quad (2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (3)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{y_i} \quad (4)$$

$$P = \left( 1 - \frac{1}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{y_i} \right) \times 100\% \quad (5)$$

式中:  $N$  为样本数,  $y_i$  为真实值,  $\hat{y}_i$  为预测值,  $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ 。

## 2 结果与分析

### 2.1 Boruta 特征选择结果

本研究选取随机森林方法对 Boruta 进行训练,最终得到的特征选择结果(表1)。从表1中可知,经过 Boruta 特征选择,被拒绝的特征有腐殖质厚度( FU\_ZHI\_HD)、蓝色波段( BAND\_1)、绿色波段( BAND\_2)、红色波段( BAND\_3)、近红外波段( BAND\_4)、归一化植被指数( NDVI)、比值植被指数( RVI)、增强型植被指数( EVI)、差值植被指数( DVI)、土壤调节植被指数( SAVI) 等 10 个特征,而最终被选择的特征集包含了土层厚度( TU\_CENG\_HD)、年龄( NL)、郁闭度

( YU\_BI\_DU)、海拔( HAI\_BA)、坡度( PO\_DU) 和坡向( PO\_XIANG) 等 6 个特征。

表1 Boruta 特征选择结果表

Tab. 1 Boruta feature selection result table

特征	结果	特征	结果
土层厚度	通过	腐殖质厚度	不通过
年龄	通过	郁闭度	通过
海拔	通过	坡度	通过
坡向	通过	蓝色波段	不通过
绿色波段	不通过	红色波段	不通过
近红外波段	不通过	差值植被指数	不通过
比值植被指数	不通过	增强型植被指数	不通过
差值植被指数	不通过	壤调节植被指数	不通过

### 2.2 模型参数优化结果

以往研究表明,参数优化对于建模具有重要意义。本研究选用 3 种机器学习方法( 随机森林方法、Gradient Boosting 方法、极端随机树方法),参考现有文献以及反复实验设置各方法对应的参数,在此基础上对比参数调整过程中对建模精度的影响。

本研究在 python 环境下运行,每个方法的参数如下:

1) 随机森林方法。其基本参数为: 树的个数  $N$ , 设置为( 50 ,100 ,150 ,200 ,250); 树的最大深度  $M$ , 设置为( 2 ,6 ,10 ,14 ,18) 。

2) Gradient Boosting 方法。其基本参数为: 树的个数  $N$ , 设置为( 50 ,100 ,150 ,200 ,250); 树的最大深度  $M$ , 设置为( 2 ,6 ,10 ,14 ,18) 。

3) 极端随机树方法。其基本参数为: 树的个数  $N$ , 设置为( 50 ,100 ,150 ,200 ,250); 树的最大深度  $M$ , 设置为( 2 ,6 ,10 ,14 ,18) 。

本研究采用网格搜索对模型参数进行优化,通过比较不同参数对负均方误差 neg\_m 的影响,负均方误差大的模型较优,从而得出模型最优参数,结果如表2所示。由表2可知,随机森林方法的负均方误差 neg\_m 最大为 -5.82,最优参数组合为树的个数  $N = 150$ , 树的最大深度  $M = 14$ ; Gradient Boosting 方法负均方误差 neg\_m 最大为 -6.03,最优参数组合为树的个数  $N = 50$ , 树的最大深度  $M = 6$ ; 极

表2 网格搜索参数优化结果

Tab. 2 Grid search parameter optimization results

树的最大深度/m	树的个数/个	随机森林的均方误差	梯度提升的均方误差	极端随机树的均方误差
2	50	-8.21	-6.49	-10.86
2	100	-8.23	-7.26	-10.73
2	150	-8.22	-6.22	-10.52
2	200	-8.12	-6.21	-10.52
2	250	-8.24	-6.23	-10.69
6	50	-6.23	-6.03	-6.86
6	100	-6.26	-6.14	-6.87
6	150	-6.25	-6.25	-6.87
6	200	-6.25	-6.29	-6.86
6	250	-6.27	-6.32	-6.88
10	50	-6.01	-6.48	-5.92
10	100	-5.95	-6.52	-5.93
10	150	-5.93	-6.54	-5.91
10	200	-5.92	-6.54	-5.92
10	250	-5.94	-6.51	-5.91
14	50	-5.94	-7.44	-5.73
14	100	-5.90	-7.42	-5.74
14	150	-5.82	-7.40	-5.73
14	200	-5.86	-7.42	-5.72
14	250	-5.89	-7.47	-5.71
18	50	-5.93	-8.63	-5.88
18	100	-5.88	-8.62	-5.83
18	150	-5.87	-8.59	-5.82
18	200	-5.86	-8.64	-5.81
18	250	-5.85	-8.69	-5.77

