doi:10.6041/j.issn.1000-1298.2020.09.022

基于图像处理和 GBRT 模型的表土层土壤容重预测

杨　玮　兰　红　李民赞　孟　超
（中国农业大学现代精细农业系统集成研究教育部重点实验室，北京 100083）

摘要：针对传统的表土层土壤容重测量方法费时、耗力的问题，利用易获得的土壤物理参数实现农田大范围表土层土壤容重的快速、准确预测。通过分析表土层土壤容重与土壤表面粗糙度、土壤阻力的关系，构建了以土壤表面粗糙度和土壤阻力为输入的 GBRT 模型，土壤表面粗糙度利用图像处理技术获得，土壤阻力使用实验室车载式阻力测量系统获得。使用同态滤波技术对土壤表面图像进行预处理，提取图像灰度直方图的熵、平均值、方差、偏度和峰度表征图像的纹理特征参数，提取图像灰度共生矩阵的能量、熵、对比度和逆方差表征图像的区域特征参数。利用灰度关联分析，从 9 个表征土壤表面粗糙度的特征参数和土壤阻力中选取与表土层土壤容重关联度大于 0.65 的变量作为模型输入，将得到的 GBRT 模型预测结果与环刀法得到的结果进行相关性分析，R^2 达到 0.878 2，平均绝对误差达到 0.021 g/cm^3。同时在相同的输入参数和运算环境下，与 BPNN 和 SVR 模型的预测精度和运算速度进行了对比，验证得到 GBRT 模型具有更高的预测精度和更短的运算时间。本文研究结果为科学指导农田表土层土壤容重的获取提供了思路。

关键词：土壤容重；图像处理；梯度提升决策回归树模型；预测；土壤表面粗糙度；土壤阻力
中图分类号：S152.5　　**文献标识码**：A　　**文章编号**：1000-1298(2020)09-0193-08　　**OSID**：

Prediction of Top Soil Layer Bulk Density Based on Image Processing and Gradient Boosting Regression Tree Model

YANG Wei　LAN Hong　LI Minzan　MENG Chao
(*Key Laboratory of Modern Precision Agriculture System Integration Research*, *Ministry of Education*, *China Agricultural University*, *Beijing* 100083, *China*)

Abstract: Aiming at the time-consuming and labor-intensive problem of traditional soil bulk density measurement of topsoil, using easily available soil physical parameters to accurately and quickly predict the bulk density of topsoil in farmland. By analyzing the relationship between soil bulk density of topsoil layer and surface roughness and resistance of soil, a gradient boosting regression tree (GBRT) model with input of surface roughness and resistance of soil was constructed. The roughness of soil surface was obtained using image processing techniques. Using homomorphic filtering technology to preprocess the surface image of soil, extract the entropy, average, variance, skewness and kurtosis of the image gray histogram to characterize the texture feature parameters of image, extract the energy, entropy, contrast and inverse variance characterize the regional characteristic parameters of the image. The soil resistance was obtained using a laboratory vehicle-mounted resistance measurement system. Using gray correlation analysis, from nine characteristic parameters that characterizing the roughness of soil surface and soil resistance, the variables with bulk density of topsoil greater than 0.65 were selected as the model input. The prediction results of the GBRT model were the same as those obtained by the ring knife method. As a result of correlation analysis, the determination coefficient R^2 reached 0.878 2, and the average absolute error reached 0.021 g/cm^3. At the same time, under the same input parameters and computing environment, compared with the prediction accuracy and operation speed of the BPNN and SVR models, it was verified that the GBRT model had better prediction accuracy and shorter operation time. The research results can provide ideas for obtaining the bulk density of topsoil and provide theoretical support for scientific and rapid guidance of farmland.

Key words: soil bulk density; image processing; GBRT model; prediction; soil surface roughness; soil resistance

收稿日期：2020-05-07　修回日期：2020-06-24
基金项目：国家自然科学基金青年基金项目(31801265)
作者简介：杨玮(1981—)，女，副教授，博士生导师，主要从事精细农业系统集成研究，E-mail：cauyw@cau.edu.cn
通信作者：李民赞(1963—)，男，教授，博士生导师，主要从事精细农业系统集成研究，E-mail：limz@cau.edu.cn

0 引言

土壤容重又称为土壤密度,是指在自然结构状态下单位体积干土的质量,是表征土壤物理性质的重要参数[1]。表土层土壤容重一般指地表以下 28 cm 深度范围内土层的容重,是影响土壤水分、盐分及养分在土壤中运移的重要因素之一[2]。土壤容重的测量方法可分为直接方法和间接方法。传统的测量方法是环刀法,属于直接测量法的一种[3]。该方法是测量土壤容重的标准方法,测量过程简单,所得结果较为精确,但是对于大范围农田土壤容重数据的获取,该方法耗时、耗力,不能满足精细农业快速测量土壤容重的需求。

