



随机森林对人工林立地条件评价研究

参与人 林杰 张飞龙
汇报人 沈竞

College of Ecology and Environment, Nanjing Forestry University
(CEE, NJFU)

第八届全国生物多样性信息学研讨会（中国 新疆 中科院地理所）

随机森林对人工林立地 条件评价研究 汇报大纲

01

研究背景

02

研究区域及数据信息库

03

漆树人工林立地质量统计方程

04

随机森林建模过程

05

结论与讨论

01研究背景

一

种群生态学与人工智能

二

人工林立地分类与评价

三

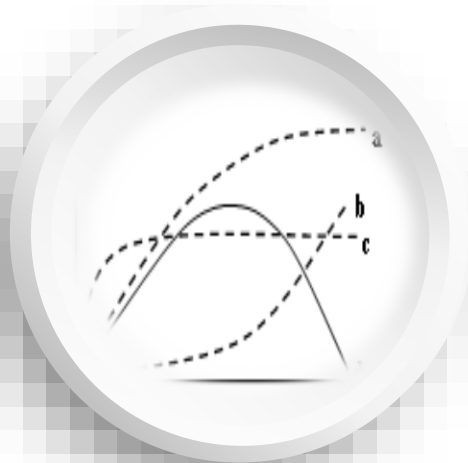
统计模型与立地分类评价

种群生态



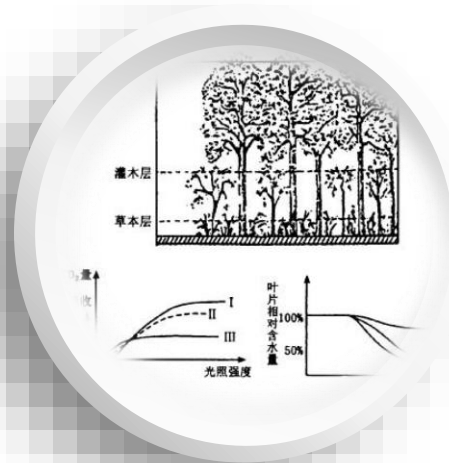
种群

种群是在一定的时间和空间内由同种个体组成的具有一定结构和功能的基本单位。



模型分析

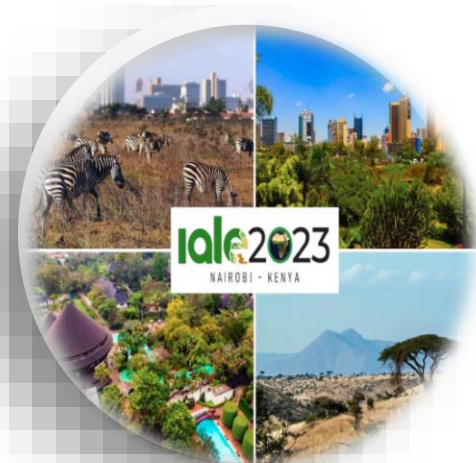
种群生态为生态学中的分支是
最多利用数学方法和模型



生态学

生态系统的恢复和重建提供理论支持和实践指导

立地条件



信息数据源（库）

人工智能

松线虫

害虫综合管理(IPM)研究和实施,以及昆虫种群生态学和系统生态学的发展,特别是IPM中系统分析和数学建模技术的广泛采用,极大地促进了计算机技术,特别是计算机模拟、数据库技术、网络技术等在害虫管理中的广泛应用。

植物保护

专家与决策系统基础的模拟(rule-based simulation),知识系统(knowledge systems),符号数学操作工具(symbolic mathematics manipulators),景观结构模拟工具(landscape simulators),目标模拟和行为模型化方法(object-oriented simulation and behavioral modeling)。

食物链

[Lotka-Volterra \(espadrine.github.io\)](https://github.com/espadrine/lotka-volterra)
算法仿真 (网络仿真)