端随机树方法负均方误差  $neg\_m$  最大为 -5.71, 最优参数组合为树的个数  $N=250$ , 树的最大深度  $M=$

14。因此, 本研究采用以上3种模型的最优参数分别建立森林蓄积量估测模型。

### 2.3 蓄积量估测模型建立

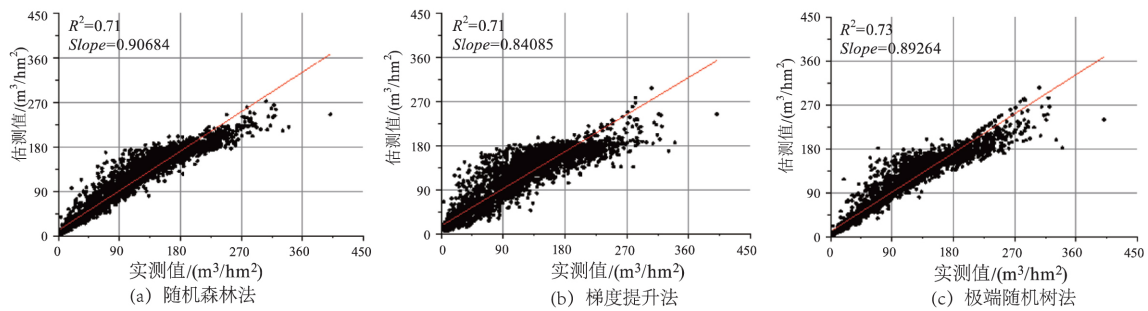
本研究的建模过程如下: 基于上述 Boruta 特征选择结果, 选择随机森林(RF)方法、梯度提升(Gradient Boosting)方法和极端随机树(Extra-Trees)方法通过网格搜索调参结果对森林蓄积量进行建模与估测。3种方法的森林蓄积量建模与估测结果如表3、图1和图2所示。

由表3可知, 从建模结果看, 随机森林(RF)方法建立的蓄积量估测模型的决定系数  $R^2$ 、均方根误差  $RMSE$ 、平均绝对误差  $MAE$  分别是 0.71、36.90( $m^3/hm^2$ )、25.65( $m^3/hm^2$ ), 其平均百分比误差  $MAPE$  以及估测精度  $P$  分别为 13.78% 和 86.22%; 梯度提升(Gradient boosting)方法建立的蓄积量估测模型的决定系数  $R^2$ 、均方根误差  $RMSE$ 、平均绝对误差  $MAE$  分别是 0.71、36.75( $m^3/hm^2$ )、25.95( $m^3/hm^2$ ), 其平均百分比误差  $MAPE$  以及估测精度  $P$  分别为 27.92% 和 72.08%; 极端随机树(Extra-Trees)方法建立的蓄积量估测模型的决定系数  $R^2$ 、均方根误差  $RMSE$ 、平均绝对误差  $MAE$  分别是 0.73、35.85( $m^3/hm^2$ )、25.05( $m^3/hm^2$ ), 其平均百分比误差  $MAPE$  以及估测精度  $P$  分别为 14.05% 和 85.95%。从估测结果看, 极端随机树(Extra-Trees)方法的决定系数  $R^2$  最大、均方根误差  $RMSE$  最小、平均绝对误差  $MAE$  最小、平均百分比误差  $MAPE$  最小、估测精度  $P$  (84.14%) 最大; 随机森林(RF)方法和梯度提升方法的估测效果差异不明显。

表3 3种方法蓄积量建模与估测评价指标

Tab. 3 Three methods of volume accumulation modeling and estimation evaluation indicators