近年来,国内外学者在快速高效预测土壤容重方面进行了较多研究。文献[4]以圆锥指数作为输入,应用 BP 神经网络建立土壤传递函数 PTF_s,预测了粘土容重和粉质壤土容重,决定系数 R^2 分别为 0.697 3 和 0.686 8。文献[5]对兰陵溪流域森林土壤数据库进行研究,利用土壤有机碳、有机质建立了与土壤容重之间的回归模型。文献[6]研究了 6 种不同的回归方法,将几何平均粒径、土壤容重和粘土含量作为模型的输入,预测土壤水容重,结果显示,神经模糊和基因表达编程模型具有较高的预测精度,R^2 分别为 0.874 和 0.995。文献[7]将预测的土壤含水率作为辅助数据,改进协同克里格算法,实现了对土壤容重的预测,预测精度 R^2 从 0.449 提高到 0.852。

然而,影响表土层土壤容重的因素是多方面的,在大多数预测土壤容重的研究中,影响因素的选取较为简单,且部分参数需要经过复杂的实验才能获得。另外,由于数据获取地块和建模数据不同,大部分由实验建立的预测模型的适用范围具有针对性和局限性,因此应用于大范围农田土壤容重的获取时存在困难,不利于方法的推广使用。

现有研究表明,表土层土壤容重与土壤孔隙度存在极显著负相关关系[8],土壤表面孔隙度可以通过土壤表面粗糙度来预测[9-11]。而图像处理在粗糙度测量中有着广泛应用[12-13]。土壤阻力和土壤容重也具有着重要关系[14-15]。为此,本文从土壤阻力和基于图像处理得到的表征土壤表面粗糙度的信息融合入手,避免复杂的参数获取实验,建立梯度提升决策回归树(Gradient boosting regression tree, GBRT)模型,以期实现利用常规易获得的土壤物理参数对表土层土壤容重进行预测。

1 数据获取

1.1 数据采集

田间实验于 2019 年 5 月在北京市海淀区中国农业大学上庄实验站进行,该站位于 116.189 4°E, 40.141 1°N。实验田长 42 m,宽 40 m,总占地面积 1 680 m^2。采集 10 列样本,每列采集 10 个,共 100 个样本采集点。首先使用手机自带的相机拍摄采样点表面的土壤图像,然后使用环刀取土法[16]采集土样并装入有序号标记的铝盒中,采土设备如图 1 所示。同时使用 GPS 测量仪记录采样点的 GPS 数据,从第 1 个采集点开始,按一条龙顺序依次采集 100 个土样。使用实验室自主研发的车载式土壤阻力测量系统[17]测得离地表约 15 cm 深处的土壤阻力和 GPS 信息并保存,测量场景如图 2 所示。为了减少天气等因素对采集样本的影响,数据采集需要保证在 24 h 之内完成。



(a) 环刀 (b) 环刀手柄 (c) 铝盒

图 1 环刀取土设备

Fig. 1 Equipment for digging soil



图 2 车载式土壤阻力测量场景

Fig. 2 Vehicle-mounted soil resistance measurement

1.2 数据处理

车载式土壤阻力测量系统采用应变片电桥法测得深松犁钩在水平方向上受到的阻力并记录 GPS 信息。采集的数据为连续数据,所以根据土样采集点的 GPS 数据在土壤阻力数据中找到对应采集点的土壤阻力信息。将采集的 100 个土样带回实验室,连同环刀和铝盒一起称量。然后把装有环刀和土壤的铝盒放入 105℃ 的电热恒温干燥箱中干燥 24 h,取出铝盒,冷却至室温,再次称量并记录数据。表土层土壤容重计算式为

$$d=\frac{100P}{V(100+W)} \tag{1}$$

式中　d——土壤容重，g/cm^3

P——环刀内湿土质量，g

V——环刀容积，cm^3

W——土壤含水率，%

剔除实验中的人为误差和损坏数据，一共取得 90 组可用于建模分析的实验数据。选取其中 75% 的数据四舍五入得到 68 个样本进行建模，剩下的 22 个样本作为验证集。

2　研究方法

2.1　图像预处理

考虑到拍摄时偏光导致图像照明不均匀，图像各部分的平均亮（灰）度产生起伏变化，同时为了提高图像的对比度，起到图像增强的作用，本文使用同态滤波技术[18]对土壤表面图像进行预处理。主要步骤如下：

使用 $f_i(x,y)$ 表示照明函数，$f_\gamma(x,y)$ 表示物体的反射函数，则在光照下获得景物图像的数学模型为

$$f(x,y)=f_i(x,y)f_\gamma(x,y)$$
$$(0\leqslant f_i(x,y)\leqslant\infty\ ,0\leqslant f_\gamma(x,y)\leqslant 1) \tag{2}$$

