

# 基于图像处理和 SVR 的土壤容重与土壤孔隙度预测

杨 玮, 兰 红, 李民赞\*, 孟 超

(中国农业大学信息与电气工程学院, 北京 100083)

**摘 要:** 土壤容重和土壤孔隙度是衡量土壤结构的重要参数。传统的土壤容重、土壤孔隙度获取方法费时费力, 且大多数预测模型的输入变量获取难度较高。该研究利用土壤粗糙度、土壤阻力与土壤容重的相关关系, 以土壤表面图像的颜色参数和纹理参数表征土壤粗糙度, 同使用车载式土壤阻力测量系统获得的土壤阻力一起, 从信息融合的角度构建了支持向量机回归 (Support Vector Regression, SVR) 土壤容重预测模型和 SVR 土壤孔隙度预测模型。图像处理使用 HSV 颜色空间进行阈值分割, 得到 HSV 颜色参数, 纹理参数使用灰度共生矩阵的能量、熵、对比度和逆方差。使用主成分分析对颜色参数和纹理参数进行主成分提取。将 SVR 模型的预测结果同环刀法测得的标准值进行相关性分析, 决定系数  $R^2$  达到了 0.867。土壤孔隙度 SVR 预测模型决定系数  $R^2$  达到了 0.743。在相同的运行环境下, 将 SVR 模型与决策树回归模型结果做了对比, 决策树回归对土壤容重和土壤孔隙度的预测精度  $R^2$  分别为 0.734 和 0.690, 验证得到 SVR 预测模型具有更好的预测精度。研究可为节省试验成本, 以及快速、有效预测土壤容重和土壤孔隙度提供方法和参考。

**关键词:** 图像处理; 土壤; 容重; 孔隙度; HSV 图像分割; SVR 预测模型

doi: 10.11975/j.issn.1002-6819.2021.12.017

中图分类号: S152.5

文献标志码: A

文章编号: 1002-6819(2021)-12-0144-08

杨玮, 兰红, 李民赞, 等. 基于图像处理和 SVR 的土壤容重与土壤孔隙度预测[J]. 农业工程学报, 2021, 37(12): 144-151.

doi: 10.11975/j.issn.1002-6819.2021.12.017 <http://www.tcsae.org>

Yang Wei, Lan Hong, Li Minzan, et al. Predicting bulk density and porosity of soil using image processing and support vector regression[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(12): 144-151. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.12.017 <http://www.tcsae.org>

## 0 引 言

土壤结构反映了土壤不同成分或性质的空间异质性, 对植物发育和土壤水平衡至关重要<sup>[1]</sup>。土壤容重和土壤孔隙度是衡量土体结构的重要参数<sup>[2]</sup>, 表征土体物理性质。二者密切相关<sup>[3]</sup>, 土壤容重小, 表明土壤比较疏松, 孔隙多, 保水保肥能力强。反之, 土壤容重大, 表明土体紧实, 结构性差, 孔隙少, 耕性、透水性、通气性不良, 保水保肥能力差。土壤容重是计算土壤孔隙率、空气含量和换算土壤相对含水率的基本数据<sup>[4]</sup>。传统测量土壤容重的方法是环刀法。该方法测量结果精准, 通常用作标准测量法, 但是需要环刀、铝盒、烘干箱等多个仪器, 整个过程耗时耗力, 不利于快速、精准地测量大范围农田土壤容重。因此, 探寻便捷高效的间接土壤容重测量法一直是精细农业研究的重要内容。

近年来, 利用机器学习等回归模型预测土壤容重的方法在国内外得到了广泛应用。Alqinna 等<sup>[5]</sup>利用简单线性回归、多元非线性回归、逐步多元线性回归、偏最小二乘和人工神经网络预测土壤容重, 并比较了每个模型的预测精度, 发现人工神经网络在预测土壤容重方面最

为出色, 与其他模型相比准确度提升了 18.4%, 同时发现, 增加模型预测变量的数量可以增强土壤容重的预测能力。Abdelbaki<sup>[6]</sup>使用美国土壤数据库评估了 48 种公开的 PTF 函数 (Pedo-Transfer Function) 预测土壤容重的准确性, 并构建了新的 PTF 预测模型, 利用有机碳或有机物含量作为模型输入, 模型的最佳性能可以达到  $R^2$  为 0.59, RMSE 为 0.13 g/cm<sup>3</sup>。Yang 等<sup>[7]</sup>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  $R^2$  达到了 0.852。Martin 等<sup>[8]</sup>基于多重累加回归树方法推导并试验了两个新的 PTF 模型, 应用于国家级土壤数据库, 模型的  $R^2$  值分别为 0.83 和 0.94。然而, 大多数用于预测土壤容重的 PTF 函数和回归模型, 其输入的变量很难获得且测量成本很高, 如各种有机物的含量、土壤有机质等。且模型仅适用于特定的农田环境条件, 仅能在有限的地理区域内使用, 不便于方法的重复和推广。

研究表明<sup>[3,9]</sup>, 土壤容重与土壤孔隙度具有极强的相关性, 且土壤孔隙度可以直接利用土壤容重由经验公式计算得到。利用土壤粗糙度预测土壤孔隙度的研究也已经取得不少成果<sup>[10-12]</sup>。随着技术的进步, 图像法在土壤粗糙度的测量中得到了不少应用<sup>[13-16]</sup>。同时有研究指出<sup>[17-19]</sup>, 耕作的牵引阻力主要取决于土壤强度。在茬口、有机质和含水量相同的情况下, 影响耕作阻力的主要因素是土壤容重, 土壤阻力可以在很大程度上定量反映土壤容重。综上, 利用图像处理法得到表征土壤粗糙度的特征参数和土壤阻力作为输入, 能够构建土壤容重和土壤粗糙度回

收稿日期: 2021-03-18 修订日期: 2021-05-15

基金项目: 国家自然科学基金资助项目 (31801265)

作者简介: 杨玮, 博士, 副教授, 博士生导师, 研究方向为精细农业系统集成。Email: cauyw@cau.edu.cn.