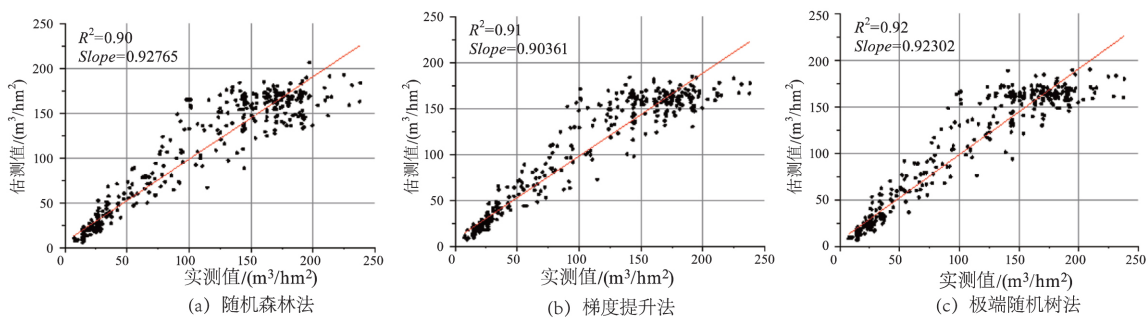
模型	决定系数		均方根误差/( $m^3/hm^2$ )		平均绝对误差/( $m^3/hm^2$ )		平均百分比误差/%		估测精度/%	
	建模	估测	建模	估测	建模	估测	建模	估测	建模	估测
随机森林	0.71	0.90	36.90	21.75	25.65	15.30	13.78	16.78	86.22	83.22
梯度提升	0.71	0.91	36.75	20.85	25.95	15.00	27.92	16.45	72.08	83.55
极端随机树	0.73	0.92	35.85	19.65	25.05	13.95	14.05	15.86	85.95	84.14



注：图中的点以实测值为X坐标，估测值为Y坐标，“—”表示拟合直线

图1 3种方法蓄积量建模结果

Fig. 1 Three methods of accumulation volume modeling results 90



注：图中的点以实测值为X坐标，估测值为Y坐标，“—”表示拟合直线

图2 3种方法蓄积量估测结果

Fig. 2 Accumulation estimation results of three methods

### 3 讨论

1) 本研究采用 Boruta 特征选择方法对可能影响森林蓄积量估测的特征因子进行筛选,筛选得出的特征集包括土层厚度、年龄、郁闭度、海拔、坡度和坡向等6个特征。Boruta 特征选择该方法虽然损失了部分细节信息,但通过去除多因子之间相关性后起到了降维作用,从而有效减少了特征的数量,提高了建模与估测的效率及模型的泛化能力。

2) 本研究率先将极端随机树方法引入到森林蓄积量估测中,采用网格搜索调参得到模型最优参数,在不区分优势树种的情况下,最终得到的森林蓄积量的估测精度达到84.14%。李世波等<sup>[18]</sup>利用随机森林方法对湖南省醴陵市的森林蓄积量进行估测,得到的估测精度为83.69%;罗蜜等<sup>[19]</sup>采用多元逐步回归构建的福建省三明市将乐国有林场的杉木蓄积量估测精度为80.67%;李亚东等<sup>[20]</sup>采用最小二乘法结合无人机航摄影像三维点云得到的

蓄积量的估测精度为82.46%。相比而言,本研究的极端随机树方法的估测精度更高,从决定系数的大小来看,本文所建模型的 $R^2$ 为0.73,优于郎晓雪等<sup>[21]</sup>基于支持向量机对香格里拉市云冷杉建立蓄积量模型的 $R^2$ 为0.67,陈新云等<sup>[22]</sup>采用多元线性回归建立的森林蓄积量模型的 $R^2$ 为0.694,以及曹霖等<sup>[23]</sup>基于随机森林方法建立的吉林省中东部森林蓄积量模型 $R^2$ 为0.5925。由此表明,基于 Boruta 特征选择方法结合极端随机树方法所建立的森林资源蓄积量估测模型具有更强的泛化能力和稳定性。

### 4 结论

本文基于浙江省龙泉市森林资源二类调查数据、高分二号(GF-2)遥感影像数据、数字高程模型数据提取多元特征组成原始特征集,应用 Boruta 特征选择方法和极端随机树(Extremely randomized

trees ,Extra-trees) 方法 ,以小班为研究单元 ,估测龙泉市部分区域森林资源的每公顷蓄积量。结果表明 ,经过 Boruta 特征选择出的特征有土层厚度、年龄、郁闭度、海拔、坡度和坡向等 6 个因子。极端随机树方法采用网格搜索之后的最优参数组合为: 树的个数  $N = 250$  ,树的最大深度  $M = 14$  ,测试精度为 84.14% , $R^2$  为 0.92 , $RMSE$  为  $19.65\text{m}^3/\text{hm}^2$  , $MAE$  为  $13.95\text{m}^3/\text{hm}^2$ 。模型优于随机森林方法和梯度提升方法 ,为精确估算森林蓄积量提供了新的方法。后续将考虑将本文研究结果尝试应用于不同时期不同区域的森林蓄积量估测。