由于照明函数 $f_i(x,y)$ 在空间上变化缓慢，其频谱处于低频区域；反射函数 $f_\gamma(x,y)$ 反映图像的细节内容，且纹理本身具有较多的细节和边缘，其频谱处于高频区域，故将空间域转变为频域。首先，对式(2)两端取对数，使原来相乘的 2 个分量变成空间域中相加的 2 个分量，即

$$z(x,y)=\ln f(x,y)=\ln f_i(x,y)+\ln f_\gamma(x,y) \tag{3}$$

再对式(3)两端分别进行傅里叶变换，得

$$Z(m,n)=I(m,n)+R(m,n) \tag{4}$$

式中，$Z(m,n)$、$I(m,n)$ 和 $R(m,n)$ 分别为 $z(x,y)$、$\ln f_i(x,y)$ 和 $\ln f_\gamma(m,n)$ 的傅里叶变换。

将式(4)两端乘以传递函数 $H(m,n)$，构成同态滤波器，使低频段受压缩，高频段得以扩展。则

$$S(m,n)=H(m,n)Z(m,n)=$$
$$H(m,n)I(m,n)+H(m,n)R(m,n) \tag{5}$$

利用傅里叶反变换，把式(5)变换回空间域，得

$$s'(x,y)=f_i'(x,y)+f_\gamma'(x,y) \tag{6}$$

由于傅里叶变换前取了对数，所以同态滤波增强后的图像函数 $g(x,y)$ 为

$$g(x,y)=\exp(s'(x,y))=$$
$$\exp(f_i'(x,y))\exp(f_\gamma'(x,y)) \tag{7}$$

亦即

$$g(x,y)=i_0(x,y)\gamma_0(x,y) \tag{8}$$

式中，$i_0(x,y)$ 和 $\gamma_0(x,y)$ 分别为输出图像的照射分量和反射分量。

2.2　表征土壤表面粗糙度的特征参数提取

特征提取是从图像中提取一组反映图像特性的基本元素或数字值[19]。表征图像表面粗糙度轮廓的特征向量可由反映纹理粗细程度的像元特征参数和反映纹理排列结构的区域特征参数两部分组成[20]。像元特征参数主要由图像灰度概率密度估计，通常由图像的灰度直方图来反映灰度概率密度，由熵、平均值、方差、偏度和峰度来表征直方图的分布情况，从而反映纹理的细致程度；区域特征参数主要由图像的灰度共生矩阵来反映。灰度共生矩阵通过研究图像中任意两点之间的灰度空间相关性来描述纹理，反映了关于方向、间隔和变换幅度的综合信息。通常选其能量、熵、对比度和逆方差等度量指标来描述图像的纹理特征。

2.3　特征参数选择

特征参数选择是从已经抽取的特征中选择能够更好地完成图像识别任务的参数，来表示原图像。实验使用灰度关联分析法[21]计算图像灰度直方图的熵、平均值、方差、偏度和峰度，以及图像灰度共生矩阵的能量、熵、对比度和逆方差，再加上土壤阻力共 10 种参数分别与表土层土壤容重的关联度，选择关联度较高的参数作为梯度提升决策回归树模型的输入变量。

灰度关联分析法的基本思想是量化特征图的几何状态，再计算出参考序列和比较序列之间的关联度。关联度较大时，比较序列的几何发展情况与参考序列更为接近[22]。利用灰色关联理论求出灰色关联度，以此来描述因素间关系的强弱和次序，为合理地选择回归模型的输入因子提供依据。

首先将表土层土壤容重作为母序列 $\boldsymbol{Y}$，将灰度直方图的熵、平均值、方差、偏度、峰度以及图像灰度共生矩阵的能量、熵、对比度、逆方差和土壤阻力作为子因素 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$、$\boldsymbol{X}_5$、$\boldsymbol{X}_6$、$\boldsymbol{X}_7$、$\boldsymbol{X}_8$、$\boldsymbol{X}_9$、$\boldsymbol{X}_{10}$，建立分析序列为

$$\boldsymbol{Y}=(y_1,y_2,\cdots,y_k,\cdots,y_n) \tag{9}$$

$$(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n)=\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{bmatrix}_{m\times n} \tag{10}$$

式中　m——样本数量

n——子因素数量

k——样本组序号

x_{nm}——第 m 个样本的第 n 个子因素

然后对所选样本序列进行无量纲化处理

$$y_{0j}=\frac{y_k}{y_1}\quad(1\leqslant j\leqslant m)\tag{11}$$

$$x_{0ik}=\frac{x_{ik}}{x_{i1}}\quad(1\leqslant i\leqslant n)\tag{12}$$

最后计算关联系数,$\boldsymbol{R}=[\xi_{ij}]_{m\times n}$,其中

$$\xi_{ij}=\frac{\min\limits_n\min\limits_m|y_{0j}-x_{ij}|+\rho\max\limits_n\max\limits_m|y_{0j}-x_{ij}|}{|y_{0j}-x_{ij}|+\rho\max\limits_n\max\limits_m|y_{0j}-x_{ij}|}\tag{13}$$