杉木碳计量

通过测定一定数量样木、样地碳储量(生物量)的数据,建立模型后就可以在同类的林分中结合相应的林分调查数据来估计单株碳含量或是整体林分碳储量

森林资源动态监测

基于遥感智能视觉平台确定小班变化的地理位置
确定变化的数量与规模

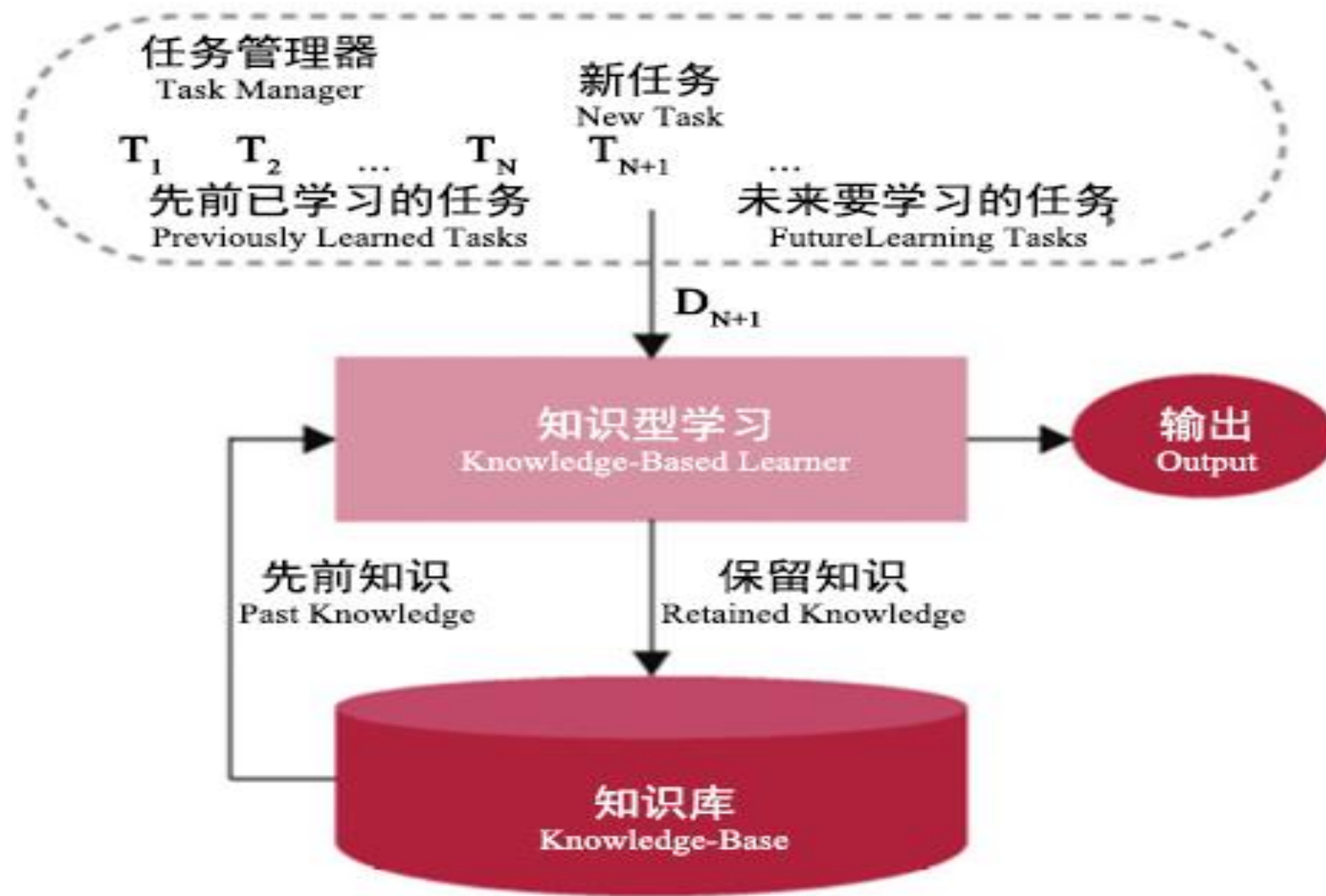


图1 人工智能分类思路

漆树人工林

植物种群生态学是研究植物种群的数量动态、特性、分化及与环境相互关系的科学,主要研究种群结构、种群动态、种群进化和适应对策等内容。

植物种群统计学

1. 种群密度
2. 初级种群参数包括 出生率、死亡率、迁入和迁出率
3. 次级种群参数 ----包括 年龄结构和分布、种群增长率等。

植物种群结构

种群结构是种群的基本特征,包括大小结构、年龄结构、空间结构和遗传结构。

植物种群动态

种群动态主要研究在一个生境中单个种的动、植物个体出生或发芽、成长及其死亡过程。。

植物组织器官间生物量分配

植物各器官间功能性状的关联,有助于确定控制功能性状的内在机制以及性状间的比例关系

现阶段植物种群生态发展

种群特征对城市大气、水体、土壤污染物的反应来**监测环境**
数据驱动的时代,准确地确定变量之间的**因果关系**

材积生长量和蓄积生长量作为立地质量的主要评价因素

林木的高度是反映林木生长状况的重要数量指

直接评价法是直接利用林分的收获量或者生长值来评价立地质量。又可以被划分为根据材积生长量或蓄积生长量评定立地质量和依据林分高来评定立地质量两种方法。材积生长量和蓄积生长量作为立地质量的主要评价因素，可直接作为经济和社会

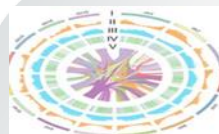
间接评价法

- 依据上层木之间的相关性评价立地质量（生长型与主要树种相关）
- 利用下层植被特性评价立地质量（群落对立地的作用）
- 通过**立地因子评价立地质量**（树木生长与立地因子联系 数量转换与函数方程）



人工林立地

人工林立地环境因子与上层木平均高的测定值及落叶松地位指数表与密度控制图



立地质量(Site quality)

对于森林的管理和经营来说，是十分重要的指标,其基本含义是指在某一立地上既定森林或者其他植被类型的生产潜力。常常用立地指数（site index）和地位级（site class）等其它指标来表示优劣差异

立地

可决定这个地区所能生长的所有植被的类型和性质森林立地学是一门识地用地的应用科学，立地是从环境的角度来考虑一个地区，

立地类型(Site type)

立地分类当前公认的**基本单位**，其概念主要是指对于某一造林目的的树种，具有相同的宜林程度和宜林性质，用同样的经营管理措施，**林木在生长过程、生长率**等方面趋向一致的所有地段的综合

立地分类（Site classification）

是依据立地特征的差异性和相似性将立地综合体详细的划分为不同的类别和等级[8]。而在同一类别和等级中的立地综合体就具有相同的生产能力和生产潜力，因此也就能够以相同的方式进行抚育和管理。

立地质量统计评价方法

统计系统分类与评价，哑变量回归和混合效应模型已经在林业上有了一些成功的应用



立地分析对漆树人工林的**选址栽植**具有一定的指导意义

在植树造林阶段避开潜在的立地质量较低的区域,选择潜在立地质量较高的区域,

使漆树生长的环境尽可能接近最佳状态,从而取得较好的经济效益.

随机森林对人工林立地条件评价研究

汇报大纲

01

研究背景

02

研究区域及数据信息库

03

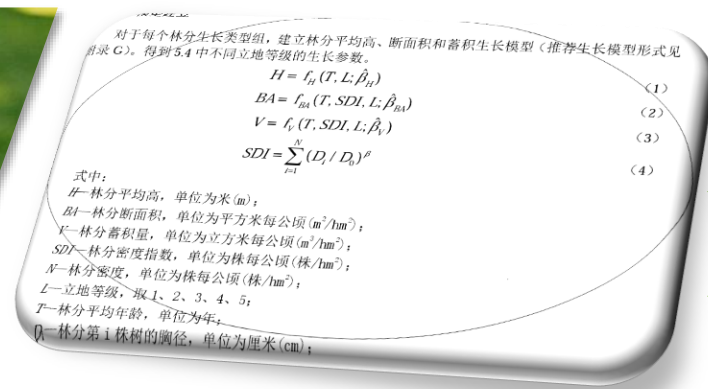
漆树人工林立地质量统计方程

04

随机森林建模

05

结论与讨论



样地设置与内容

调查与分析的内容：植被、地形、土壤
取样、
植物叶片取样
面积、编号、个数

研究区域概况

自然地理环境、地质地貌、气候条件、土壤条件、植被概况

每木调查

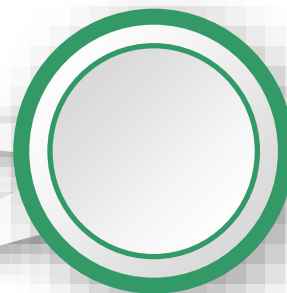
个体坐标位置

经纬度、海拔、坡度、坡向



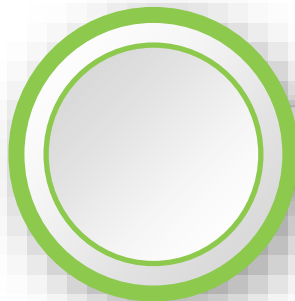
环境因子

土壤因子



树木特征

树高、胸径、冠幅
径级：胸径5cm为一个径级



环境因子

地形因子



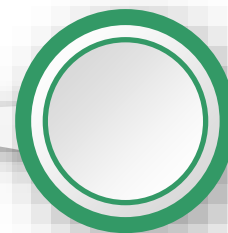
种群（物种）组成

空间格局与种间关联



树叶叶片采集

叶片元素分布于土壤养分



习性

漆树为漆树科漆树属双子叶植物，原产于中国和印度次大陆]，漆树立地评价对充分挖掘地力、树木生产潜力，促进林业向高产、优质、高效益发展有着重要的现实意义]。

立地评价

漆树立地评价还可以作为当地林业部门指导生产、实行科学管理的重要依据和手段，可以对当地生产潜力，建立合理的经营结构，从而为区域长远的造林目标提供可靠依据

因地制宜

漆树所产的生漆、籽油、木材、漆蜡等具有较高的利用价。在平利县漆树造林的实践中，一直存在死亡率过高的问题，因地制宜、适地适树是降低漆树死亡率的关键



样地

平利县漆树人工林大多为7-30年生，本次样地调查时间为2018年6月和2020年8月，共设置了70个样地

样方—漆树调查

漆树分布零散，故样方面积设为

15m×15m。样地设置原则为：土壤类型全覆盖、地貌类型全复盖且样点基本均匀分布

利用GPS记录样方的地理位置，坐标参考为WGS-84，记录的数据有漆树的胸径、树龄、海拔、坡向、坡度等因子

土壤采样及分析

每个样地内按对角线方向设置3个采样点，每个采样点在0~30cm深度内取样约500g，不同样点间土样充分混合后放入自封袋，所取样品混合均匀后带回。土壤样品经过风干、去杂、碾碎，分别过100和300目数筛，采用四分法称取，供土壤理化性质分析。同时在3个采样点处挖土壤剖面，并用100cm³环刀在每点取原状土供土壤物理性质测定。

土壤样品中除全氮（N）和有机质（OM，由全碳所测值换算）利用碳氮分析仪测定外，其余理化性状指标采用国家林业行业标准《森林土壤测定方法》LY/T1215—1275—1999中相关测定项目的方法进行测定。

树芯采样及分析

对样地内的漆树每木检尺，选出胸径最大的两株漆树并用生长锥钻取木芯，测量树龄，计算年平均胸径生长量，将生长量年平均胸径生长量平均值的样地定义为低立地质量，其余样地则定义为高立地质量。