#### 参考文献:

- [1] 吴达胜. 基于多源数据和神经网络模型的森林资源蓄积量动态监测[D]. 杭州: 浙江大学, 2014.
- [2] 尤静妮. 基于高分遥感纹理信息的森林蓄积量估测研究[D]. 西安: 西安科技大学, 2017.
- [3] 姚新华, 金佳, 徐飞飞, 等. 太湖流域果树提取的光谱和纹理特征选择研究[J]. 中国生态农业学报, 2019, 27(10): 1596–1606.
- [4] 刘明艳, 王秀兰, 冯仲科, 等. 基于主成分分析法的老秃顶子自然保护区森林蓄积量遥感估测[J]. 中南林业科技大学学报, 2017, 37(10): 80–83.
- [5] 周如意. 基于 Landsat-8 遥感影像的森林蓄积量估测[D]. 杭州: 浙江农林大学, 2019.
- [6] 王海宾, 彭道黎, 高秀会, 等. 基于 GF-1 PMS 影像和 k-NN 方法的延庆区森林蓄积量估测[J]. 浙江农林大学学报, 2018, 35(6): 1070–1078.
- [7] Wu Dasheng. Estimation of forest volume base on LM-BP neural network model[J]. Comput. Model. New Technol, 2014, 18(4): 131–137.
- [8] 汪康宁, 马婷, 吕杰. 基于随机森林算法的凉水自然保护区蓄积量反演研究[J]. 西南林业大学学报, 2016, 36(5): 125–129.
- [9] 郭海山, 高波涌, 陆慧娟. 基于 Boruta-PSO-SVM 的股票收益率研究[J]. 传感器与微系统, 2018, 37(3): 51–53.
- [10] Wei Jing, Huang Wei, Li Zhanqing, et al. Estimating 1-km-resolution PM<sub>2.5</sub> concentrations across China using the space-time random forest approach[J]. Remote Sensing of Environment, 2019, 231: 111221.
- [11] Djavan De Clercq, Wen Zongguo, Fei Fan. Determinants of efficiency in anaerobic bio-waste co-digestion facilities: A data envelopment analysis and gradient boosting approach[J]. Applied Energy, 2019, 253: 113570.
- [12] 洪汝锋. 龙泉市林权制度改革研究[D]. 杭州: 浙江农林大学, 2012.
- [13] 杨永恬. 基于多源遥感数据的森林蓄积量估测方法研究[D]. 北京: 中国林业科学研究院, 2010.
- [14] 姚啸. 面向对象的高分遥感影像分类在森林蓄积量估测中的应用研究[D]. 西安: 西安科技大学, 2015.
- [15] 王雪军. 基于多源数据源的森林资源年度动态监测研究[D]. 北京: 北京林业大学, 2013.
- [16] Kursa M B, Rudnicki W R. Feature selection with the Boruta package[J]. Journal of Statistical Software, 2010, 36(11): 1–13.
- [17] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees[J]. Machine Learning, 2006, 63(1): 3–42.
- [18] 李世波, 林辉, 王光明, 等. 基于 GF-1 的森林蓄积量遥感估测[J]. 中南林业科技大学学报, 2019, 39(8): 70–75.
- [19] 罗蜜, 孙玉军, 张博, 等. 应用衍生纹理指数对杉木林分蓄积量的估测[J]. 东北林业大学学报, 2019, 47(7): 43–49.
- [20] 李亚东, 曹明兰, 李长青, 等. 无人机森林航摄影像三维点云估测林分蓄积量研究[J]. 中南林业科技大学学报, 2019, 39(3): 56–60.
- [21] 郎晓雪, 许彦红, 舒清态, 等. 香格里拉市云冷杉林蓄积量遥感估测非参数模型研究[J]. 西南林业大学学报: 自然科学, 2019, 39(1): 146–151.
- [22] 陈新云, 李利伟, 刘承芳, 等. 热带原始森林类型分类和蓄积量遥感反演研究[J]. 林业资源管理, 2019(2): 39–46.
- [23] 曹霖, 彭道黎, 王雪军, 等. 应用 Sentinel-2A 卫星光谱与纹理信息的森林蓄积量估算[J]. 东北林业大学学报, 2018, 46(9): 54–58.