式中 ξ_{ij}——关联系数

$\boldsymbol{R}$——关联系数的 $m\times n$ 矩阵

ρ——分辨系数,一般取 0.5

$\min\limits_n\min\limits_m|y_{0j}-x_{ij}|$——2 级最小差

$\max\limits_n\max\limits_m|y_{0j}-x_{ij}|$——2 级最大差

各因素关联度计算公式为

$$\psi_{0j}=\frac{1}{m}\sum_{k=1}^{m}\xi_{ij}\tag{14}$$

式中 ψ_{0j}——关联度

2.4 梯度提升决策回归树

梯度提升决策回归树(GBRT)是由 FRIEDMAN[23] 于 2001 年提出的一种函数空间优化算法。该算法有 2 个重要思想,分别为提升和梯度。提升,即对一个任务而言,先由一棵决策树对任务进行判断,再用下一棵决策树对先前过程中的错误进行判断,并重复循环该过程直到得到满意的结果为止。梯度,即利用梯度作为前一步中的预测误差,梯度容易获取,而一般的损失函数在计算每一次迭代过程中的增益时难以获取,将梯度代替每一步迭代过程中的损失下降,并作为下一步拟合的目标值,易于操作和实现。主要步骤为:

(1)确定学习目标

在 GBRT 模型构建过程中,通常采用平方误差作为模型的学习目标,计算公式为

$$L(y_i,f(x))=\min\sum_{i=1}^{N}(y_i-f(x))^2\tag{15}$$

式中 y_i——真实值

$f(x)$——模型预测值

N——样本的数量

(2)构建弱决策回归树

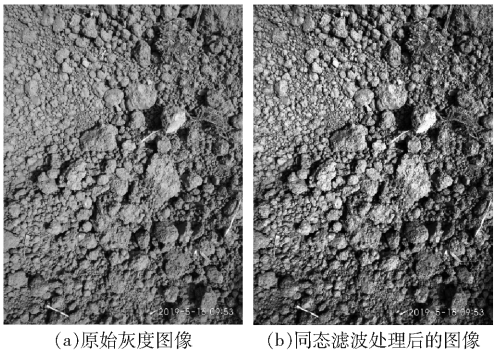
GBRT 是以分类回归树作为基函数的前向分布算法与加法模型。构建弱决策回归树的过程即为构建一棵分类回归树的过程。构建过程分为决策树生成和决策树剪枝。决策树生成是指基于训练集数据来生成决策树,生成的决策树越大越好。决策树剪枝是指利用测试集数据,以最小化评估指标对已生成的树进行剪枝。

(3)GBRT 的集成

模型最终的预测结果是将所有弱决策树的预测结果进行一定比例的求和。

3 结果与分析

分别对 90 组土壤表面图像利用同态滤波技术进行预处理。原始 RGB 图像颜色本身容易受到光照的影响,不能很好地反映图像的本质信息。图像所携带的信息主要表现在灰度形式上,颜色对提取信息并不是必需的。而且 3 通道处理转换为单通道处理以后,运算量可以大大减少,所以需要先将原始图像转换为灰度图像。部分处理后的图像与原始灰度图像对比如图 3 所示。可以看到,与图 3a 原始灰度图像相比,图 3b 经过同态滤波处理后的图像,其凹凸处的明暗对比更加明显,达到了图像增强的效果。



(a)原始灰度图像 (b)同态滤波处理后的图像

图 3 原始灰度图像与同态滤波处理后的图像

Fig. 3 Original grayscale image and image after homomorphic filtering

对每幅经同态滤波技术处理后的图像进行表征土壤表面粗糙度的特征参数提取。计算图像灰度直方图的熵(设为熵 1)、平均值、方差、偏度、峰度和图像灰度共生矩阵的能量、熵(设为熵 2)、对比度、逆方差的结果如表 1 所示。

使用灰度关联分析进行特征参数的提取。计算 90 组实验数据的图像灰度直方图的熵、平均值、方差、偏度和峰度,图像灰度共生矩阵的能量、熵、对比度和逆方差,以及土壤阻力共 10 个子因素与表土层土壤容重之间的关联度,结果如表 2 所示。关联度从大到小排序为 $\boldsymbol{X}_{10}$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$、$\boldsymbol{X}_8$、$\boldsymbol{X}_5$、$\boldsymbol{X}_7$、$\boldsymbol{X}_2$、$\boldsymbol{X}_9$、$\boldsymbol{X}_6$、$\boldsymbol{X}_1$。

