文章编号：2095-2163(2020)04-0158-04 中图分类号：TP757 文献标志码：A

基于机器学习的森林蓄积量研究综述

黄宇玲
（浙江农林大学 信息工程学院，浙江 临安 311300）

摘 要：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。本文首先对森林蓄积量的研究进展做了简单介绍；其次阐述了常见的 4 种机器学习算法、研究进展和成果；介绍了机器学习算法在森林蓄积量方面的研究。

关键词：机器学习；森林蓄积量；随机森林

A review of the research on forest volume based on machine learning

HUANG Yuling
（College of Information Engineering，Zhejiang A&F University，Lin'an Zhejiang 311300，China）

【Abstract】 Forest volume reflects the forest information of forest ecosystem，which is closely related to forest biomass，biodiversity and carbon storage. It is an important indicator to reflect the quantity of forest resources，and has become the focus of forestry research. In this paper，firstly，the research progress of forest storage is briefly introduced；secondly，four common machine learning algorithms and their research progress and achievements are described；secondly，the research of machine learning algorithm in forest storage is introduced.

【Key words】 machine learning；Forest stock；Random forest

0 引 言

目前，对于森林蓄积量的传统监测方法中，常以森林资源一、二类调查为主，这类调查方式存在调查周期长，以及需要大量的人力、物力和财力等问题。然而，遥感技术能够给地面调查提供很好的支撑与补充，是宏观、快速、经济地实现森林蓄积量估测的有效途径。

机器学习源于人工智能和统计学[1]，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中。应用回归模型对森林蓄积量进行估测，已经成为了森林蓄积量研究的重点和难点。因此，研究机器学习算法在森林蓄积量估测的应用，不仅仅对精确地估算森林蓄积量有着重要的现实意义，对提高森林资源监测效率也有着很大的影响。

1 森林蓄积量研究概述

目前对于蓄积量的估测方法，主要有：以方法或者模型为重点的森林蓄积量的估测，以及基于遥感数据的森林蓄积量的反演。

基于模型估测的方法一般通过输入年龄、郁闭度、海拔以及坡度等建模因子建立森林蓄积量模型，以估测区域的森林蓄积量。森林资源一、二类调查一直是传统森林蓄积量的测定方法。Jahangir M 等[2]采用多元逐步线性回归方法和回归树分析方法对伊拉克北部地区建立了森林蓄积量模型，结果显示回归树方法建立的森林蓄积量模型较优，其均方根误差（Root Mean Square Error，RMSE）为 88.7m^3/ha。Breidenbach 等[3]用非参数学习的朴素贝叶斯方法（Bayesian Analysis）与多元线性回归方法进行比较，对 Forbach 的森林蓄积量建立估测模型进行预测，研究结果表明朴素贝叶斯方法的精度高于多元线性回归方法。Yim 等[4]选用 K-近邻法（K-Nearest Neighbor，KNN）建立森林蓄积量估测模型，对面积不同的两个县域进行蓄积量估测，研究结果显示 K-近邻法对于小面积的森林蓄积量的反演有着良好的效果。杨明星等[5]基于 Sentinel-A 影像，通过相关性分析对研究的自变量因子进行筛选，以相关性分析特征，结果采用随机森林方法建立了思茅松林蓄积量遥感估测模型，模型的估测精度为 75.46%，得到的估测效果较好，且表明随机森林方