\*通信作者: 李民赞, 博士, 教授, 博士生导师, 研究方向为农业电气化与自动化。Email: limz@cau.edu.cn

归预测模型。杨玮等<sup>[20]</sup>对土壤表面图像做了预处理,并提取了灰度图像的特征参数,与土壤阻力一起构建了梯度提升决策回归树(Gradient Boosting Regression Tree, GBRT)预测模型,实现了对土壤容重的预测。文献的重点在于图像预处理和灰度图像特征提取,但是缺少对图像的颜色特征参数的探究,也没有构建土壤孔隙度预测模型。另外,GBRT 模型的调参过程比较复杂,从图像增强、特征提取、特征选择,再到模型构建,整个过程的技术和算法要求较高,因此有必要探究更便捷、高效的方法,来实现对土壤容重和土壤孔隙度的预测。支持向量机回归(Support Vector Regression, SVR)模型作为机器学习的经典模型,在非线性预测的应用中已被良好认可,且模型应用广泛,调参快捷,精度较高。

本文基于实验室的设备和研究基础,利用高质量的土壤表面图像,免去图像增强处理步骤,直接提取图像的颜色特征参数,将其和车载式土壤阻力测量系统获得的阻力作为回归模型的输入变量,建立支持向量机回归预测模型,预测土壤容重,并补充构建了土壤孔隙度预测模型,同时实现对土壤容重和土壤孔隙度的预测。本研究可降低参数获取难度,节省试验步骤,简化方法和模型构建过程,为大范围农田土壤容重和孔隙度的获取提供思路和方法。

## 1 数据获取

### 1.1 试验田概况

中国农业大学上庄实验站地处北京市海淀区北部的上庄镇,位于 115.7°~117.4°E, 39.4°~41.6°N, 面积 16 410 km<sup>2</sup>。该区域为典型的暖温带半湿润大陆性季风气候,夏季高温多雨,冬季寒冷干燥。年平均气温 11.5 °C, 年平均降水量约 540.7 mm, 年蒸发量为 466.7 mm。土壤质地包括轻壤土、中壤土、粉砂壤土和砂壤土,有机质质量分数达 0.34%~1.70%, 土壤肥沃,质地适中。

### 1.2 数据采集

数据采集试验于 2019 年 5 月进行,试验田长 42 m, 宽 40 m, 总占地面积 1 680 m<sup>2</sup>。采集 10 列样本,每列采集 10 个,共 100 个样本。拍摄土壤表面图像的设备为小米 CC9 智能手机。使用后置相机垂直离地大约 30 cm 左右拍摄土壤表面图像。然后使用环刀取土法<sup>[21]</sup>采集采样点的土壤。

首先用环刀垂直插入取样点,然后用小铲子挖出装满土的环刀,削去环刀口多余的土,保证采集面平整。将装满土的环刀放入带有标号的铝盒中,盖好盖,防止水分蒸发。从第 1 个点开始,按 S 路线依次采集。使用实验室自主研发的车载式土壤阻力测量系统<sup>[22]</sup>从第 1 个采集点开始,按照环刀取土的路线以 2 m/s 的速度测得离地表约 15 cm 深处的土壤阻力并保存。完成试验的时间大约 5 h,并保证土壤样本不受时间、天气等其他环境因素的影响。

### 1.3 数据处理

使用车载式土壤阻力测量系统记录插入土壤的犁钩在水平方向受到的阻力,采集的数据为连续数据,包括阻力和时间。采集点之间的间距为 4 m,车速为 2 m/s,所以纵向采集的每列中,两个采样点阻力采集的时间间隔为 2 s,变换采集线路时,犁钩抬离地面,阻力为 0。找到对

应采集点的阻力数据并整理到 Excel 中。将采集的 100 个土样带回实验室,连同环刀和铝盒一起称质量。然后把装有环刀和土壤的铝盒放入电热鼓风干燥箱(DHG-9003, 上海一恒科学仪器有限公司)中,在 105 °C 下干燥 24 h。取出铝盒,冷却至室温(26 °C),再次称质量并记录数据。分别计算土壤含水率、土壤容重和土壤孔隙度:

$$W = \frac{\rho_{\text{湿}} - \rho_{\text{干}}}{\rho_{\text{干}}} \quad (1)$$

式中  $W$  代表土壤含水率,%;  $\rho_{\text{湿}}$  代表湿土的质量, g;  $\rho_{\text{干}}$  代表干土的质量, g。

$$d = \frac{100\rho_{\text{湿}}}{V(100+W)} \quad (2)$$

式中  $d$  代表土壤容重, g/cm<sup>3</sup>;  $V$  代表环刀容积, cm<sup>3</sup>, 本试验的环刀容积为 100 cm<sup>3</sup>。

$$f = 100(1 - \frac{d}{s}) \quad (3)$$

式中  $f$  代表土壤孔隙度,%;  $s$  代表土壤密度, g/cm<sup>3</sup>, 一般取 2.65。

## 2 研究方法

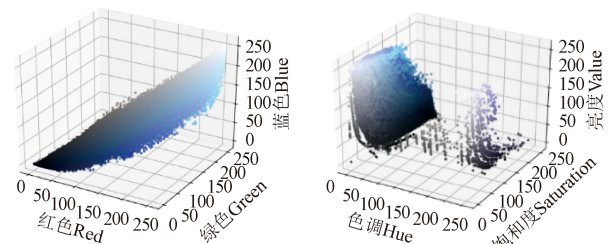
### 2.1 图像分割

手机拍摄的土壤表面图像为彩色图像。很多图像处理方法都是先把图像转换为灰色图像,丢失了图像彩色信息的细节成分<sup>[23]</sup>。图像分割是图像识别和图像理解的基本前提,分割效果直接影响图像特征参数的提取结果。所以,基于图像颜色信息进行分割,可以充分发挥彩色空间中颜色的贡献<sup>[24]</sup>。