图2 研究区和采样点分布

随机森林对人工林立地条件评价研究 汇报大纲

01

研究背景

02

研究区域及数据信息库

03

漆树人工林立地质量统计方程

04

随机森林建模

05

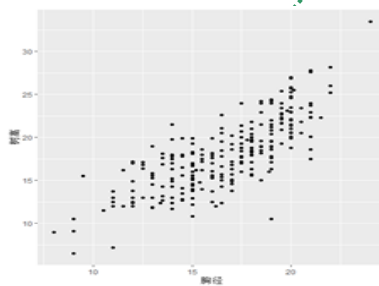
结论与讨论

漆树人工林立地评价



70块样地

收集和漆树生长有关的地形因子,植被因子和土壤养分因子



线性混合效应模型

对胸径-年龄数据作散点图,当地的胸径-树龄呈现线性关系,

$$h_{ij} = \gamma_{00} + \gamma_{10}d_{ij} + U_{0j} + U_{1j}d_{ij} + \varepsilon_{ij} \quad (3)$$

基于线性混合效应模型的胸径生长方程

将样地划分为四种立地类型,分别为坡度二平地,坡度三坡下,坡度四坡下,坡度二坡上;对四种立地类型引入随机效应,或者分别对坡度和坡向引入随机截距或随机斜率,可以衍生出不同的混合效应模型. 所有模型的拟合使用 R 软件中的 nlme 扩展包完成

决定系数
(R²)
均方根误差
(RMSE)两个
指标,

模型检验

AIC 代表赤池信息量准则 (英语: Akaike information criterion, 简称 AIC), 是评估统计模型的复杂度和衡量统计模型“拟合”资料的优良性

一种是将坡向和坡度结合起来,视为一个随机效应

随机效应设置

是将坡向和坡度视为两个随机效应,对模型中的随机效应引入随机斜率或随机截距,剔除不能收敛的无效模型

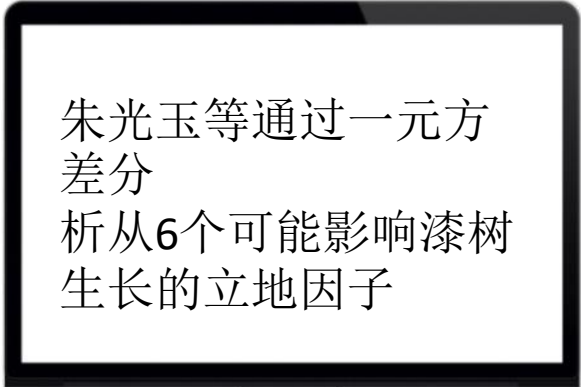
Model [↗]	自由度 [↗]	AIC [↗]	BIC [↗]	logLik [↗]	似然比 [↗]	p-value [↗]	R ²	↗ RMSE [↗]
基础模型 [↗]	3 [↗]	1174.121 [↗]	1184.601 [↗]	-584.061 [↗]	19.92559 [↗]	<0.001 [↗]	0.604 [↗]	2.5945 [↗]
M1 [↗]	4 [↗]	1156.196 [↗]	1170.168 [↗]	-574.098 [↗]	3.760532 [↗]	0.15255 [↗]	0.642 [↗]	2.4536 [↗]
M2 [↗]	6 [↗]	1156.435 [↗]	1177.394 [↗]	-572.218 [↗]	13.86782 [↗]	<0.001 [↗]	0.649 [↗]	2.4172 [↗]
M3 [↗]	4 [↗]	1166.303 [↗]	1180.275 [↗]	-579.152 [↗]	8.708968 [↗]	<0.05 [↗]	0.626 [↗]	2.5089 [↗]
M4 [↗]	6 [↗]	1161.594 [↗]	1182.553 [↗]	-574.797 [↗]	NA [↗]	NA [↗]	0.642 [↗]	2.4407 [↗]
M5 [↗]	6 [↗]	1151.695 [↗]	1172.653 [↗]	-569.848 [↗]	NA [↗]	NA [↗]	0.659 [↗]	2.3643 [↗]

图3 不同混合效应模型拟合结果

将坡度随机参数加入到基础模型中,只考虑随机截距,构造模型 1(M1),在 M1 的基础上考虑随机斜率,构造模型 2(M2);将坡向随机参数加入到基础模型中,只考虑随机截距,构造模型 3(M3),在 M3 的基础上考虑随机斜率,构造模型 4(M4);将坡度和坡向因子合并为一个随机因子,加入到基础模型中,并考虑随机斜率和随机截距,构造模型 5(M5).M1 和基础模型的似然比极显著($P<0.001$),说明加入了混合效应的M1表现优于基础模型.在以上5种混合效应模型中,M5 的 AIC 最小,说明 M5 的拟合优良性相比其他四个模型更好.M5 的 RMSE 为 2.3643,11说明 M5 的误差最小,相比基础模型降低了 8.9%;决定系数为 0.659,相比基础模型提高了 5.5%.

混合效应模型可以解释 66%的变异性,还有 34%的变异性未能解释,可能存在的原因如下:

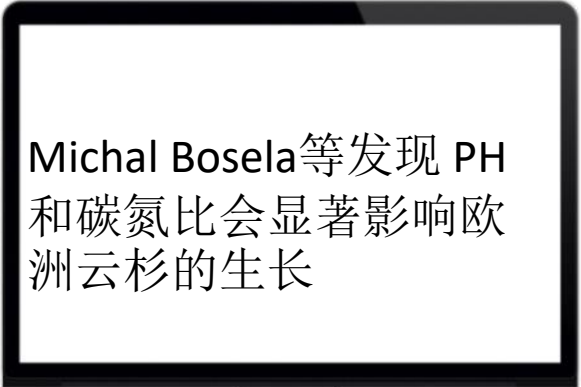
(1)场地测量和采样误差;(2)缺失了能够解释变异的立地因子;(3)采样数量不够充足;(4)基因变异;(5)人为造林.



朱光玉等通过一元方差分析从6个可能影响漆树生长的立地因子

海拔

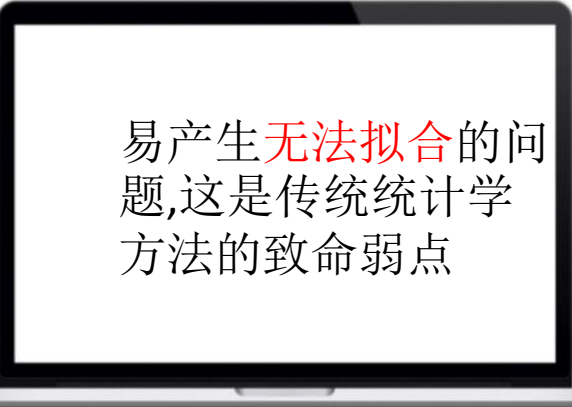
海拔是影响树木生长的一个重要因素
选取的样地海拔范围较广,每隔 100m 分一个等级,一共 12 个等级,而本次研究的 8 个样地海拔相近,不能体现出海拔的差异,故无法研究海拔对当地漆树生长的影响。(因子使用不充分)