作者简介：黄宇玲（1995-），女，硕士研究生，主要研究方向：资源与环境信息系统。

收稿日期：2019-12-05

法在森林蓄积量建模估测研究方面具有一定的可行性与推广性。王海宾等[6]选用平均残差平方和(Residual Mean Squares,RMS)方法,对可能影响森林蓄积量的自变量因子进行筛选,利用 K-近邻方法对延庆区县域的森林蓄积量建模估测,并与偏最小二乘回归(Partial Least-Squares Regression,PLSR)方法进行对比。研究结果表明,基于 K-近邻方法得到的森林蓄积量估测的均方根误差 RMSE 为 12.80m^3/hm^2,优于偏最小二乘回归方法建立的森林蓄积量估测的均方根误差 RMSE(21.90m^3/hm^2)。刘明艳等[7]通过主成分分析方法对可能影响森林蓄积量的自变量因子进行降维,降维处理之后得到的数据集作为多元线性回归模型的输入,建立了老秃顶子自然保护区森林蓄积量估测模型,多元线性回归方程调整后的决定系数(R-squared,R^2)为 0.810,结果表明拟合度很好,估测精度达到 92.18%,研究结果满足林业调查中对蓄积量估测的要求。

近年来,随着遥感技术的快速发展,遥感影像的空间分辨率有了大幅的提高,许多学者对于遥感影像在林业科学中的研究也日趋深入。李世波[8]采用国产高分一号遥感影像数据,通过移动窗口来解决像元与样地之间的对应关系,选用多元线性逐步回归法、偏最小二乘回归法以及随机森林方法,对湖南省醴陵市的森林蓄积量进行估测。研究结果表明:利用高分一号遥感影像数据,结合随机森林方法建立的森林蓄积量模型,其估测效果较趋向于真实分布。刘俊等[9]基于(Advanced Land Observing Satellite,ALOS)卫星的 2.5 m 遥感影像计算,在不同窗口情况下的纹理特征以及纹理参数,研究区域为北京市怀柔区柞树林,建立了多元逐步线性回归柞树蓄积量模型,最终筛选出了最优反演模型为多元逐步回归模型,最优纹理生成窗口为 11×11。蔡学成等[10]利用中巴资源卫星遥感数据,通过多元线性回归方法对贵州省黎平县、从江县和榕江县建立蓄积量估测模型,最终结果显示整体估测能力较好,有一定的利用前景。张翔宇等[11]基于资源三号卫星影像,以宁波市北仑区新路林场为研究区域,采用主成分回归分析法、偏最小二乘法和多元逐步回归方法,分别建立蓄积量反演模型。最终发现,基于多元逐步回归模型反演的森林蓄积量估测精度最高。张苏等[12]以分辨率为 2 m 的高分一号卫星遥感数据为主要数据源,采用多元线性回归方法与支持向量机方法,对福建省将乐县亚热带针叶林蓄积量进行估测,最终表明支持向量机方法的森林蓄积量模型预估结果优于多元线性回归方法。

2　机器学习算法

2.1　随机森林算法

Breiman[13]在 2001 年提出的随机森林方法,是一种基于决策树的机器学习方法,也是一种 Bagging(又称套袋,是一种可以提高算法准确性的方法)集成学习方法[14],通过多个弱分类器组合在一起,最终的结果是通过投票或者取平均值,从而让模型整体的结果具有较高的准确度和泛化性能。随机森林方法的重点在“随机”和“森林”上,“随机”使得其具有抗过拟合能力,“森林”使得它结果更加精准。“随机”主要是指两个方面的随机:一是样本随机,即通过自助法重采样技术,从最初的训练样本集 N 中拿出样本再放回去,一直重复随机的取出 K 个样本,K 个样本作为新的训练样本集合($\boldsymbol{N}=\boldsymbol{K}$);二是对于特征的选择是随机的,即随机森林方法在建立每一棵决策树的时候,每棵决策树选择出来的特征仅仅是随机选出来的少数特征,在这些被选出的少数特征中,选择其中一个最优的特征来作为决策树的左右子树划分,继而将随机效果扩大,进一步增强了模型的泛化能力。随机森林方法中的“森林”是指由许许多多的决策树建立之后形式了森林。随机森林方法的学习器使用 CART 树(即分类回归树),当数据集的因变量属于连续性数值时,这种树的方法就是一个回归树,其可以采用叶子节点观察得到平均值来作为预测值;当输入的数据集为离散型数值时,这种树的方法就是一个分类树,每个叶子节点的投票结果就是分类结果。CART 树是一种二叉树,即每一个非叶子节点只能出 2 个分支,因此当某个非叶子节点是多个(2 个以上)的离散变量时,那么该变量就有可能被多次使用。