原始土壤表面图像如图 1 所示。首先选取合适的颜色空间,分别制作图像在 RGB 颜色空间和 HSV 颜色空间上的可视化图像,如图 2a 和图 2b 所示。



图 1 土壤原始图像  
Fig.1 Original image of soil

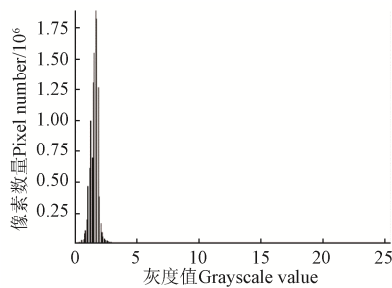


a. RGB 空间可视化图像  
a. RGB space visualization image  
b. HSV 空间可视化图像  
b. HSV space visualization image  
图 2 土壤图像在 RGB 和 HSV 颜色空间上的可视化图像  
Fig.2 Visualization of soil image in RGB and HSV color space

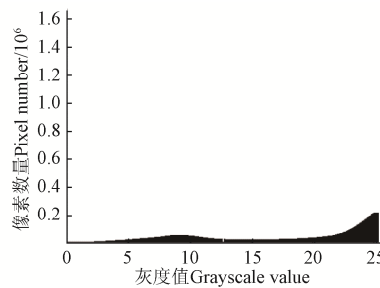
由图 2 可以看出, 在 RGB 颜色空间上, 图像的主体颜色跨越了几乎整个范围的 Red (红色,  $R$ )、Green (绿色,  $G$ )、和 Blue (蓝色,  $B$ ) 值。由于图像的一部分延伸到几乎整个坐标系, 所以根据 RGB 值的范围在 RGB 空间分割土壤图像并不容易。而在 HSV 颜色空间上, 图像的色彩在 Hue (色调,  $H$ )、Saturation (饱和度,  $S$ )、Value (亮度,  $V$ ) 上有明显的分离, 能够用于图像分割。因此, 选择 HSV 颜色空间进行图像分割。

将 RGB 图像转换为 HSV 图像<sup>[25]</sup>, 原理如下:

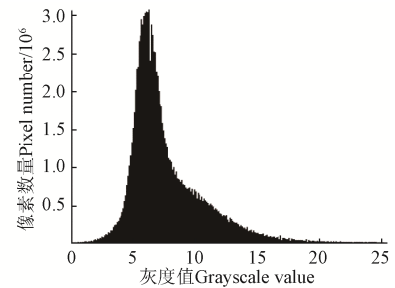
$$H = \begin{cases} 0^\circ, \Delta = 0 \\ 60^\circ \cdot \left( \frac{G' - B'}{\Delta} \bmod 6 \right), C_{\max} = R' \\ 60^\circ \cdot \left( \frac{B' - R'}{\Delta} + 2 \right), C_{\max} = G' \\ 60^\circ \cdot \left( \frac{R' - G'}{\Delta} + 4 \right), C_{\max} = B' \end{cases} \quad (4)$$



a.  $H$  分量的直方图  
a. Histogram of  $H$  component



b.  $S$  分量的直方图  
b. Histogram of  $S$  component



c.  $V$  分量的直方图  
c. Histogram of  $V$  component

图 3 HSV 分量直方图

Fig.3 HSV component histogram

图像分割结果如图 4 所示。可以看出, 经过 HSV 图像分割后的土壤表面图像与原始图像相比, 起伏更加突出, 颗粒感更强, 可以更明显地表现土壤图像的粗糙度特征。



图 4 HSV 图像分割结果

Fig.4 HSV image segmentation results

## 2.2 图像颜色特征提取

颜色距是一种有效的颜色特征表示方法<sup>[26-27]</sup>, 图像的颜色信息主要分布于低阶距中, 包括一阶距 (均值)、二阶距 (准差) 和三阶距 (斜度)。图像的三个分量  $H$ 、 $S$ 、 $V$  的前三阶颜色矩组成一个 9 维直方图向量, 图像的颜色特征表示如下:

$$S = \begin{cases} 0, C_{\max} = 0 \\ \frac{\Delta}{C_{\max}}, C_{\max} \neq 0 \end{cases} \quad (5)$$

$$V = C_{\max} \quad (6)$$

式中  $R'$  为  $R/255$ ,  $G'$  为  $G/255$ ,  $B'$  为  $B/255$ , 代表把 RGB 空间的 Red、Green 和 Blue 值分别转为  $[0, 1]$  区间;  $\bmod 6$  代表对 6 取余;  $C_{\max}$  为  $\max(R', G', B')$ ;  $\Delta$  为  $C_{\max}$  与  $\min(R', G', B')$  的差。

