

# 基于机器学习 XGBoost 的机制砂细粉含量预测<sup>①</sup>



李姝彤, 李 伟, 高 尧, 杨 明, 丁健刚

(长安大学 信息工程学院, 西安 710064)

通信作者: 李 伟, E-mail: [grandy@chd.edu.cn](mailto:grandy@chd.edu.cn)

**摘 要:** 机制砂是由碎石或者砾石经制砂机反复破碎加工至粒径小于 2.36 mm 的人工砂. 在实验中把机制砂中的石粉含量和含泥量称为细粉含量, 细粉含量表征机制砂的洁净程度. 本文提出了一种基于 XGBoost 网络的机制砂细粉含量预测方法. 首先, 利用完全封闭的图像采集设备对机制砂细粉制成的溶液进行图像采集, 保证外界光线不会对图像拍照造成影响, 之后进行图片裁剪、读取 RGB 值、转 LCH 颜色空间等预处理, 然后构建 XGBoost 网络模型, 通过贝叶斯原理进行参数的循环迭代, 之后进行模型优化, 使模型的  $r2\_score$  更高, 最终实现对机制砂细粉含量的预测. 结果表明: 该模型预测的数据的  $r2\_score$  可以达到 0.967 762, 相比于传统的多元线性回归模型、BP 神经网络、传统 XGBoost 网络预测的  $r2\_score$  为 0.896 144、0.914 598、0.950 670, 预测精度有明显提高. 在实际应用中, 该方法可以缩短机制砂细粉含量测量时间, 简化机制砂细粉含量测量步骤, 是一种新型的预测机制砂细粉含量的方法.

**关键词:** 机制砂; 细粉含量; 预测模型; XGBoost; 贝叶斯优化

引用格式: 李姝彤, 李伟, 高尧, 杨明, 丁健刚. 基于机器学习 XGBoost 的机制砂细粉含量预测. 计算机系统应用, 2023, 32(3): 256-264. <http://www.c-s-a.org.cn/1003-3254/8992.html>

## Prediction of Fine Powder Content in Manufactured Sand Based on Machine Learning XGBoost

LI Shu-Tong, LI Wei, GAO Yao, YANG Ming, DING Jian-Gang

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

**Abstract:** Manufactured sand refers to artificial sand whose particle size is less than 2.36 mm after the repeated crushing of gravels by sand-making machines. In experiments, stone powder and mud contents in the manufactured sand are called fine powder content, which represents the cleanliness of the manufactured sand. In this study, a method for predicting the fine powder content in the manufactured sand based on the XGBoost network is proposed. First, a completely closed image acquisition device is used to collect images of a solution made of fine powders in the manufactured sand, so as to guarantee that the outside light will not affect shooting. Then pre-treatment is carried out, such as picture cropping, RGB value reading, and LCH color space shifting, and an XGBoost network model is built. Through the Bayes principle, loop iteration of parameters is conducted, and the model is optimized, so as to make the  $r2\_score$  of the model higher and finally predict the fine powder content in the manufactured sand. The results show that the  $r2\_score$  of the data predicted by this model can reach 0.967 762. In addition, the  $r2\_score$  predicted by the traditional multiple linear regression models, BP neural network, and traditional XGBoost network is 0.896 144, 0.914 598, and 0.950 670. In contrast, the prediction accuracy of the proposed model is significantly improved. In practical application, this method can shorten the measurement time and simplify the measurement steps of the fine powder content in the manufactured sand. Therefore, it is a new method for predicting the fine powder content in manufactured sand.

**Key words:** manufactured sand; fine powder content; prediction model; XGBoost; Bayesian optimization

<sup>①</sup> 收稿时间: 2022-08-01; 修改时间: 2022-09-01; 采用时间: 2022-09-27; csa 在线出版时间: 2022-11-18

CNKI 网络首发时间: 2022-11-21

机制砂是粒径小于 2.36 mm 的颗粒碎块。经过开采天然岩石,对天然岩石进行除土,对其进行机械的破碎,之后用 2.36 mm 的筛子进行筛分得到机制砂。在工业应用中,混凝土的用处非常广泛,例如:各种建筑桥梁的建设过程中,混凝土就有非常重要的用途,而机制砂是构成混凝土非常重要的原材料。机制砂的质量严重影响着混凝土的性能,特别是机制砂在进行开采岩石,除土,破碎的过程中容易引入一些非机制砂的物质,这些物质的含量无论高低都会对机制砂的性能产生影响。机制砂的杂质含量可以用细粉含量<sup>[1]</sup>来表示。目前对于机制砂的细粉含量的测量是用传统的规范《公路工程集料试验规程》提及的细集料砂当量试验,但该方法程序繁琐且时间久,由于人为的观察、环境及其他因素,导致该检测方法的准确性和高效性难以保证。