Michal Bosela等发现 PH 和碳氮比会显著影响欧洲云杉的生长

土壤因子

土壤因子(如 PH,容重,碳氮比等)也是影响树木生长不可忽视的因素
在8块样地随机选择了24个样点测量了土壤指标,各项指标的变动范围极小



易产生无法拟合的问题,这是传统统计学方法的致命弱点

传统统计的缺点

哑变量模型中引进新的哑变量
过拟合的问题

随机森林对人工林立地条件评价研究

汇报大纲

01

研究背景

02

研究区域及数据信息库

03

漆树人工林立地质量统计方程

04

随机森林建模过程

05

结论与讨论

04随机森林建模过程

一

机器学习与立地分析

二

随机森林基本理论

三

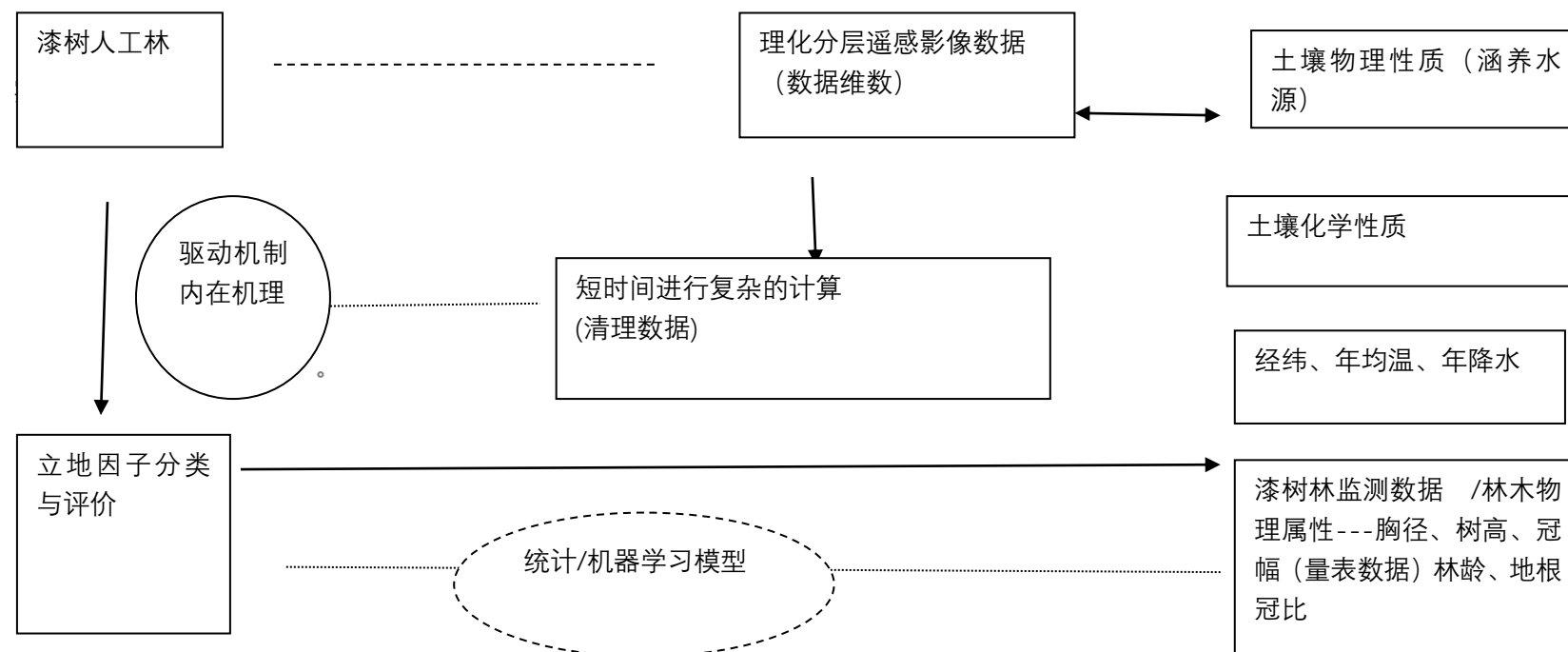
随机森林进行漆树人工林立地分析

研究基础：

- 1) 真实的数据集
- 2) 分类的方法

。

机器学习与立地分析



结合近年来较为先进的机器学习方法，构想漆树的立地质量评价：

第一步：比较常用的 6 个树高曲线拟合效果的基础上，将各个影响漆树生长的立地因子纳入树高曲线中

第二步：通过改进的含有立地因子的树高曲线模型，预测基准年龄下各区域漆树的立地指数

第三步：景观生态专题类型图----通过反距离权重插值法插值到整个区域，并设定立地指数区间，对立地质量进行分级，反映漆树立地质量的空间分布，为漆树的潜在立地生产力的评价提供依据

第四步：超参数选择

随机森林算法是一种基于决策树的集成学习算法。假设有N个训练样本, 而M表示训练样本所用的特征数目, 随机森林的基本原理和精度评价

训练集
择特征数目 μM 来确定决策树上一节点决策结果, 其中 μ 为特征选取比例, 从N个训练样本中放回抽样N次, 形成训练集, 用未抽到的样本作预测并评估误差。

袋外数据(OOB, Out-Of-Bag)

测试数据, 可以对随机森林的准确性得分(accuracy score)进行估算



$$accuracy(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n 1(\hat{y}_i = y_i)$$

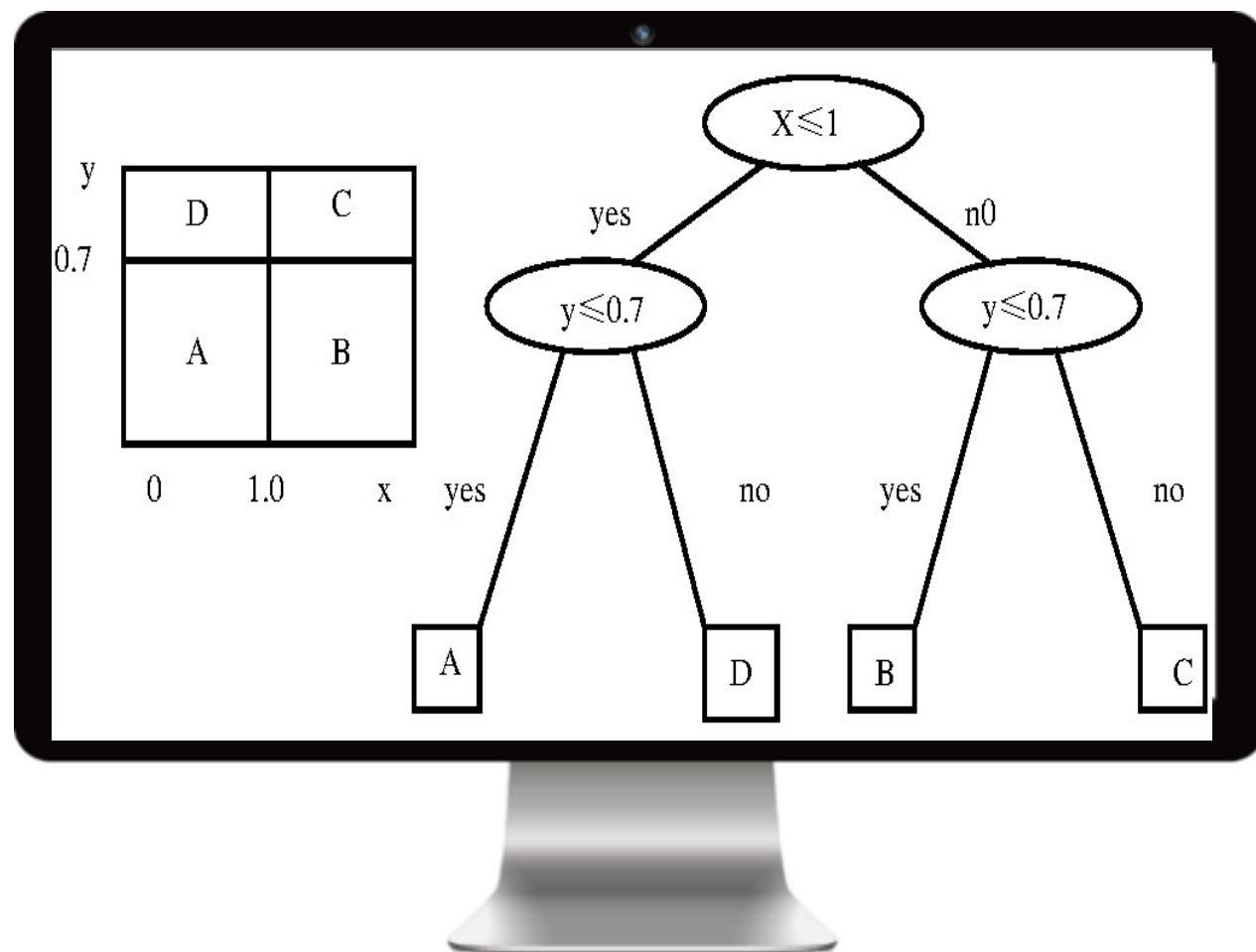
每棵决策树正常成长, 即没出现剪枝

$$MDG = \frac{1}{M} \sum_{m=1}^M \sum_{t \in \varphi_m} 1(j_t = j) D_{Gj}$$

$$Gini(t) = 1 - \sum_{j=1}^k [p(i|t)]^2$$

对分类树节点作t分割, 计算使用立地因子j前与使用后Gini系数的减小值(D_{Gj}), 对所有节点的 D_{Gj} 求和后再对所有分类树M取平均, 得到平均基尼指数降低程度(mean decrease Gini, MDG)。