一般关联度大于等于 0.8 时,子序列与母序列关联度很好;处于 0.6~0.8 之间关联度较好;小于 0.6 时,表示该子序列与母序列基本不相关。从表 2 可知,除图像灰度直方图的熵以外,其他的子序列与表土层土壤容重的关联度都大于 0.6。考虑到模型输入变量的选择对构建出的模型精度的影响,最终选定关联度大于 0.65 的变量,即土壤阻力,灰度直

表 1　90 组特征参数提取结果

Tab. 1　Extraction results of 90 sets of feature parameters

样本序号	熵 1	平均值	方差	偏度	峰度	能量	熵 2	对比度	逆方差
1	7.995 2	128.324 7	5 452.238 8	−0.008 8	−1.202 8	0.005 2	5.362 7	22.078 9	0.263 8
2	7.994 9	128.554 9	5 473.644 1	−0.009 4	−1.204 3	0.005 5	5.319 6	19.155 8	0.288 1
3	7.994 3	128.596 5	5 468.249 5	−0.009 9	−1.203 9	0.006 3	5.212 3	14.093 4	0.323 3
4	7.993 8	128.368 2	5 456.381 9	−0.010 7	−1.204 2	0.005 9	5.339 5	20.818 9	0.273 8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	7.983 2	129.020 1	5 490.325 2	−0.012 6	−1.201 7	0.009 4	4.903 4	6.835 7	0.431 4

表 2　表土层土壤容重与各子因素关联度计算结果

Tab. 2　Calculation results of correlation between soil bulk density of top soil layer and each sub-factor

子序列	子因素	与母序列土壤容重的关联度
X_1	熵 1	0.549 4
X_2	平均值	0.614 1
X_3	方差	0.710 2
X_4	偏度	0.684 9
X_5	峰度	0.672 7
X_6	能量	0.605 5
X_7	熵 2	0.655 5
X_8	对比度	0.682 3
X_9	逆方差	0.614 0
X_{10}	土壤阻力	0.881 2

方图的方差、偏度、峰度，图像灰度共生矩阵的熵和对比度，共 6 个子因素作为梯度提升决策回归树的输入变量。计算 6 个子因素之间的相关系数，结果如表 3 所示。可以看到变量之间的相关性都比较低，因此可以作为模型的输入参数。

表 3　6 个子因素的相关系数

Tab. 3　Correlation coefficient of six sub-factors

	X_{10}	X_3	X_4	X_8	X_5	X_7
X_{10}	1					
X_3	0.212 7	1				
X_4	0.266 1	0.142 8	1			
X_8	0.072 4	0.054 4	0.232 6	1		
X_5	0.134 0	0.037 8	0.347 7	0.255 1	1	
X_7	0.190 7	0.054 4	0.355 1	0.202 1	0.224 8	1

GBRT 模型主要有 7 个参数需要调整：弱决策树的数量 n_estimators，树深度 max_depth、内部节点再划分所需最小样本数 min_samples_split，叶子节点最少样本数 min_samples_leaf、学习率 learning_rate，最大特征数 max_features 和子采样比例 subsample。模型的构建和测试均在 Python 3.7 的环境中完成。首先选取一个固定的学习率，利用训练样本来训练模型，根据训练样本的损失来确定弱决策树的数量。通常对训练集进行交叉验证[24-25]获取，持续增大决策树的数量直到交叉验证的结果不再降低为止，此时获取的数量就是在确定学习率下能够获取的最优决策树数量。本实验以学习率 0.1 进行 5 折交叉验证，验证集的平均绝对误差（MAE）随决策树的数量变化趋势如图 4 所示。

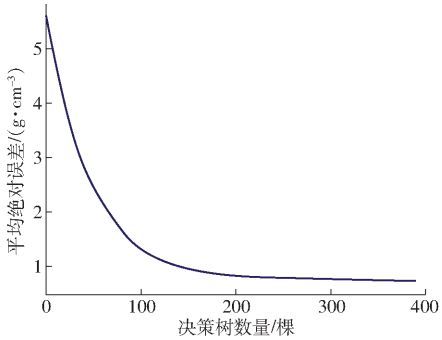


图 4　交叉验证下平均绝对误差随决策树数量的变化趋势

Fig. 4　Changing trend of average absolute error with number of trees under cross-validation

由图 4 可以看出，最佳的弱决策树数量为 380 棵。即在 380 棵后继续增长弱决策树数量，会出现平均绝对误差负增长情况，模型停止树的增长，同时不再计算平均绝对误差。

在获取了最佳弱决策树的数量后，即确定了一个集成器模型的基本规模。接下来确定树的深度和内部节点再划分最小样本数。使用网格搜索与交叉验证函数 GridSearchCV( ) 实现自动调参，只要把参数输进去，就能给出最优化的结果和参数。每一次调参之前都把上一步调出的参数代入模型。设 n_estimators为 380，max_depth 的范围为（3，14），搜索步长为 2，内部节点再划分所需最小样本数 min_samples_split范围为（100，801），搜索步长为 200。最佳搜索结果为：max_depth 为 5，min_samples_split 为 100。由于决策树深度 5 是一个比较合理的值，所以可以确定。而 min_samples_split 和决策树其他的参数存在关联，所以需要继续调参。