目前,关于机制砂,大多数人研究的是机制砂对混凝土的影响或者改进细集料砂当量试验的方法,而对于机制砂细粉含量的数据预测还是比较少的。胡胜荣<sup>[2]</sup>研究了路面性能和细集料的关系,路面主要研究的是沥青混合料的路面。陆宇等<sup>[3]</sup>研究了机制砂砂当量对沥青混合路面性能的影响,该研究对于机制砂的工业应用具有非常深远的意义。马兴等<sup>[4]</sup>研究了砂当量试验仪的内容,主要研究了该试验仪的试验原理和功能,建立了研究所用的模型。朱立无等<sup>[5]</sup>制定了《砂当量试验仪校准规范》,该规范中规定的校准的方法非常合理和科学,可以为之后研究机制砂砂当量的测量提供一个较为标准的对比参考。

用预测网络预测机制砂细粉含量的现阶段研究较少,但是如今对于数据预测相关内容却越来越常见。如:2019年,为解决原砂含泥量检测的不精确和步骤繁琐的问题,郭景宝等<sup>[6]</sup>设计了一种原砂含泥量检测装置,但是其并未解决全部问题,虽然提高了检测精度,但是并未放慢了检测过程。2019年,刘泽宇<sup>[7]</sup>研究了浊度和含泥量的对应关系,一般情况下,含泥量越高,配置成的溶液越浑浊,所以浑浊度和含泥量是具有对应关系的,刘泽宇具体化了这种关系。2019年郭景宝<sup>[8]</sup>主要研究原砂的相关内容,之前进行原砂含泥量的检测都较为复杂,为了缩短检测时间,其研究出一种新型检测含泥量的装置。2020年,马旭梁等<sup>[9]</sup>提出了一种可以进行速度较快的检测原砂含泥量的方法,解决了之前研究所出现的检测精度不高的问题。2021年,王风雅<sup>[10]</sup>提出一种利用计算机图像分析技术定量测定原砂含泥量

的方法。首先将泥砂分离,采用较为传统的方法:涡洗法,之后以数字图像分析技术为基础,根据不同的含水量其洗砂液颜色会发生变化,采用数字图像采集技术对涡洗之后的原砂浊液进行图像采集,再结合数字图像处理分析技术,对原砂含泥量进行预测。

传统的细粉含量的测量用到《公路工程集料试验规程》中的细集料砂当量试验(国外的某些地方也会用到亚甲蓝试验),需要经过试验制备(需各种化学试剂:氯化钙、甘油、甲醛)、测含水率、配置冲洗液、多次震荡等步骤,程序非常繁琐。综上所述,若将预测网络运用到机制砂细粉含量预测,将可以很好地解决机制砂细粉含量预测繁杂、效率低、耗时间等问题。故本文针对机制砂溶液图片的亮度、色度、色调等特征提出一种 XGBoost 预测网络来解决机制砂细粉含量预测难题。针对本文的预测方法,具体实现思路是对机制砂细粉所制得的溶液拍照,提取其 RGB 颜色特征,转换颜色空间至 LCH,将 L, C, H 这 3 个分量取均值,输入 L, C, H 值和细粉含量值进入 XGBoost 网络进行训练,得到预测值。

## 1 算法和预处理

### 1.1 Boosting 算法

在一般的监督式学习中,会有固定的偏差,Boosting 算法可以减小这种偏差,提高精确度。Boosting 算法示意图如图 1 所示。

#### 1) 弱学习

在做预测工作时,训练的网络的精确度不是很高,可以说明该预测或识别网络属于弱学习器。该学习效率不是很高,需要再进行模型优化来达到学习的效果。

#### 2) 强学习

强学习是相对于弱学习而言的,在做识别和预测工作时,训练的网络的精度很高,可以说明该预测或识别网络属于强学习器。

#### 3) Boosting 算法

在进行模型训练时,需要确定样本集,样本集是进行网络训练的基础。进行预处理之后需要进行加强融合,加强融合可以结合不同分类器的优点,对于模型训练有很大的帮助作用。

### 1.2 XGBoost 网络

#### 1) 目标函数

损失函数和惩罚函数共同组成 XGBoost 模型的目

标函数<sup>[11-13]</sup>, 提高预测精度通过最小化损失函数来达到. 目标函数的化简方法较为简单, 一般情况下, 对目标函数进行泰勒展开, 泰勒展开式可以使目标函数更加直观, 可以使模型较快运行, 较快得出预测结论. 图2为机制砂细粉含量预测流程.