随机森林



单棵树生长方法

生长单棵分类树的原则是递归分区。最简单的树是二叉树,即树中每个节点最多有两个分支节点

自助采样

在统计量重采样技术中,一种新方法是自助法 (bootstrap)。自助法是从原始的样本容量为 N 的训练样本集合中随机抽取 N 个样本生成新的训练样本集

随机森林生成

随机森林是通过自助法重采样技术,生成很多个树分类器

变量重要性评价

$p(i|t)$ 为类别 i 在节点 t 处的概率, k 为分类结果数,在本研究中取值为 2.

随机森林的基本
原理和精度
评价



变量重要性评价



反距离权重插值法



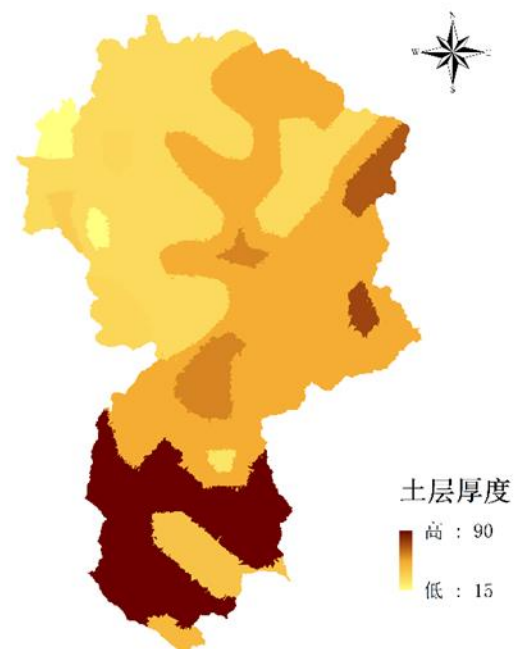
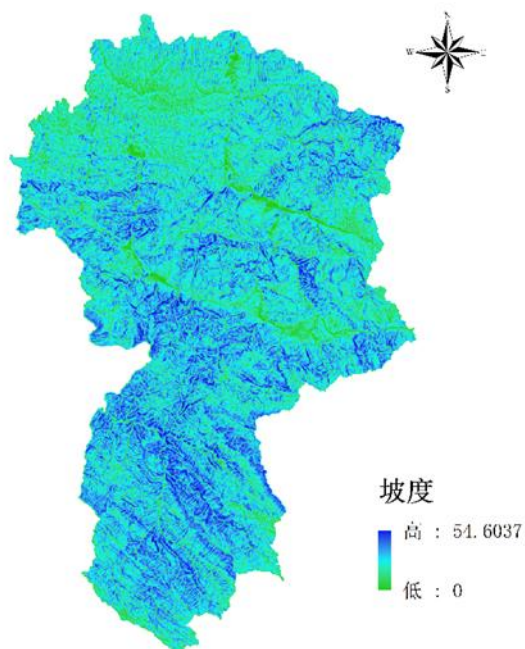
随机森林模型验证



- 根据陕西省平利县的地形,植被,土壤养分特点,以漆树为研究对象,选取了覆盖整个研究区域70个样地,并从可能影响漆树生长的立地因子中选取了9个较为具有代表性的因子
- 较为先进的随机森林算法引入到漆树的立地质量评价中
- 通过调整参数明显地提高了模型的准确性和稳定性

$$accuracy(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n 1(\hat{y}_i = y_i)$$

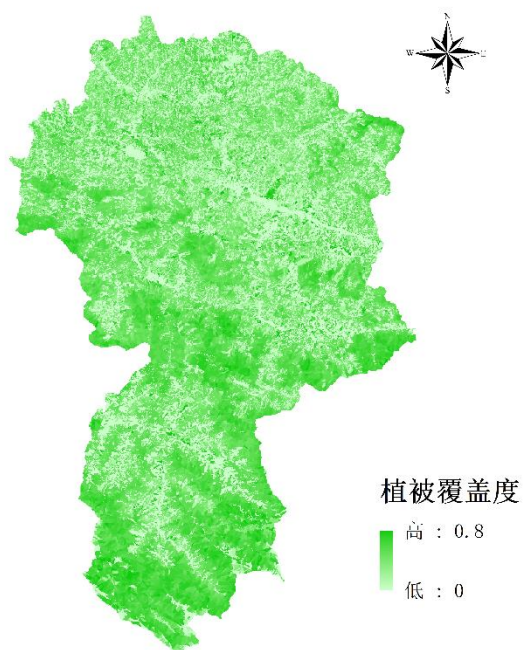
随机森林的基本 原理和精度评价



变量重要性评价

$$MDG = \frac{1}{M} \sum_{m=1}^M \sum_{t \in \varphi_m} 1(j_t = j) D_{Gj}$$

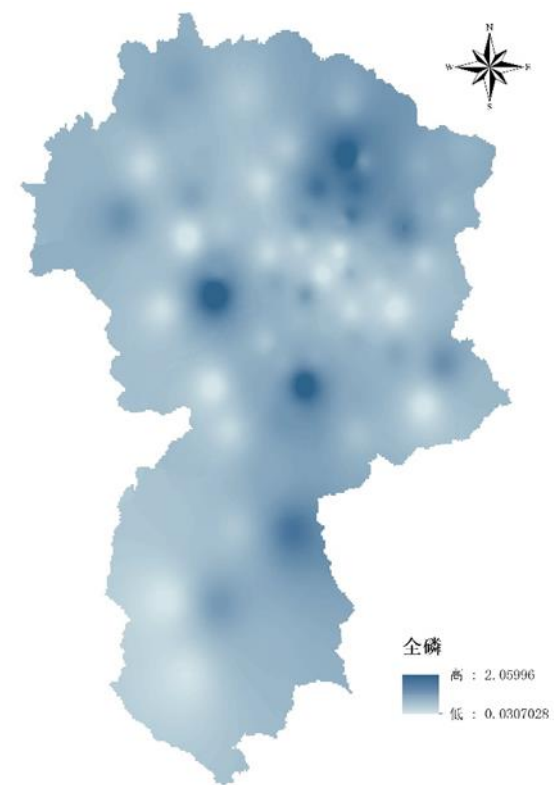
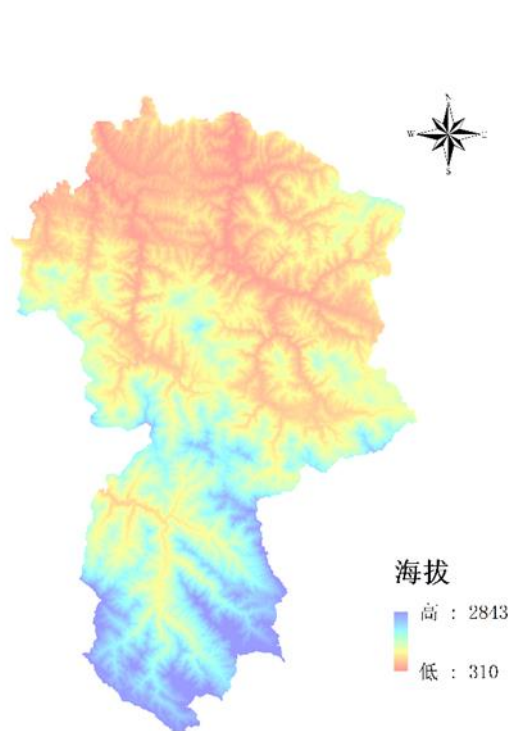
$$Gini(t) = 1 - \sum_{i=1}^k [p(i|t)]^2$$