根据图像 HSV 分量的直方图来确定最佳的分割阈值, 如图 3 所示。结合分布直方图的分量范围, 编写阈值滑动调整程序。在对阈值进行调整的同时, 获得指定颜色范围内的掩码, 对原始图像进行按位与操作, 保留掩码区域。通过程序调试得到的最佳分割阈值为:  $h\_min=0$ ,  $h\_max=30$ ,  $s\_min=10$ ,  $s\_max=60$ ,  $v\_min=0$ ,  $v\_max=255$ 。对应的掩码范围边界值为:  $lower = (0, 10, 0)$ ,  $upper = (30, 60, 255)$ 。

$$F_{color} = [\mu_H, \sigma_H, \delta_H, \mu_S, \sigma_S, \delta_S, \mu_V, \sigma_V, \delta_V] \quad (7)$$

$$\mu_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (8)$$

$$\sigma_i = \left( \frac{1}{N} \sum_{j=1}^N (P_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (9)$$

$$\delta_i = \left( \frac{1}{N} \sum_{j=1}^N (P_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (10)$$

式中  $F_{color}$  表示图像的颜色特征,  $P_{ij}$  表示彩色图像第  $j$  个像素的第  $i$  个颜色分量;  $N$  表示图像中像素的个数;  $\mu$ 、 $\sigma$ 、 $\delta$  分别表示一阶距、二阶矩和三阶矩下标  $H$ 、 $S$ 、 $V$  分别代表图像的三个分量。

## 2.3 图像纹理特征提取

灰度共生矩阵 (Gray-Level Co-occurrence Matrix, GLCM) 通过研究图像的灰度空间, 使用不同像素之间的联合概率密度得出图像特征, 不但能反映整体图像像素的分布特性, 也能反映图像中像素值相近的局部分布特性<sup>[28]</sup>。由于灰度共生矩阵的数据量较大, 一般不直接作为区分纹理的特征, 而是利用其构建的一些统计量来表征纹理特征。本文选择灰度共生矩阵的能量、熵、对比度和逆方差四个常用参数来作为表征图像表面粗糙度的

特征参数。

2.4 主成分分析

使用主成分分析（Principal Component Analysis, PCA）对获取的图像颜色特征参数和纹理特征参数进行降维处理。主成分分析可以将原来众多的、具有一定关联性的指标，进行重新组合，形成新的、互相无相关性的综合指标，同时对原有指标的信息做到最大化的保留，并省去部分重复信息，简化复杂数据的信息<sup>[29-30]</sup>。

2.5 支持向量机回归模型

支持向量机回归（SVR）是由支持向量机（Support Vector Machine, SVM）衍生得到的模型<sup>[31-32]</sup>。SVR 回归模型通过找到一个回归超平面，让一个集合的所有数据到该平面的距离最近<sup>[33]</sup>。具体方法如下：

设样本集为：  $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ 。 $x \in R^n, y \in R, R$  表示实数集。 $y$  使用  $x$  的函数  $f(x)$  代替，则样本集中  $y$  与  $x$  可通过如下方程表示：

$$f(x) = \omega \cdot x + b \tag{11}$$

式中  $\omega$  和  $b$  为超平面的系数。

若原始数据与支持向量机回归拟合良好，则使得

$$\min \frac{1}{2} \|\omega\|^2$$

s.t.  $\begin{cases} \omega \cdot x_n + b - y_n \leq \varepsilon \\ y_n - \omega \cdot x_n - b \leq \varepsilon \end{cases}, n=1,2,\dots,l$  (12)

式中  $\varepsilon$  为任意的一个正数。

引入拉格朗日对数对式（11）进行变化得到变形为

$$f(x) = \omega \cdot x + b = \sum_{n=1}^l (a_n - a_n^*) (x_n \cdot x) + b \tag{13}$$

式中  $a_n$  和  $a_n^*$  为样本支持向量，大多数取值为 0。

为了解决维数灾难问题，当核函数满足 Mercer 条件<sup>[34]</sup>，便可获得内积函数  $K(x_n, x_j) = \psi(x_n) \cdot \psi(x_j)$ 。同时引入拉格朗日变化得到最后的回归函数

$$f(x) = \omega \cdot x + b = \sum_{n=1}^l (a_n - a_n^*) K(x_n \cdot x) + b \tag{14}$$

本文使用线性核函数  $K(x, x_i) = x^T x_i$ ，在 python 中调用 sklearn 中的 svm 包，引入 SVR 模型。试验均在 Python3.7 的环境中进行。

3 结果与分析

3.1 图像特征参数提取结果

分别利用阈值调整程序对田间试验拍摄的 100 幅土壤表面图像确定分割阈值，并进行 HSV 图像分割处理。对每幅经过 HSV 分割处理之后的图像提取颜色特征参数和灰度共生矩阵纹理特征参数。图像处理得到的所有参数，即  $H$  分量的一阶距 ( $H\_mean$ )、 $S$  分量的一阶距 ( $S\_mean$ )、 $V$  分量的一阶距 ( $V\_mean$ )、 $H$  分量的二阶矩 ( $H\_std$ )、 $S$  分量的二阶矩 ( $S\_std$ )、 $V$  分量的二阶矩 ( $V\_std$ )、 $H$  分量的三阶矩 ( $H\_thirdMoment$ )、 $S$  分量的三阶矩 ( $S\_thirdMoment$ )、 $V$  分量的三阶矩 ( $V\_thirdMoment$ )、及灰度共生矩阵的能量、熵、对比度和逆方差，依次设为  $X_1 \sim X_{13}$ ，对应关系如表 1 所示，计算结果如表 2 所示，表 2 中颜色特征的最大值和最小值之间有明显差异，因此可以通过选择的这些颜色特征来对土壤容重和孔隙度进行区分。13 个变量之间的相关性如表 3 所示。