随机森林方法的基本流程如下:

(1)随机选择样本。假如给出一个数量为 N 的训练样本集,通过从训练样本集中拿出样本再放回去,如此反复地采样,直到得到K个样本($N=K$,K个样本中可能会存在相同的样本)构成一个新的训练集。利用新的训练集训练出一个决策树,作为决策树根节点处的样本集。

(2)随机选择特征。在建立决策树时,通过把每个特征的信息增益进行计算,选择信息增益结果里最大值的特征作为划分下一个子节点的走向。

(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)来进行分裂,一直到不能再分裂为止(并且决策树形成过程中没有进行剪

枝现象)。

(4)随机森林预测结果。通过步骤(1)~(3)的持续执行建立大量的决策树,进一步构成随机森林。把测试样本输入到随机森林中,利用对每一棵决策树的分类或者回归操作,得到最终的分类或者回归估测结果。

随机森林方法的主要优点:

(1)在测试集上的表现很好,由于样本以及特征都是随机选择的,因此随机森林不容易陷入过拟合。

(2)可以处理高维度数据,不需要进行特征选择,对数据集的适合能力强;处理对象可以是离散型数据,也可以是连续型数据,并且数据不需要进行规范化操作。

(3)在训练过程中,能够检测到特征间的相互影响且得出特征的重要性,具有一定的参考意义。

(4)每棵树都可独立、同时生成,容易做成并行化方法

(5)由于实现简单、精度高、抗过拟合能力强,当面对非线性数据时,适合作为基准模型。

2.2 梯度提升算法

梯度提升(Gradient Boosting)方法是一种较新的非参数机器学习方法[15],主要用于回归和分类问题的机器学习技术。其以弱预测模型(通常是决策树)集合的形式产生预测模型,目前,梯度提升方法在林业科学领域中的研究与应用相对较少。

Gradient Boosting 算法是一种可以使用任何损失函数(只要损失函数是连续可导的)的 Boosting 算法,其构建的模型抗噪音能力更强。Gradient Boosting 以弱预测模型(通常是决策树)集合的形式产生预测模型[15]。其在建立子树时,利用之前子树构建结果形成的残差作为输入数据,再构建下一棵子树。最终的估测按照子树构建的顺序进行估测,并将估测结果相加。Gradient Boosting 可以处理连续型数据和离散型数据,并且在相对少的调参情况下,模型的估测效果也会不错,模型的鲁棒性比较强。但由于各子树之间存在关联关系,难以并行训练模型。

2.3 Catboost 提升算法

Catboost 算法是由 Prokhorenkova L[16](Yandex 公司)在 2017 年首次提出的,设计的初衷是为了更好的处理梯度提升树(Gradient Boosting Decision Tree, GBDT)特征中的 categorical features。Catboost 采用的策略在降低过拟合的同时保证所有数据集都可用于学习,具有性能卓越、鲁棒性与通用性更好、易于使用而且更实用的优点。Catboos 的基本流程是先对所有样本进行随机排序,对每一条样本数据都会训练一个单独的模型 M(M 由不包含这条数据的训练集训练得到),依次类推,都累加到原来的模型上,得到最终的模型。

2.4 Stacking 集成学习算法

Stacking(有时也称之为 stacked generalization)是一种集成学习技术,通过元分类器或元回归聚合多个分类或回归模型[17]。Stacking 集成学习算法集成了各种不同的算法,较彻底地利用不同算法,从不同的数据空间和数据结构角度对数据进行不同估测,增强了算法模型的稳健性,得到的结果一般优于单一算法模型。该算法一般由两层组成:第一层为基础层次模型,第二层为元模型。基础层次模型是选择完整的训练集进行训练,元模型是基于基础层次模型的输出来进行训练。基础层次模型通常是由不同的学习算法组成的,因此集成通常是异构的。Stacking 先从初始训练集中基于各种不同的算法学习出初级学习器,然后生成一个新的数据集,用于训练次级学习器。在新数据集中,每个初级学习器对原始样本的预测标记被作为新样本的输入特征,而原始样本的原始标记被作为新样本的输出特征。