反距离权重插值法

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i z(s_i)$$

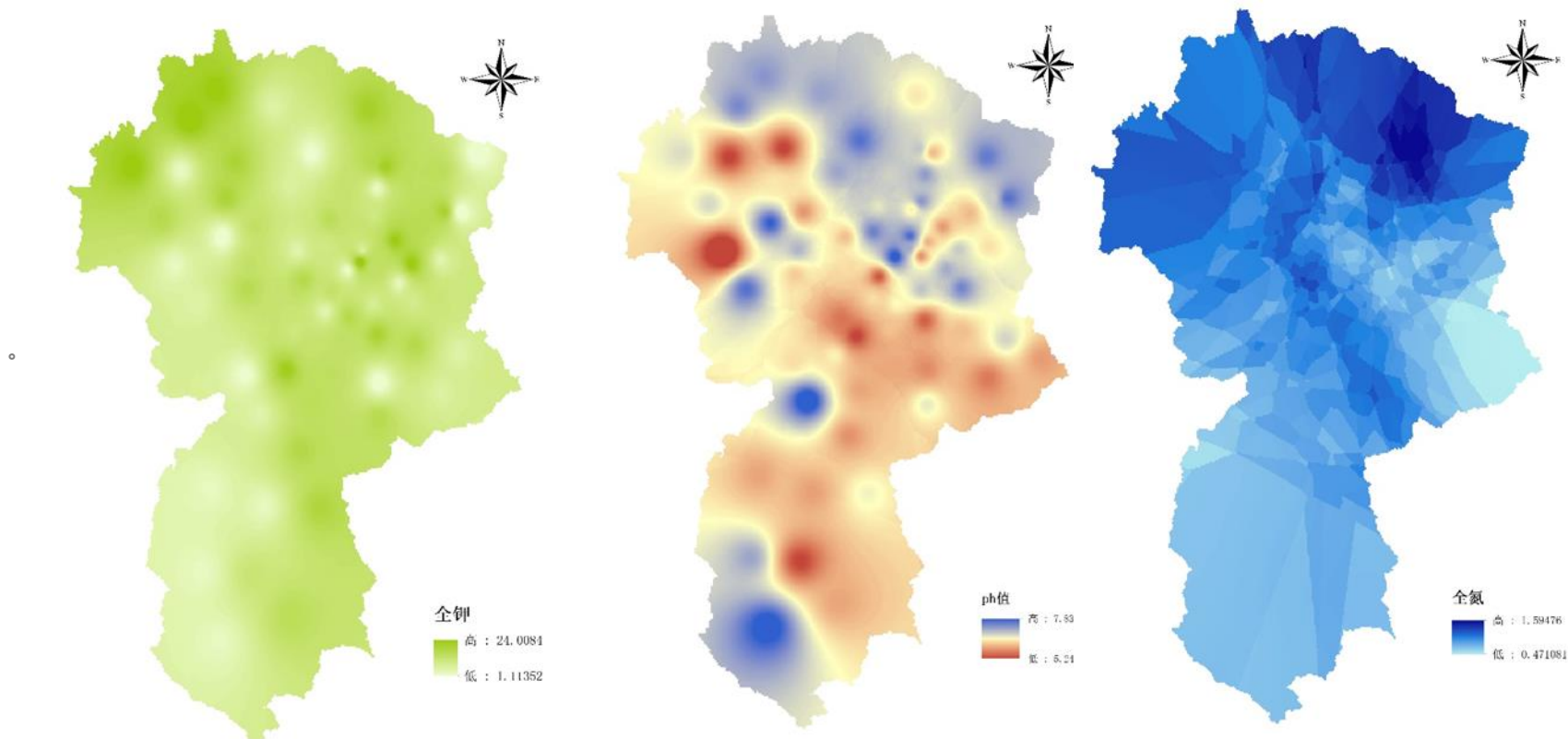
$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^N d_{i0}^{-p}, \sum_{i=1}^N \lambda_i = 1$$



反距离权重插值法

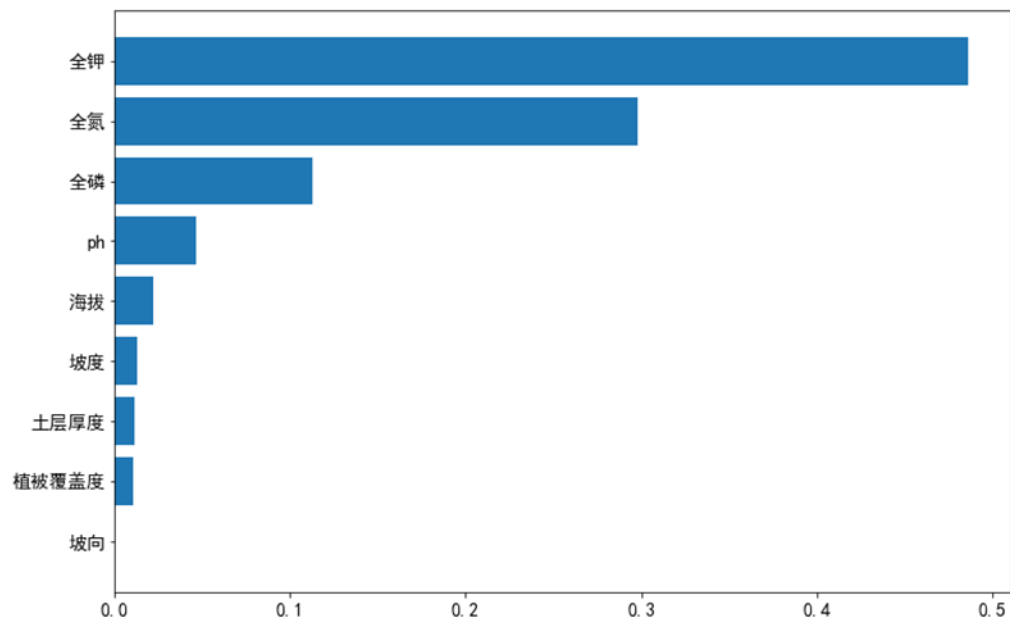
$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i z(s_i)$$

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^N d_{i0}^{-p}, \sum_{i=1}^N \lambda_i = 1$$