表 1 图像特征参数与替代字母对应表  
Table 1 Correspondence table of image feature parameters and substitute letters

替代字母 Substitute letter	特征参数 Feature parameter
$X_1$	$H\_mean$
$X_2$	一阶距 $S\_mean$ $V\_mean$
$X_3$	
$X_4$	
$X_5$	二阶矩 $H\_std$ $S\_std$ $V\_std$
$X_6$	
$X_7$	
$X_8$	三阶矩 $H\_third\ Moment$ $S\_third\ Moment$ $V\_third\ Moment$
$X_9$	
$X_{10}$	
$X_{11}$	灰度共生矩阵的能量
$X_{12}$	灰度共生矩阵的熵
$X_{13}$	灰度共生矩阵的对比度
	灰度共生矩阵的逆方差

表 2 100 幅图像颜色特征参数和纹理特征参数计算统计结果  
Table 2 Statistical results of color feature parameters and texture feature parameters calculation of 100 images

统计指标 Statistical indicators	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$
最大值 Max	16.125	47.977	215.253	9.547	27.998	113.819	13.156	28.294	115.108	0.244	28.226	5.068	0.858
最小值 Min	5.135	21.342	91.676	3.423	11.100	58.688	4.941	14.088	75.800	0.010	1.372	2.351	0.365

表 3 13 个图像特征参数之间的相关性  
Table 3 Correlation among 13 image feature parameters

参数 Parameters	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$
$X_1$	1												
$X_2$	0.421	1											
$X_3$	0.645	0.655	1										
$X_4$	-0.406	-0.548	-0.747	1									
$X_5$	-0.780	-0.246	-0.778	0.667	1								
$X_6$	-0.805	-0.427	-0.779	0.718	0.943	1							
$X_7$	-0.011	-0.437	-0.354	0.793	0.140	0.216	1						
$X_8$	-0.743	-0.074	-0.677	0.587	0.981	0.894	0.071	1					
$X_9$	-0.791	-0.197	-0.510	0.550	0.874	0.921	0.077	0.878	1				
$X_{10}$	-0.368	0.251	-0.244	0.041	0.487	0.278	-0.206	0.550	0.319	1			
$X_{11}$	-0.345	-0.452	-0.164	0.324	0.147	0.324	0.271	0.084	0.350	-0.460	1		
$X_{12}$	0.193	-0.489	0.051	0.097	-0.406	-0.187	0.336	-0.502	-0.281	-0.923	0.604	1	
$X_{13}$	-0.084	0.454	-0.130	-0.062	0.372	0.159	-0.285	0.452	0.168	0.839	-0.764	-0.944	1

### 3.2 主成分分析结果

可以看到图像处理得到的 13 个参数, 部分参数之间存在相关性, 不利于后期模型的构建。利用主成分分析算法在数据信息损失最小的基础之上, 对原始特征进行

降维, 得到全新的正交特征并作为主成分。因图像颜色特征包含  $H$ 、 $S$ 、 $V$  三个分量, 纹理特征包括 4 个特征参数, 所以选定将 13 维特征变为 7 维特征。主成分分析得到的 7 维特征参数如表 4 示。

表 4 主成分分析得出的 7 维特征参数

Table 4 7-dimensional feature parameters derived from principal component analysis

样本编号 Sample number	F1	F2	F3	F4	F5	F6	F7
1	-15.943 867	10.120 447	-5.370 081	-4.885 529	-0.531 156	0.455 391	-0.889 505
2	27.704 339	-13.382 078	-3.871 811	2.295 241	0.160 221	1.797 315	-0.493 482
3	46.612 955	-4.319 465	-4.474 657	0.859 269	0.932 793	0.098 121	-0.010 338
4	32.471 711	5.841 787	3.363 701	-3.985 743	1.013 543	-0.636 549	0.901 573
...	...	...	...	...	...	...	...
100	70.504 153	12.387 427	-3.460 819	0.597 178	6.469 628	-4.948 285	0.860 971

注: F1~F7 表示主成分分析得到的 7 维特征参数。

Note: F1-F7 represent 7-dimensional feature parameters derived from principal component analysis.

### 3.3 模型预测与结果分析

scikit-learn(sklearn)是机器学习中常用的第三方模块, 建立在 numpy、scipy 和 matplotlib 模块之上, 对机器学习方法进行了封装。模型训练程序从 sklearn.svm 中引入 SVR, 内核类型选择 linear 线性内核, 惩罚参数等其他参数默认。使用 75% 的数据作为建模集, 25% 的数据作为验证集。模型精度的评价标准选择决定系数  $R^2$  和均方根误差 RMSE。将 7 个特征参数 F1~F7, 和土壤阻力一起共 8 个因素作为输入, 土壤容重数据作为输出, 训练得到土壤容重预测模型。使用验证集对模型进行验证, 验证集得到的预测值和环刀法得到的土壤容重标准值进行相关性分析, 结果如图 5a 所示。

为了比较不同机器学习模型在同样的运算环境和相同变量输入条件下, 对土壤容重和土壤孔隙度的预测精度, 选择常用且易于理解的决策树回归模型做了性能对比, 并记录两种模型的运算时间。决策树回归模型的特征选择标准使用 MSE, 特征划点选择标准使用 best, 最大特征数和决策树最大深度默认。决策树回归模型对土壤容重预测值和真实值的相关性如图 5b 所示。