3 机器学习在森林蓄积量方面的研究进展

部分学者已尝试将机器学习算法应用于森林蓄积量估测。其中,杨柳等[18]以鹫峰林场森林为研究对象,利用 3 种机器学习方法(BP 神经网络、最小二乘支持向量机、随机森林方法)分别构建了森林蓄积量多光谱估测模型,最终结果显示采用随机森林方法建立的多光谱蓄积量模型的精度最高,为森林蓄积量遥感反演估测提供了一种新的方法。向安民等[19]对黑龙江省某林业局采用 K-近邻(K-Nearest Neighbor,KNN)方法进行森林蓄积量估测研究,与最小二乘估计和稳健估计建模进行对比,KNN 方法建立的森林蓄积量估测精度达到 97.3%,并且 KNN 方法能够有效克服建模变量间的复共线性问题。李圣娇等[20]的数据源为 Landsat8 影像,对香格里拉高山松森林蓄积量建立了偏最小二乘法遥感估测模型。

4 结束语

本文对森林蓄积量的研究进展以及 4 种机器学习算法做了简介,阐述了目前机器学习算法在森林蓄积量方面的研究进展。此外,由于当前森林蓄积量的研究重点是建立森林蓄积量估测模型,因此,本

文详细介绍了随机森林算法、梯度提升算法、Catboost 算法和 Stacking 集成学习算法的 4 种模型。尽管机器学习算法在其它领域已被广泛应用，但在林学研究邻域内，还有许多研究难点需要克服与探讨。本文认为随着机器学习算法的不断研究深入，其在森林蓄积量的研究、甚至是在林学研究领域将会取得更多成果和发展。

参考文献

[1] 王钰，周志华，周傲英. 机器学习及其应用[M]. 清华大学出版社，2006.

[2] MOHAMMADI J, SHATAEE S, BABANEZHAD M. Estimation of forest stand volume, tree density and biodiversity using Landsat ETM+ Data, comparison of linear and regression tree analyses[J]. Procedia Environmental Sciences, 2011, 7: 299-304.

[3] BREIDENBACH J, KUBLIN E. Estimating Timber Volume using Airborne Laser Scanning Data based on Bayesian Methonds[J]. Forest Science, 2009, 52: 611-622.

[4] YIM J S, KIM Y H, KIM S H, et al. Comparsion of the k-nearest neighbor technique with geographical calibration for estimating forest growing stock volume[J]. Canadian Journal of Forest Reasearch, 2010, 41(1): 73-82.

[5] 杨明星，徐天蜀，牛晓花，等. 基于 Sentinel-1A 雷达影像的思茅松林蓄积量估测[J]. 西部林业科学，2019，48(2)：52-58.

[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] 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+117.

[8] 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+86.

[9] 刘俊，毕华兴，朱沛林，等. 基于 ALOS 遥感数据纹理及纹理指数的柞树蓄积量估测[J]. 农业机械学报，2014，45(7)：245-254.

[10] 蔡学成，杨政熙. 基于中巴资源卫星数据的森林蓄积量估测研究[J]. 农业与技术，2013，33(12)：86-88.

[11] 张翔雨，王瑞瑞. 基于资源三号卫星遥感影像的森林蓄积量估测[J]. 湖北农业科学，2019，58(12)：74-78.

[12] 张苏，周小成，黄洪宇，等. 基于 SVR 的 GF1 号遥感影像森林蓄积量估测[J]. 贵州大学学报(自然科学版)，2019，36(3)：21-26.

[13] BREIMANL. Random forests, machine learning 45[J]. Journal of Clinical Microbiology.2001, 2: 199-228.