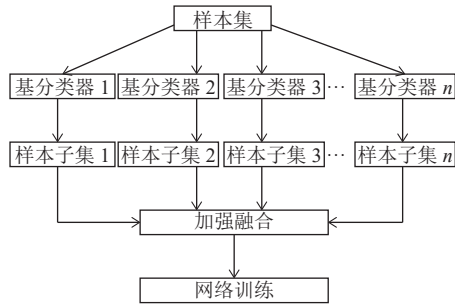


图1 Boosting 算法示意图

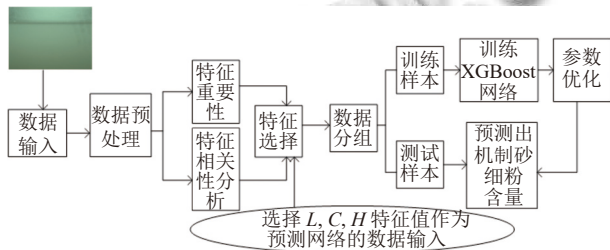


图2 机制砂细粉含量预测流程

### 2) 基于决策树<sup>[14,15]</sup>的目标函数

决策树是一种运用概率分析进行决策的方法, 决策树的决策运用到很多概率论的内容, 是根据概率论的内容来进行决策. 图3为决策树在预测中的流程.

决策树的复杂度是由叶子节点的多少决定, 目标函数正则项可以用式(1)表示:

$$\Omega(f) = \gamma T + \frac{1}{2} \gamma \sum_{j=1}^T w_j^2 \quad (1)$$

也就是说, 决策树的模型复杂度由两个变量决定: 一个是叶子节点数量, 表示生成的决策树的叶子节点数量之和; 另外一个为目标函数正则项, 该正则项将所有节点的权重进行计算, 可以很好地作为模型复杂度的构成因子.

### 1.3 改进型 XGBoost 网络-贝叶斯优化

XGBoost 算法精确度还有待提升, 运用贝叶斯优化可以进行模型参数的选择. 本文用贝叶斯优化的方法进行模型改进<sup>[16-20]</sup>, 将模型的  $r2\_score$  由 XGBoost 的 0.95 提升至改进型 XGBoost 的 0.96 左右.

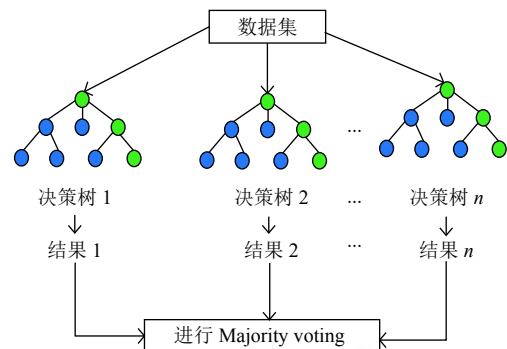


图3 决策树在预测中的流程

#### 1) 贝叶斯定理

贝叶斯定理的公式如下:

$$p(f|D_i) = \frac{p(D_i|f)p(f)}{p(D_i)} \quad (2)$$

其中,  $f$  表示目标函数, 但是其在初始迭代阶段是未知的,  $D_i$  表示已知的集合, 用类似位置坐标表示, 即  $(x_1, y_1), (x_2, y_2), \dots$ , 其中  $x_i$  表示用来进行决策的向量.

#### 2) 贝叶斯优化步骤

贝叶斯定理是贝叶斯优化参数优化的基础, 同时, 下一个采样的超参数组合也是参数优化的基础. 贝叶斯优化可以较为充分的利用某一个采样点的位置信息, 它的方式是通过学习目标函数形状, 不断迭代, 当结果向全局最大提升时, 迭代的参数就是本文所要确定的结果. 贝叶斯优化流程如图4所示.

假设一组超参数组合是  $X=x_1, x_2, \dots, x_n$ , 超参数和函数的关系如下:

$$x^* = \arg \min f(x) \quad (3)$$

其中,  $x$  表示超参数, 在 XGBoost 中, 超参数就是学习率、树的最大深度等这些参数.