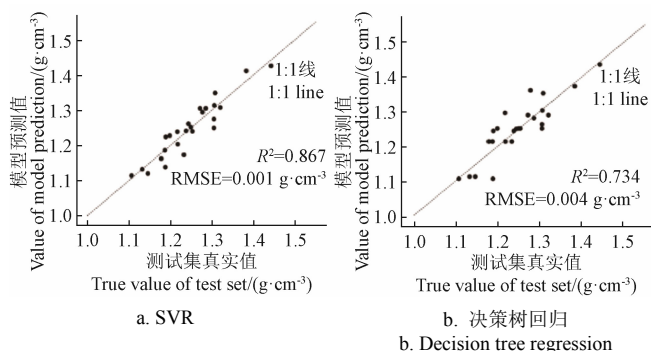


图 5 SVR 土壤容重预测模型和决策树回归土壤容重预测模型的预测值和真实值

Fig.5 Predicted values of SVR and decision tree regression soil bulk density prediction model and true values

将土壤孔隙度作为输出, 训练得到土壤孔隙度预测模型。SVR 土壤孔隙度模型预测值和真实值的相关性如图 6a 所示, 决策树回归土壤孔隙度预测模型的预测值和真实值的相关性如图 6b 所示。

可以看到, SVR 模型对土壤容重预测的决定系数  $R^2$  达到了 0.867, 均方根误差为  $0.001 \text{ g/cm}^3$ , 优于决策树回归对土壤容重的预测。不过运算时间用了 6.810 s, 相较于决策树回归模型 2.153 s 的运算时间, 多出了 4.7 s 左右。对土壤孔隙度的预测上, SVR 模型的决定系数  $R^2$  为 0.743, 均方根误差为 2.284%。决策树回归模型的决定系数  $R^2$  为 0.690, 均方根误差为 3.345%, 运算时间上 SVR 模型所用时间为 3.144 s, 决策树回归模型的运算时间为 4.302 s。综合可得, 将图像特征参数和土壤阻力作为输入, SVR 模型对土壤容重和土壤孔隙度的预测都表现出了比决策树高的精度, 虽然在预测土壤容重时运算时间多出几秒, 但是在可接受范围内, 符合模型精度和运算需求, 因此 SVR 模型具有更好的性能。

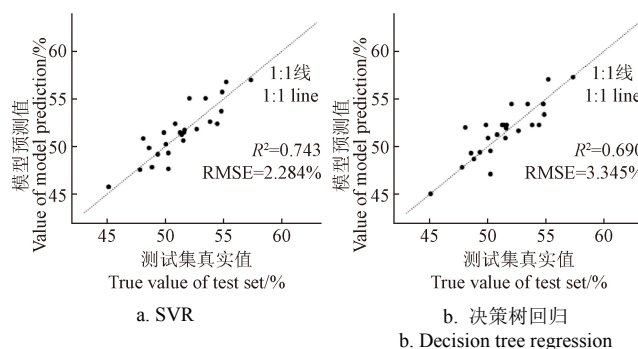


图 6 SVR 土壤孔隙度预测模型和决策树回归土壤孔隙度预测模型的预测值和真实值

Fig.6 Predicted values of SVR and decision tree regression soil porosity prediction model and true values

SVR 的理论基础是利用内积核函数代替高维空间的非线性映射。当核函数已知时, 可以简化高维空间的求解难度。同时, SVR 基于小样本统计理论, 避开了从归纳到演绎的传统过程, 实现了高效的从训练样本到预测样本的推理, 简化了回归问题。SVR 的最终决策函数由少数的支持向量所确定, 计算的复杂性取决于支持向量的数目, 可以抓住关键样本, 剔除大量冗余样本, 非支持向量样本数量的改变对模型没有影响, 增强了模型的鲁棒性。相对于其他机器学习模型, SVR 非常擅长解决复杂的、中小规模训练集的非线性问题, 且模型调参比

较容易, 内核类型确定以后其他参数可选择默认。但是随着样本量的增加, SVR 模型的计算复杂度会成倍增加, 从而耗费较大的机器内存和运算时间, 因此在考虑成本和效率时, 需要注意模型训练样本量不能过大。

## 4 结 论

1) 利用图像的 HSV 颜色空间参数、纹理参数和土壤阻力预测土壤容重与土壤孔隙度是可行的。颜色参数和纹理参数可以作为表征土壤粗糙度的特征参数。将土壤表面图像在 HSV 颜色空间上做分割处理, 可以增强图像的粗糙表现程度。对颜色参数和纹理参数进行主成分分析, 然后同土壤阻力一起作为输入变量, 构建支持向量机回归 (Support Vector Regression, SVR) 土壤容重预测模型和 SVR 土壤孔隙度预测模型, 能够实现对土壤容重和土壤孔隙度的预测。

2) SVR 模型在利用图像颜色参数、纹理参数和土壤阻力作为输入时, 预测土壤容重的模型精度可达 0.867, 预测土壤孔隙度的模型精度可达 0.743。将模型结果同决策树回归模型的预测结果做对比, 决策树回归模型对土壤容重和土壤孔隙度的预测精度  $R^2$  分别为 0.734 和 0.690, 因此 SVR 模型具有更好的预测精度。虽然预测土壤容重时耗费的时间稍多, 但是并不影响模型的精度和适用性。在中小规模数据样本情况下, SVR 可以实现良好的预测效果, 为省时省力获取土壤容重和土壤孔隙度提供了方法。