[14] 赵帅，李妍君，熊伟丽. 基于 KPCA-Bagging 的高斯过程回归建模方法及应用[J]. 控制工程，2019，26(1)：131-136.

[15] 张宏鸣，刘雯，韩文霆，等. 基于梯度提升树算法的夏玉米叶面积指数反演[J]. 农业机械学报，2019，50(5)：251-259.[16] PROKHORENKOVA L, GUSEV G, VOROBEV A, et al. CatBoost: unbiased boosting with categorical features[C]// Advances in neural information processing systems. 2018: 6638-6648.

[17] 史佳琪，张建华. 基于多模型融合 Stacking 集成学习方式的负荷预测方法[J]. 中国电机工程学报，2019，39(14)：4032-4042.

[18] 杨柳，冯仲科，岳德鹏，等. 结合纹理因子和地形因子的森林蓄积量多光谱估测模型[J]. 光谱学与光谱分析，2017，37(7)：2140-2145.

[19] 向安民，刘凤伶，于宝义，等. 基于 k-NN 方法和 GF 遥感影像的森林蓄积量估测[J]. 浙江农林大学学报，2017，34(3)：406-412.

[20] 李圣娇，舒清态，徐云栋，等. 基于偏最小二乘回归模型的高山松蓄积量遥感估测[J]. 江苏农业科学，2015，43(8)：182-185.

(上接第 157 页)

4　结束语

BIM、AR 和三维激光扫描技术在轨道交通上的合理结合运用，体现了信息化和数字化在轨道交通工程的重要应用价值。三维扫描技术的应用使模型建立更加高效、精准、便捷。同时 AR 技术的结合运用，可以增加用户和工程人员的体验感、交互感。

本文以实现轨道交通标准化、智能化管理的宏观把控和精益协调为目标，综合运用各项技术，并深入开展应用实践，借助对轨道工程的施工建设、后期运维阶段进行全方位的精准服务，实现了轨道交通复杂过程的虚拟预演、模块化维修及信息管理等功能。为轨道交通工程的全生命周期的智慧化管理带来切实可行的方案，提升了轨道交通工程的管理水平，推进了轨道交通信息化、数字化的进程，在工程应用中有着广阔的前景。

参考文献

[1] 王代兵，杨红岩，邢亚飞，等. BIM 与三维激光扫描技术在天津周大福金融中心幕墙工程逆向施工中的应用[J]. 施工技术，2016，46(23)：10-13.

[2] 胡跃军，罗坤，乔鸣宇. 基于 BIM 的智能建造技术探索[J]. 研究论文，2019(16)，52-53.

[3] 何建军，危鼎，姚守俨，等. 三维扫描技术结合 BIM 在佘山深坑酒店项目的应用[J]. 土木建筑工程信息技术，2015，7(4)：31-38.

[4] 黄阵仙，张爱青，方伟国. 基于建筑安全管理的 BIM+AR 技术应用研究[J]. 福建建筑，2019，9(255)：119-121.

[5] 刘照球，李云贵，吕西林，等. 基于 BIM 建筑结构设计模型集成框架应用开发[J]. 同济大学学报(自然科学版)，2010，38(7)：948-953.

[6] 张鹏，何东海. BIM 技术在幕墙工程中的应用[J]. 施工技术，2013，42(8)：105-106.

[7] 田飞腾. 基于安卓平台的增强现实导航系统设计与实现[D]. 西安建筑科技大学，2018.

[8] 刘剑锋. 增强现实技术在建筑设计中的应用研究[D]. 河南大学，2018.

[9] 清华大学 BIM 课题组. 设计企业 BIM 实施标准指南[M]. 中国建筑工业出版社，2013.

[10] 谭军，李哲林，姜立军，等. 增强现实技术在建筑设计仿真中的应用[J]. 吉林大学学报(工学版)，2013.

[11] 张启福，孙现申. 三维激光扫描仪测量方法与前景展望[J]. 北京测绘，2011(1)：394