凸优化可以很好地解决二阶导数大于零出现的各种问题, 例如每次计算超参数的时候, 会进行人工操作来计算结果, 耗费许多人力资源. 为了解决这一问题, 本文假设高斯分布为搜索的模型空间, 利用高斯过程, 每次计算得到的结果都比上次计算得到的结果更优, 得到该结果是利用迭代的方法贝叶斯优化的步骤.

① 输入:  $f$  (每次输入一组超参数得出的一个值)、 $X$  (超参数的定义域)、 $S$  (可以进行参数选择的算法)、 $M$  (对初始化数据进行拟合得到的模型).

② 初始化: 初始化获取数据集  $D=(x_1, y_1), \dots, (x_n, y_n)$ , 其中  $y_i=f(x_i)$ .

③ 循环迭代参数: 若是按照手动选参的方法, 每次

选中参数, 需要计算它的  $f(x)$ , 这样需耗费较多的人力. 为了使选参过程更加快速, 本文固定选参次数, 具体步骤如下: 利用高斯过程, 设模型  $M$  服从高斯分布, 且已知了数据集  $D$ , 通过计算得出具体的模型具体函数表示.

高斯过程是多元高斯分布向无穷维的扩展, 高斯过程函数的分布可以由均值函数和协方差函数组成:

$$f(x) \sim GP(m(x), k(x_i, x_j)) \quad (4)$$

其中,  $GP$  表示高斯过程,  $m(x)$  是均值函数,  $k(x_i, x_j)$  表示协方差函数.

在式 (4) 中,  $k(x_i, x_j)$  表示的协方差函数即是高斯过程的核函数, 核函数是高斯过程的核心, 其起衡量任意两个点之间的距离的作用:

$$k(x_i, x_j) = \sigma^2 \exp\left(-\frac{\|x_i - x_j\|_2^2}{2l^2}\right) \quad (5)$$

其中,  $x_i, x_j$  表示本文中模型运行时不同时间点,  $\|x_i - x_j\|_2^2$  表示一个二范式, 表示  $x_i, x_j$  之间的距离.

将模型建立成如下的形式:

$$y = f(x) + N(0, \sigma^2) \quad (6)$$

其中,  $f(x)$  表示高斯过程函数的分布,  $N(0, \sigma^2)$  表示模型建立过程中的噪音,  $\sigma^2$  表示高斯过程的方差.

建模之后, 本文尝试抽样进行样本计算, 该优化方法在局部最优解上不断采样.

④ 确定最优参数: 通过多次迭代找到最优参数, 达到模型优化.

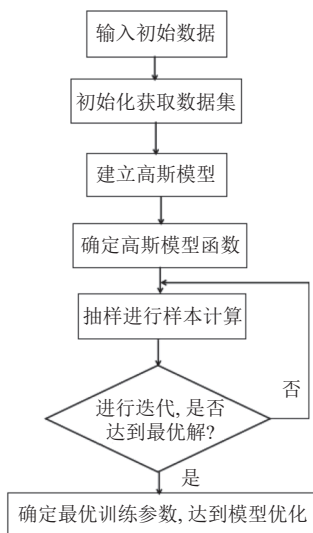


图4 贝叶斯优化流程

### 1.4 图像预处理

图片预处理流程如图5所示.



图5 图片预处理流程

#### 1) 图片裁剪

图6为原始溶液图像, 图7为不同细粉含量的原始图片裁剪之后的图片.



图6 拍摄原始溶液图像

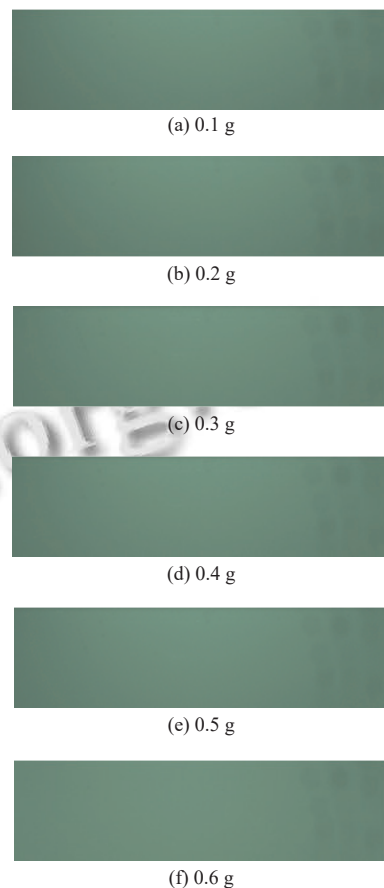


图7 不同细粉含量的原始图片裁剪之后的图片

#### 2) 图片裁剪区域的确定

图8-图10分别为不同区域  $L, C, H$  值与细粉含量的关系.