## [参 考 文 献]

- [1] Dexter A R. Advances in characterization of soil structure[J]. Soil & Tillage Research, 1988, 11(3/4): 199-238.
- [2] 于博, 赵兰坡, 高聚林, 等. 松辽平原高产农田土壤孔性对冻结的响应及机理研究[J]. 水土保持学报, 2012, 24(4): 80-85.  
Yu Bo, Zhao Lanpo, Gao Julin, et al. Study on response and mechanism of soil porosity to freezing in high-yield farmland in Songliao Plain[J]. Journal of Soil and Water Conservation, 2012, 24(4): 80-85. (in Chinese with English abstract)
- [3] 赵洪书. 从容重计算土壤总孔隙度的初步探讨[J]. 土壤学报, 1964, 12(2): 230-234.  
Zhao Hongshu. Preliminary discussion on calculating soil total porosity from bulk density[J]. Bulletin of the Soil Science Society of China, 1964, 12(2): 230-234. (in Chinese with English abstract)
- [4] 郭国双. 谈谈土壤容重的测定[J]. 灌溉排水学报, 1983, 2(2): 39-40, 38.
- [5] Alqinna M I, Jaber S M. Predicting soil bulk density using advanced pedotransfer functions in an arid environment[J]. Transactions of the ASABE (American Society of Agricultural and Biological Engineers), 2013, 56: 963-976.
- [6] Abdelbaki A M. Evaluation of pedotransfer functions for predicting soil bulk density for U. S. soils[J]. Ain Shams Engineering Journal, 2016, 9(4): 1611-1619.
- [7] Yang Q Y, Luo W Q, Jiang Z C, 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] Martin M P, Seen D L, Boulonne L, et al. Optimizing pedotransfer functions for estimating soil bulk density using boosted regression trees[J]. Soil Science Society of America Journal, 2011, 73(2): 485-493.
- [9] 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 of physical and chemical properties of surface soil in Pohao Lake Wetland, Donglan, Guangxi[J]. Journal of Hechi University, 2019, 39(5): 44-50. (in Chinese with English abstract)
- [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 index of predicting surface soil porosity under different roughness scales[J]. Journal of Basic Science and Engineering, 2009, 17(1): 69-76. (in Chinese with English abstract)
- [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 with English abstract)
- [12] Sun Y R, Lin J H, Schulze L P, 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.
- [13] Taconet O, Ciarletti V. Estimating soil roughness indices on a ridge-and-furrow surface using stereo photogrammetry[J]. Soil and Tillage Research, 2007, 93(1): 64-76.
- [14] 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. Surface roughness detection system model based on image processing technology[J]. Mechanical Management and Development, 2010, 25(2): 137-138. (in Chinese with English abstract)
- [15] 王玉景. 基于图像的车削表面粗糙度测量[D]. 哈尔滨: 哈尔滨理工大学, 2007.  
Wang Yujing. Image-based Turning Surface Roughness measurement[D]. Harbin: Harbin University of Science and Technology, 2007. (in Chinese with English abstract)
- [16] Moreno Ra, Alvarez M C D, Alfonso A M T, et al. Shadow analysis of soil surface roughness compared to the chain set method and direct measurement of micro-relief[J].