使用 min_samples_split 和叶子节点最少样本数 min_samples_leaf 一起调参。设置 min_samples_split 的范围为（600，1 900），搜索步长为 200。min_samples_leaf 的范围为（60，101），搜索步长为 10。得到的搜索结果为：min_samples_leaf 最佳为 60，min_samples_split 最佳为 800。

再对最大特征数 max_features 进行网格搜索。将以上调出的参数代入。设搜索范围为(1,8),步长为 2,得到的搜索结果为 5。对子采样比例 subsample 进行调参。设置搜索数组为(0.6,0.7,0.75,0.8,0.85,0.9),调参最佳结果为 0.8。对学习率设置搜索数组为(0.005,0.01,0.05,0.1),调参最佳结果为 0.01。模型参数的最终结果如表 4 所示。

表 4 模型参数调节结果

Tab.4 Model parameters adjustment result

参数	数值
弱决策树数量/棵	380
树深度	5
内部节点再划分所需最小样本数	60
叶子节点最少样本数	800
学习率	0.01
最大特征数	5
子采样比例	0.8

使用 75% 的数据进行模型的构建,使用 25% 的数据进行模型的验证。模型精度的评价标准选择决定系数 R^2 和平均绝对误差 MAE。将土壤阻力,灰度直方图的方差、偏度、峰度,图像灰度共生矩阵的熵和对比度共 6 个子因素作为输入,构建梯度提升决策回归树模型。使用验证集对模型进行验证,验证集的预测值和环刀取土法得到的表土层土壤容重进行相关性分析,结果如图 5 所示。

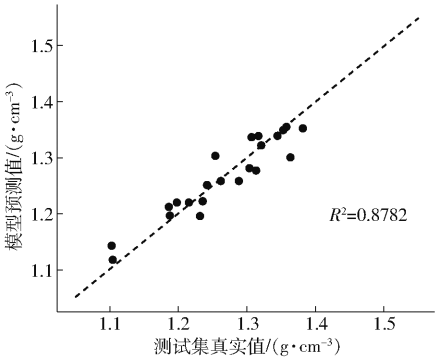


图 5 模型验证集的预测值与环刀法测得值的相关性

Fig.5 Correlation between predicted value of model verification set and value measured by ring knife method

为了验证在相同输入参数条件下模型的性能和时间复杂度,选择 BPNN 模型(训练目标为 0.001,学习率为 0.05,激活函数为 sigmoid)和支持向量机回归模型 SVR($C=0.5$,γ 为 0.2,kernel 为 linear)的精度和计算时间加以对比,使用 Python 中的 time 模块计算模型的运算时间。BPNN 和 SVR 模型验证集的预测值与环刀法测得值的相关性分别如图 6、7 所示。

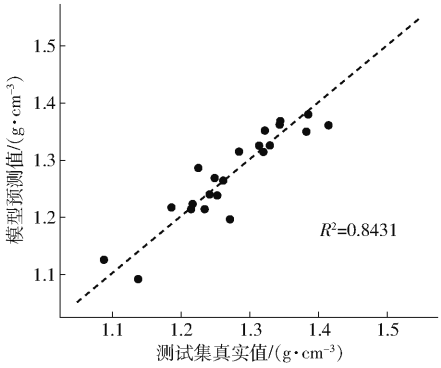


图 6 BPNN 模型验证集的预测值与环刀法测得值的相关性

Fig.6 Correlation between predicted value of BPNN model verification set and value measured by ring knife method

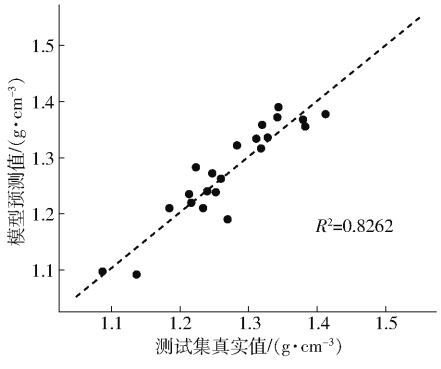


图 7 SVR 模型验证集的预测值与环刀法测得值的相关性

Fig.7 Correlation between predicted value of SVR model verification set and value measured by ring knife method

3 个算法在相同的运算环境和输入参数条件下,模型预测精度和运算时间如表 5 所示。

表 5 3 种模型预测结果

Tab.5 Prediction results of three models

模型	R^2	MAE/(g·cm⁻³)	t/s
GBRT	0.878 2	0.021 0	1.554 8
BPNN	0.843 1	0.035 6	2.658 4
SVR	0.826 2	0.042 9	2.763 1