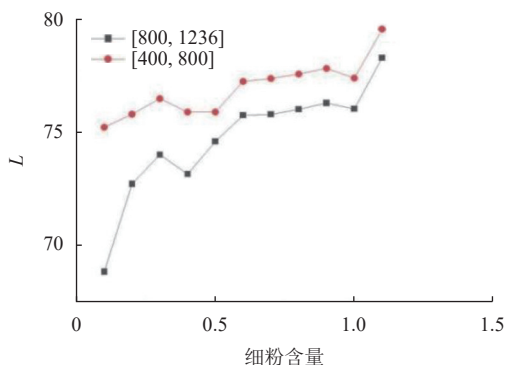


图8 不同区域 L 值与细粉含量的关系

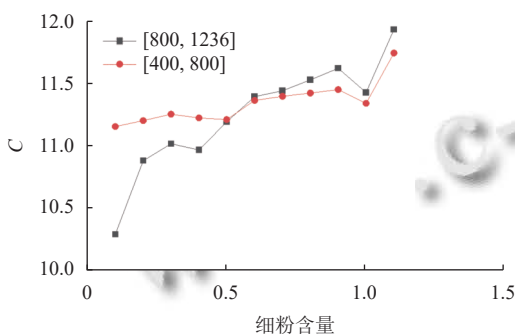


图9 不同区域 C 值与细粉含量的关系

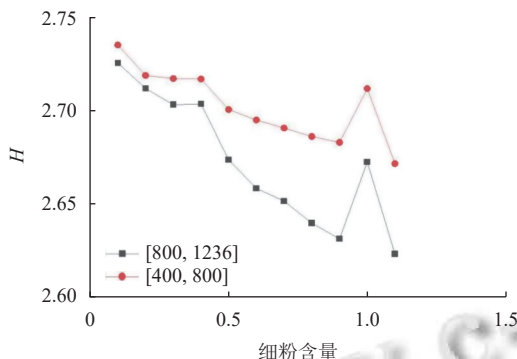


图10 不同区域 H 值与细粉含量的关系

图像大小为 1628×1236, 现有溶液的位置为 [400, 1236; 0, 628], 选取  $x$  的位置为 [0, 1628], 将  $y$  位置分为两个区域 [400, 800] 和 [800, 1236].

当  $y$  的取值范围为 [800, 1236] 时, 特征值  $L$ 、 $C$ 、 $H$  值在不同细粉含量之间存在较大的波动, 且各特征值与细粉含量的线性关系较差; 当  $y$  的取值范围为 [400, 800] 时, 特征值  $L$ 、 $C$ 、 $H$  值在不同细粉含量之间波动较小, 且各特征值与细粉含量的线性关系较好. 所以选取特征区域时, 选择  $x$  的取值范围为 [0, 1628],  $y$  的取值范围为 [400, 800].

2) 读取特征区域的 RGB 值

一张图片是由许多个像素组成的, 彩色图片的每个像素有 3 个通道, 分别为 RGB 通道, 其分别对应  $R$ 、 $G$ 、 $B$  值, 图像的  $R$ 、 $G$ 、 $B$  值可以表征图像内容.

3) 将图片的 RGB 值全部转换成 LCH 值

LCH 颜色空间首次将亮度和色度分开表示, 且该颜色空间是国际上通用颜色体系之一. 将 RGB 转换到 LCH, 需要用中间颜色空间 Lab 来转换, 转换公式如下:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (7)$$

$$\begin{cases} L^* = 116f\left(\frac{Y^*}{100}\right) - 16 \\ a^* = 500 \left[ f\left(\frac{X^*}{95.047}\right) - f\left(\frac{Y^*}{100}\right) \right] \\ b^* = 200 \left[ f\left(\frac{Y^*}{100}\right) - f\left(\frac{Z^*}{108.883}\right) \right] \end{cases} \quad (8)$$

其中,  $f$  函数为:

$$f(t) = \begin{cases} t^{\frac{1}{3}}, & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & \text{otherwise} \end{cases} \quad (9)$$

$$\begin{cases} X^* = \frac{X}{255 \times 0.950456} \\ Y^* = \frac{Y}{255} \\ Z^* = \frac{Z}{255 \times 1.088754} \end{cases} \quad (10)$$

$$\begin{cases} h^* = \arctan \frac{b^*}{a^*} \\ c^* = \sqrt{a^{*2} + b^{*2}} \end{cases} \quad (11)$$

4) 对转换得到的 LCH 值取均值

LCH 颜色空间中,  $L$  表示亮度,  $C$  表示饱和度,  $H$  表示色相, 对该 3 项取均值做预测网络的输入值.