- Biogeosciences, 2010, 7(8): 2477-2487.
- [17] 沈昌蒲. 机械化土壤耕作[M]. 北京: 中国农业出版社, 1995, 15-18.
- [18] 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.
- [19] 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. Effect of soil bulk density on maize root hydraulic conductivity under dry and wet conditions[J]. Bulletin of the Soil Science Society of China, 2003, 40(5): 779-782. (in Chinese with English abstract)
- [20] 杨玮, 兰红, 李民赞, 等. 基于图像处理和 GBRT 模型的表土层土壤容重预测[J]. 农业机械学报, 2020, 51(9): 193-200.
- Yang Wei, Lan Hong, Li Minzan, et al. Prediction of top soil layer bulk density based on image processing and Gradient Boosting Regression Tree Model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 193-200. (in Chinese with English abstract)
- [21] 孔凡伟. 如何精确土壤容重[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 with English abstract)
- [22] 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]. 农业机械学报, 2019, 50(S1): 102-107.
- Meng Chao, Yang Wei, Zhang Miao, et al. Vehicle-mounted real-time soil conductivity and mechanical resistance measurement system[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(S1): 102-107. (in Chinese with English abstract)
- [23] 王明迁, 李丹阳, 郝威凯, 等. 基于 HSV 颜色模型的图像识别技术研究[J]. 科技资讯, 2020, 18(35): 1-2, 8.
- Wang Mingqian, Li Danyang, Hao Weikai, et al. Research on image recognition technology based on HSV color model[J]. Science & Technology Information, 2020, 18(35): 1-2, 8. (in Chinese with English abstract)
- [24] 包全磊. 基于 HSV 空间的彩色图像分割[J]. 软件导刊, 2010, 9(7): 171-172.
- Bao Quanlei. Color image segmentation based on HSV space[J]. Software Guide, 2010, 9(7): 171-172. (in Chinese with English abstract)
- [25] 潘思赣, 李振. 基于颜色空间与神经网络车牌识别技术研究[J]. 中国交通信息化, 2020(S1): 132-136.
- Pan Sigan, Li Zhen. Research on license plate recognition technology based on color space and neural network[J]. China ITS Journal, 2020(S1): 132-136. (in Chinese with English abstract)
- [26] 徐婷婷, 吉晓东, 李文华, 等. 基于颜色和纹理特征的胶囊内镜图像分类[J]. 现代电子技术, 2018, 41(19): 58-62.
- Xu Tingting, Ji Xiaodong, Li Wenhua, et al. Capsule endoscopic image classification based on color and texture feature[J]. Modern Electronics Technique, 2018, 41(19): 58-62. (in Chinese with English abstract)
- [27] 徐衍鲁. 基于颜色特征的图像检索技术综述[J]. 电脑知识与技术, 2017, 13(13): 156-158, 163.
- Xu Yanlu. Overview of Image retrieval technology based on color features[J]. Computer Knowledge and Technology, 2017, 13(13): 156-158, 163. (in Chinese with English abstract)
- [28] 岳磊. 基于分块颜色矩和灰度共生矩阵的图像检索[J]. 微计算机信息, 2012, 28(8): 162-164.
- Yue Lei. Image retrieval based on block color moment and gray level co-occurrence matrix[J]. Microcomputer Information, 2012, 28(8): 162-164. (in Chinese with English abstract)
- [29] 刘兰芳, 谭秉霖, 张可, 等. 基于主成分分析的湖南省城市雾霾灾害风险评估[J]. 灾害学, 2021, 36(1): 76-81.
- Liu Lanfang, Tan Binglin, Zhang Ke, et al. Risk assessment of urban smog disasters in hunan province based on principal component analysis[J]. Journal of Catastrophology, 2021, 36(1): 76-81. (in Chinese with English abstract)
- [30] 肖建忠, 肖雨彤, 李佳锶. 基于主成分分析的我国天然气市场高质量发展评价[J]. 中国石油大学学报: 社会科学版, 2021, 37(1): 20-27.
- Xiao Jianzhong, Xiao Yutong, Li Jiasi. Evaluation of high quality development of my country's natural gas market based on principal component analysis[J]. Journal of China University of Petroleum (Edition of Social Sciences), 2021, 37(1): 20-27. (in Chinese with English abstract)
- [31] Abdi M J, Giveki D. Automatic detection of erythemato-squamous diseases using PSO-SVM based on association rules[J]. Engineering Applications of Artificial Intelligence, 2013, 26(1): 603-608.
- [32] Liu Z W, Cao H R, Chen X F, et al. Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings[J]. Neurocomputing, 2013, 99: 399-410.
- [33] 赵静, 王选仓, 丁龙亭, 等. 基于灰色关联度分析和支持向量机回归的沥青路面使用性能预测[J]. 重庆大学学报, 2019, 42(4): 72-81.
- Zhao Jing, Wang Xuancang, Ding Longting, et al. Asphalt pavement performance prediction based on gray correlation analysis and support vector machine regression[J]. Journal of Chongqing University, 2019, 42(4): 72-81. (in Chinese with English abstract)
- [34] 黄啸. 支持向量机核函数的研究[D]. 苏州: 苏州大学, 2008.
- Huang Xiao. Research on Kernel Function of Support Vector Machine[D]. Suzhou: Soochow University, 2008. (in Chinese with English abstract)

## Predicting bulk density and porosity of soil using image processing and support vector regression

Yang Wei, Lan Hong, Li Minzan<sup>\*</sup>, Meng Chao

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

**Abstract:** Soil structure is essential for plant development and moisture balance, generally representing the spatial heterogeneity of different components or properties of soil. In this case, bulk density and porosity of soil are important parameters to evaluate the soil structure. In the traditional measurement, the ring knife is normally used to measure soil bulk density. But this commonly-used measurement requires multiple instruments, such as ring cutters, aluminum boxes, and drying boxes, although the measured data is accurate to serve as a standard requirement. Particularly, the whole process is time-consuming and labor-intensive, unsuitable for the rapid and accurate measurement of soil bulk density in a large range of farmland in recent years. Therefore, it is highly urgent to explore a convenient, efficient, and indirect measurement of soil bulk density, especially for the input variables for most prediction models in precision agriculture. In this study, prediction models of soil bulk density and porosity were constructed with the soil resistance using image processing and Support Vector Regression (SVR). The color and texture parameters of the soil surface image were also used to characterize the soil roughness, according to the correlation between roughness, resistance, and bulk density of soil. A measuring system was developed to mount a vehicle for soil resistance. In image processing, HSV color space was used for the threshold segmentation, while the first-order distance, second-order moment, and third-order moment of HSV three components were taken as color parameters. The specific texture parameters included the energy, entropy, contrast, and inverse variance of the gray-level co-occurrence matrix. Principal component analysis was used to extract the principal components of color and texture parameters for the non-correlation between the input parameters. The correlation analysis was then made between the prediction of the SVR model and the standard value measured by the ring knife. Specifically, the coefficient of determination  $R^2$  of the SVR model reached 0.867 for the prediction of soil bulk density, the coefficient of determination  $R^2$  of decision tree regression model reached 0.734 for the prediction of soil bulk density, and the SVR model root mean square error was 0.001 g/cm<sup>3</sup>, indicating better performance than that of decision tree regression. Nevertheless, the calculation time took 6.810 s, about 4.7 s longer than the 2.153 s calculation time of decision tree regression. In soil porosity, the coefficient of determination  $R^2$  of the SVR model was 0.743, and the root mean square error was 2.284. The coefficient of determination  $R^2$  was 0.690 for the decision tree regression model, the root mean square error was 3.345. The calculation time of the SVR model was 3.144 s, less than the duration of the decision tree regression model at 4.302 s. It demonstrated that the SVR model can widely be expected to predict soil bulk density and porosity using color and texture parameters combined with soil resistance as input variables. In the case of small and medium-sized data samples, SVR model can achieve good prediction results. The finding can provide a sound reference for the rapid and effective prediction of bulk density and porosity in soil.

**Keywords:** image processing; soils; bulk density; porosity; HSV image segmentation; Support Vector Regression prediction model