由表 5 可以看出,GBRT 模型的预测结果最优,且运行时间最少,所以选择 GBRT 模型预测表土层土壤容重具有更好的性能。

GBRT 相比于其他回归树模型,变量的相对重要性是所有回归树变量相对重要性的均值,在回归的过程中,可以对特征进行进一步筛选,相关度低的特征可以不用。同时可以很好地处理缺失特征的数据,因为每个节点只依赖一个特征,如果某个特征不存在,这棵树依然可以用来决策,只是少一些路径,而 BPNN 和 SVR 没有这个特点。GBRT 对于特征空间的异常也有很好的鲁棒性,在低维度情况下,数据的规模对模型的影响也不大,因为对弱决策回归树的要求不高。GBRT 在建模过程中数据也不需要进行归一化,特征的作用只是用来分裂节点,叶节点的

值与特征值无关，是否归一化并不影响叶节点值和梯度提升的进程。

4　结论

（1）利用同态滤波技术对土壤表面图像进行预处理，提取反映图像纹理细致程度的像元特征参数和反映图像纹理排列结构的区域特征参数来表征表土层土壤表面粗糙度，从土壤阻力和土壤表面粗糙度的信息融合入手，利用灰度关联算法筛选出与表土层土壤容重关联度高的特征参数，使用梯度提升决策回归树模型对表土层土壤容重进行预测，并与 BPNN 和 SVR 模型的预测结果进行了对比，验证得到 GBRT 模型的预测结果具有更高的预测精度和更短的运行时间。

（2）图像处理和梯度提升决策回归树在表土层土壤容重预测上具有良好的应用前景。在低维数据下，GBRT 模型可以取得良好的预测效果，避免了传统容重测量的繁琐步骤，为表土层土壤容重的预测研究提供了参考。但是其调参过程较为繁琐，而且在处理高维度数据时会加大计算的复杂度。

参　考　文　献

[1] 郭国双. 谈谈土壤容重的测定[J]. 灌溉排水，1983(2)：39－40，38.
GUO Guoshuang. Talk about the determination of soil bulk density[J]. Journal of Irrigation and Drainage，1983(2)：39－40，38. (in Chinese)

[2] 郭李娜，樊贵盛. 基于灰色理论－BP 神经网络方法的表层土壤容重预测[J]. 节水灌溉，2018(2)：93－97.
GUO Lina，FAN Guisheng. Prediction of surface soil bulk density based on grey theory and BP neural network[J]. Water Saving Irrigation，2018(2)：93－97. (in Chinese)

[3] AL-SHAMMARY A A G，KOUZANI A Z，KAYNAK A，et al. Soil bulk density estimation methods：a review[J]. Pedosphere，2018，28(4)：581－596.

[4] 王巧利，林剑辉，许彦峰. 基于 BP 神经网络的土壤容重预测模型[J]. 中国农学通报，2014，30(24)：237－245.
WANG Qiaoli，LIN Jianhui，XU Yanfeng. The model for predicting soil bulk density based on the BP neural network[J]. Chinese Agricultural Science Bulletin，2014，30(24)：237－245. (in Chinese)

[5] 田耀武，黄志霖，肖文发，等. 三峡库区兰陵溪流域森林土壤有机碳、有机质与容重间的回归模型[J]. 华南农业大学学报，2016，37(1)：89－95.
TIAN Yaowu，HUANG Zhilin，XIAO Wenfa，et al. Regression model between forest soil organic carbon，organic matter and bulk density in the Lanlingxi watershed of the Three Gorges reservoir area[J]. Journal of South China Agricultural University，2016，37(1)：89－95. (in Chinese)

[6] SHIRI J，KESHAVARZI A，KISI O，et al. Using soil easily measured parameters for estimating soil water capacity：soft computing approaches[J]. Computers and Electronics in Agriculture，2017，141：327－339.

[7] YANG Q，LUO W，JIANG Z，et al. Improve the prediction of soil bulk density by cokriging with predicted soil water content as auxiliary variable[J]. Journal of Soils and Sediments，2016，16(1)：77－84.

[8] 许仁智，齐国翠，谢彦军，等. 广西东兰坡豪湖湿地表层土壤理化性质相关性分析[J]. 河池学院学报，2019，39(5)：44－50.
XU Renzhi，QI Guocui，XIE Yanjun，et al. Correlation analysis on physical and chemical properties of surface soil of Haohu Wetland in Donglanpo，Guangxi[J]. Journal of Hechi University，2019，39(5)：44－50. (in Chinese)

[9] SUN Yurui，LIN Jianhui，SCHULZE L，et al. Estimating surface porosity by roughness measurement in a silt-loam field[J]. Journal of Plant Nutrition & Soil Science，2006，169(5)：630－632.