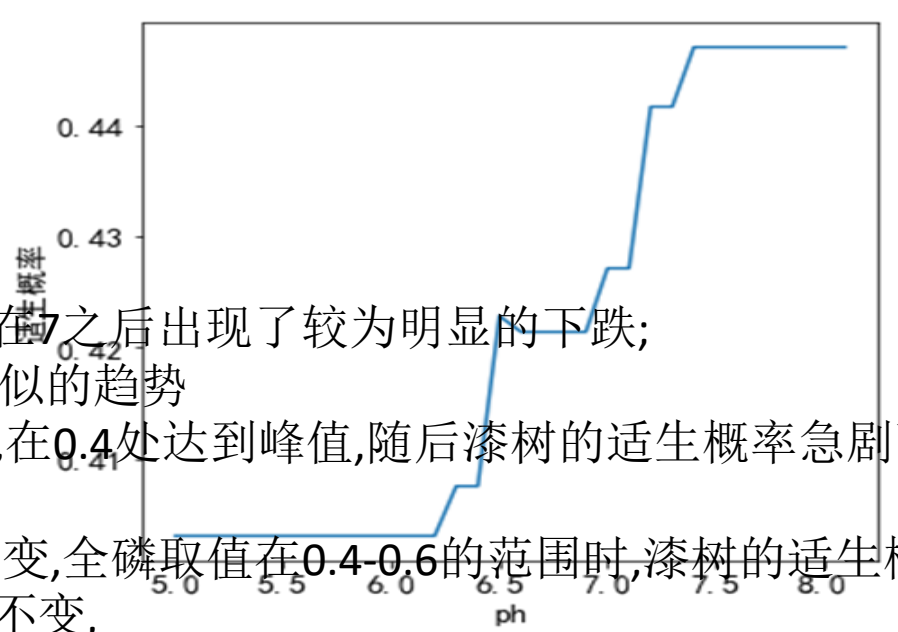
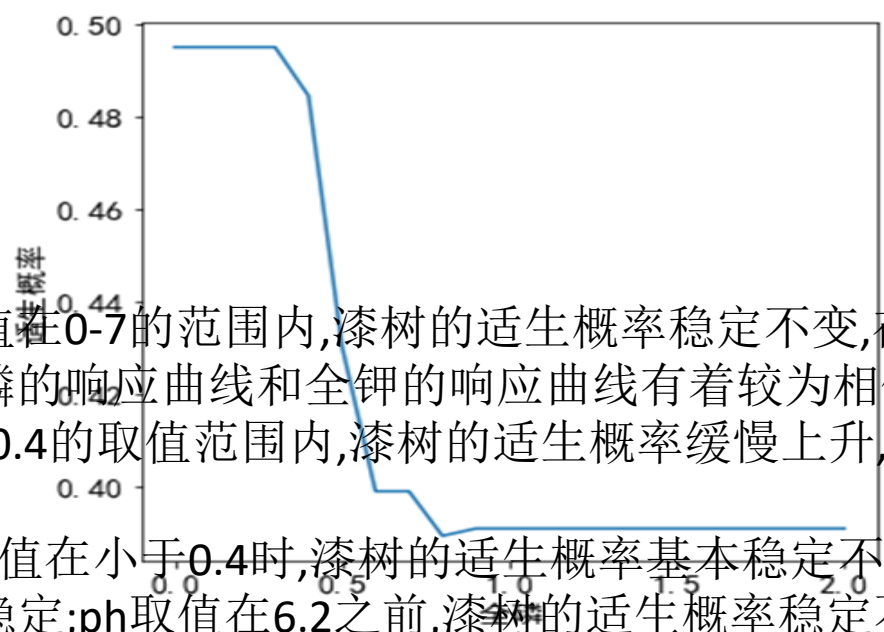
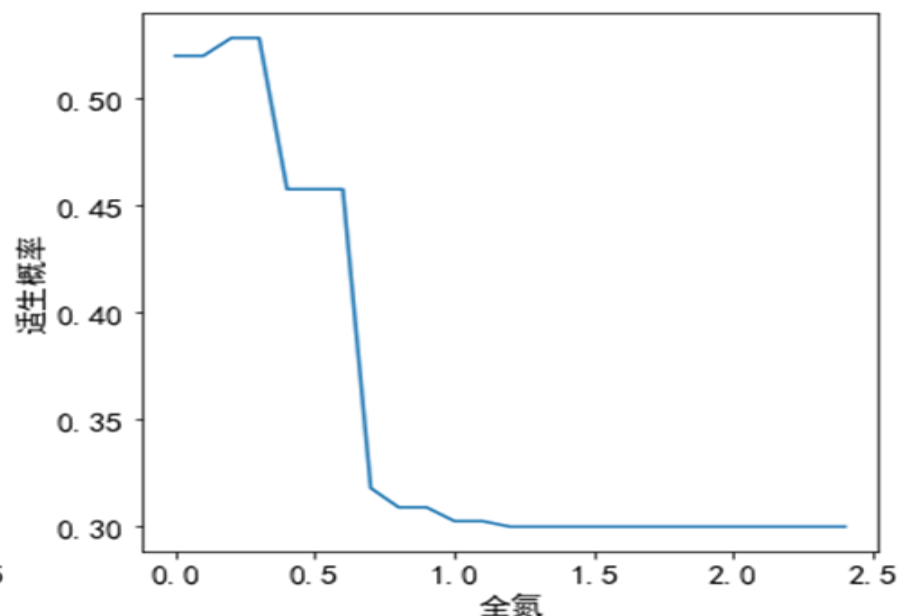
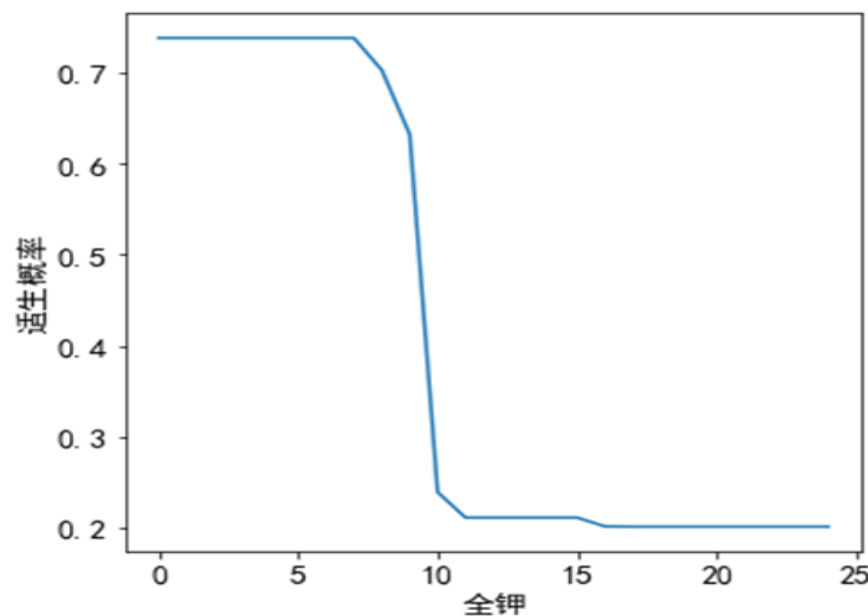
随机森林模型验证

立地因子重要性评估



重要性的值越大,说明该环境因子对漆树立地质量的影响越显著

对漆树立地质量影响较大的前四个因子,通过随机森林模型可以计算出单个环境因子不同取值下对应的漆树适生概率,漆树的适生概率越高,说明此时环境因子的取值越适合漆树生长,进而判断单个环境因子的最佳取值范围。
描绘了单个环境因子变动对漆树立地质量的影响



- 全钾的取值在0-7的范围内,漆树的适生概率稳定不变,在7之后出现了较为明显的下跌;
- 全氮和全磷的响应曲线和全钾的响应曲线有着较为相似的趋势
- 全氮在0-0.4的取值范围内,漆树的适生概率缓慢上升,在0.4处达到峰值,随后漆树的适生概率急剧下跌并在1之后趋于平稳;
- 全磷的取值在小于0.4时,漆树的适生概率基本稳定不变,全磷取值在0.4-0.6的范围时,漆树的适生概率急剧下跌,并在0.6之后趋于稳定;ph取值在6.2之前,漆树的适生概率稳定不变,
- ph取值在6.2-7.5时,漆树的适生概率明显上升,并在7.5之后趋于稳定.根据响应曲线的趋势,漆树立地质量高的适宜值为:全钾取值7,全氮取值0.4,全磷取值0.4左右

实验的样本数较少,所以**没有**划分训练集和测试集利用袋外数据的准确性得分来衡量模型的精度,这种方法**适用于样本数较少**的情

衡量模型的泛化精度,**避免过拟合**

在没有调整模型参数的情况下,模型的准确性得分为0.73;利用网格搜索法逐个尝试可能的参数取值,根据模型准确性得分的变动情况**调整模型参数**。发现在决策树数量(`n_estimators`)为50,决策树最大深度(`max_depth`)为5,随机抽取的**最大特征数量(`max_features`)为2**时,模型的分类准确度提高到0.92,继续增大决策树数量,决策树最大深度,最大特征数量,模型的准确性没有提升,甚至出现了大幅度降低的情况

参数取值



建立了最终的随机森林模型

实验结果表明: 该模型的泛化精度达到了76%,具有较高的分类准确率,在立地因子已知的情况下,可以对漆树的立地质量进行预测

随机森林对人工林立地条件评价研究

汇报大纲

01

研究背景

02

研究区域及数据信息库

03

漆树人工林立地质量统计方程

04

随机森林建模过程

05

结论与讨论

平利县漆树立地质量空间分布（景观专题生态类型图）

漆树的适生概率分为3个立地质量等级

- 绿色----较高 ----一级 (0.5-0.76)
- 灰色-----中等 ----二级 (0.2-0.5)
- 黄色-----较低 ----三级 (0-0.2)

区位特征

- 一级地区仅占整个区域的2%,二级地区占整个区域的50%,三级地区占整个研究区面积的48%
- 从图中可以看出低立地质量区域和中等立地质量的区域均匀地分布在整個研究区域,较高立地质量区域的分布较为零散,大部分位于平利县北部和中部地区,少量位于平利县南部地区.