[10] 张慧娟，孙宇瑞，林剑辉，等. 不同粗糙度尺度下预测表层土壤孔隙率量化指数比较研究[J]. 应用基础与工程科学学报，2009，17(1)：69－76.
ZHANG Huijuan，SUN Yurui，LIN Jianhui，et al. Comparative study on the quantitative indexes of predicted surface soil porosity at different roughness scales[J]. Journal of Basic Science and Engineering，2009，17(1)：69－76. (in Chinese)

[11] 陈世超，林剑辉，孙宇瑞，等. 基于土壤表面粗糙度预测降雨影响下的表层土壤孔隙度[J]. 北京林业大学学报，2013，35(2)：69－74.
CHEN Shichao，LIN Jianhui，SUN Yurui，et al. Prediction of surface soil porosity under the influence of rainfall based on soil surface roughness[J]. Journal of Beijing Forestry University，2013，35(2)：69－74. (in Chinese)

[12] 张胜宾，黄培奎，赵祚喜. 土壤表面粗糙度检测方法研究[J]. 农机化研究，2016，38(11)：257－262.
ZHANG Shengbin，HUANG Peikui，ZHAO Zuoxi. Research on testing method of soil surface roughness[J]. Journal of Agricultural Mechanization Research，2016，38(11)：257－262. (in Chinese)

[13] 陈斌，宋桂珍. 基于图像处理技术的表面粗糙度检测系统模型[J]. 机械管理开发，2010，25(2)：137－138.
CHEN Bin，SONG Guizhen. Model of surface roughness detection system based on image processing technology[J]. Mechanical Management and Development，2010，25(2)：137－138. (in Chinese)

[14] HOSSEINI M，NAEINI S A M，DEHGHANI A A，et al. Estimation of soil mechanical resistance parameter by using particle swarm optimization，genetic algorithm and multiple regression methods[J]. Soil and Tillage Research，2016，157：32－42.

[15] 刘晚苟,山仑,邓西平. 干湿条件下土壤容重对玉米根系导水率的影响[J]. 土壤学报,2003,40(5):779 - 782.
LIU Wan'gou, SHAN Lun, DENG Xiping. Effiects of soil bulk density on hydraulic conductivity of maize roots under drying and wet confitions[J]. Acta Pedologica Sinica, 2003, 40(5): 779 - 782. (in Chinese)

[16] 孔凡伟. 如何精测土壤容重[J]. 黑龙江农业科学,2010(10):178.
KONG Fanwei. How to accurately measure soil bulk density[J]. Heilongjiang Agricultural Sciences, 2010(10): 178. (in Chinese)

[17] 孟超,杨玮,张淼,等. 车载式土壤电导率与机械阻力实时测量系统[J/OL]. 农业机械学报,2019,50(增刊):102 - 107.
MENG Chao, YANG Wei, ZHANG Miao, et al. Development of real-time measurement system for vehicle-mounted soil conductivity and mechanical resistance[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 102 - 107. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=2019s017&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.S0.017. (in Chinese)

[18] 韩美林,杨琳. 基于同态滤波的彩色图像增强算法的改进与实现[J]. 科技经济市场,2018(8):14 - 15.
HAN Meilin, YANG Lin. Improvement and implementation of color image enhancement algorithm based on homomorphic filtering[J]. Science & Technology Ecnony Market, 2018(8): 14 - 15. (in Chinese)

[19] GONZALEZ R C, WOODS R E. 数字图像处理[M]. 阮秋琦,阮宇智,译. 2 版. 北京:电子工业出版社,2005.

[20] 崔敏. 纹理图像分析技术在表面粗糙度测量中的应用[J]. 新技术新工艺,2013(8):77 - 79.
CUI Min. Application of texture image analysis technology in surface roughness measurement[J]. New Technology New Process, 2013(8): 77 - 79. (in Chinese)

[21] 邓聚龙. 灰色系统基本方法[M]. 武汉:华中工学院出版社,1987:23 - 33.

[22] 刘梦雅,毛剑琳. 基于灰度关联分析法的卷积神经网络算法优化[J]. 电子科技,2018,31(6):84 - 88,95.
LIU Mengya, MAO Jianlin. Optimization of convolutional neural network algorithm based on gray relational analysis[J]. Electronic Science and Technology, 2018, 31(6): 84 - 88, 95. (in Chinese)

[23] FRIEDMAN J H. Greedy function approximation: agradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189 - 1232.

[24] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.

[25] 张欣怡,袁宏俊. 正则化和交叉验证在组合预测模型中的应用[J]. 计算机系统应用,2020,29(4):18 - 23.
ZHANG Xinyi, YUAN Hongjun. Application of regularization and cross-validation in combination forecasting model[J]. Computer Systems & Applications, 2020, 29(4): 18 - 23. (in Chinese)