平利县漆树高立地质量地区集中在平利县中部和北部,南部仅有极少部分地区立地质量较高,平利县南部以山地为主,北部以丘陵地区为主,是全县农业生产的精华之地,因此平利县北部地区的土壤养分状况优于平利县南部地区,根据前人研究N,P,K肥可以提高漆树的产量,因此本研究中随机森林对漆树立地质量空间分布的预测结果,与平利县的实际情况相吻合,但本研究得到的平利县中部和北部地区的适宜区面积偏小,这可能与采样点的位置和数量有关.

随机森林模型可以计算立地因子的重要性,给出相应的权重,研究表明,对漆树立地质量影响较大的立地因子由大到小排序依次是全钾,全氮,全磷,ph,海拔,坡度,土层厚度,植被覆盖度,坡向.其中后六个立地因子的权重不到10%,相对于土壤因子对漆树立地质量影响较小.全钾,全氮,全磷是对漆树立地质量影响最大的三个因子

与已有研究相符合 刘崇欣的研究表明N,K,P肥可以提高漆树的产量,研究结果和实际情况相符.王秦生的研究指出,在年平均气温8°C以上,年降雨量600毫米以上,相对湿度60%以上的微酸性,中性和微碱性的土壤上可以正常生长,说明漆树对土壤ph的适应范围较广

董波等人的研究表明,海拔是影响漆树立地质量的主要立地因子之一,且漆树的分布多集中于400-2000m,而平利县的海拔范围是310-2843m,且绝大部分地区的海拔处在400-2000m内,在这个范围内的漆树立地质量相差不大,因此海拔的因子重要性相对较低,此外,何忠祥的研究表明,不同漆树品种的适宜海拔不同,平利县根据漆树的生物学特性,在不同的海拔区间种植了不同品种的漆树,例如大红袍适宜在海拔1000m以下的地区栽植,而红皮高适宜在海拔1500m以下栽植,本研究采样的漆树可能涉及到不同的漆树品种,从而使海拔对漆树立地质量的影响没有明显体现;

关于土层厚度,植被覆盖度对漆树立地质量的影响

(数据可以进一步补充)

研究区域尺度较小,而世界气候数据网站worldclim的气象数据分辨率过大,在研究区范围内无法保证精度,没有考虑各种气象因子对漆树立地质量的影响,考虑扩大研究区尺度,并收集当地气象站的详细数据或采用混淆矩阵的

因子选择

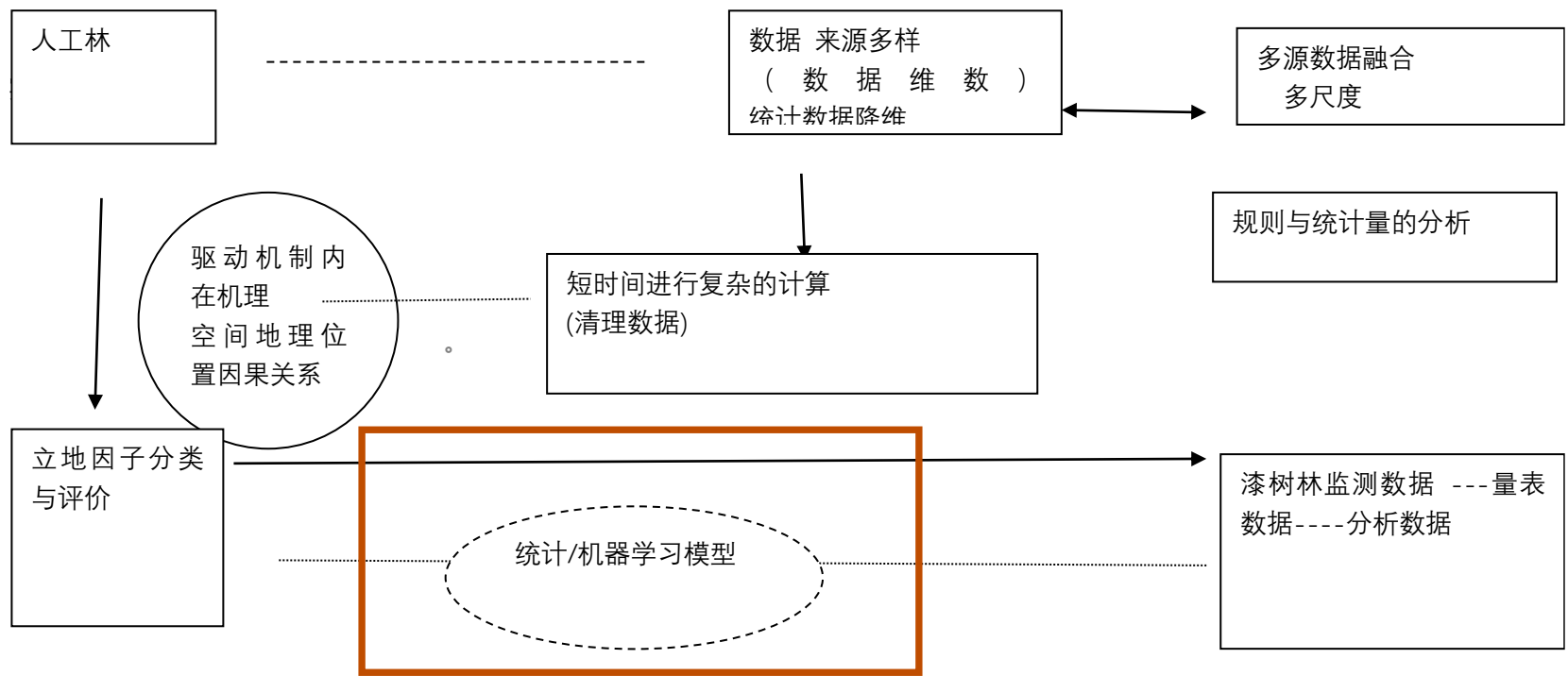
改进的随机森林模型

全氮,全钾,pH的栅格图层是基于前人在70余个样地的调查结果插值得到的,样地的数量会显著影响漆树立地质量空间分布的预测结果,

- 样地数量仍然较少,需要通过后续的大规模采样使预测精度得到进一步提升;随机森林算法在本研究中解决回归问题时的表现不佳,预测胸径年平均生长量时的袋外精度仅为0.5,存在明显的过拟合现象,改为分类问题之后,精度得到了显著提升;

随机森林属于黑箱模型,只能通过不断调整模型提高预测精度,却很难获取模型内部的运行情况===统计与机器学习的结合

人工智能助力生物（态）信息学





南京林业大学
NANJING FORESTRY UNIVERSITY

感谢 😊

COLLEGE OF ECOLOGY AND ENVIRONMENT, NANJING FORESTRY UNIVERSITY
(CEE, NJFU)