2 实验仪器和流程

2.1 实验仪器

本文实验所用的叶轮搅拌机, 玻璃杯, 相机如图 11-图 13 所示.

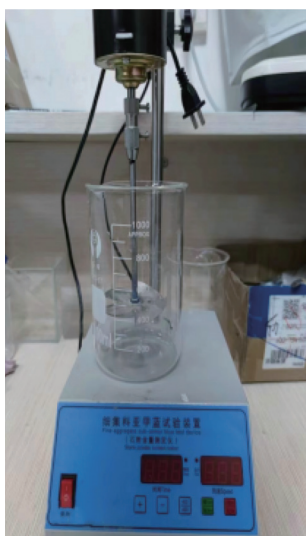


图 11 叶轮搅拌机



图 12 玻璃杯



图 13 相机

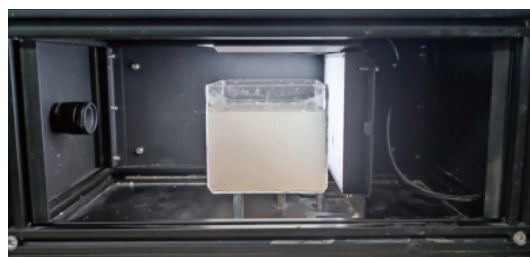


图 14 图像采集暗箱

## 2.2 试验流程

### 1) 模型训练试验流程

用烘箱开 300 °C 烘干机制砂, 时间可用机制砂烘干情况来定; 之后用 0.075 mm 的筛子筛出石粉和泥,

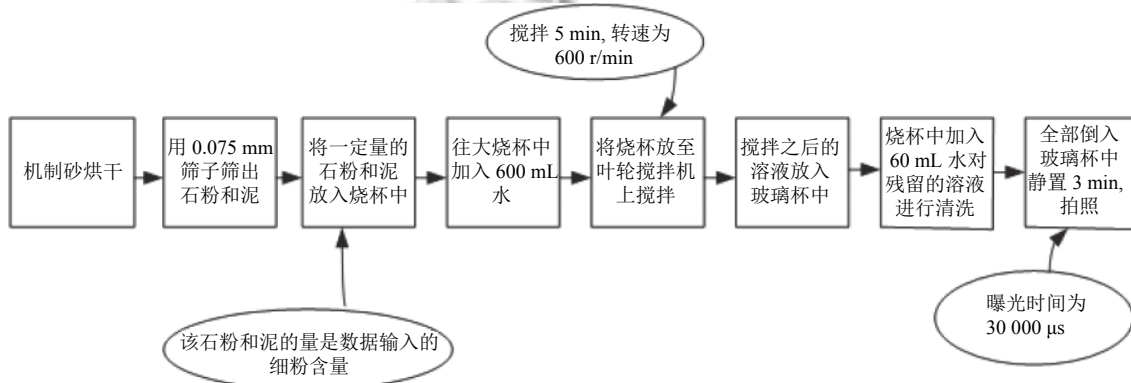


图 15 机制砂实验流程

静置时间过短, 细粉还没有完全沉在玻璃杯的底部; 静置时间过长, 部分细粉会沉到玻璃杯底部导致检

取出一定量的石粉和泥放入烧杯中, 在烧杯中加入 600 mL 水, 并放到叶轮搅拌机上, 设置搅拌时间为 5 min, 转速为 600 r/min, 开始搅拌; 搅拌之后倒入玻璃杯中, 用 60 mL 的水对玻璃杯进行清洗, 将清洗的溶液倒入玻璃杯中, 开始计时, 静置 3 min 拍照, 拍照时采取曝光时间为 30 000 μs, 采取上侧光源和右侧光源 (图 14 中的玻璃杯上侧和右侧都有光源且都可以固定) 同时开启, 之后拍照, 将拍照之后的图片进行预处理得到预测网络的输入数据. 机制砂实验流程如图 15 所示.

测不准确. 选取在搅拌 300 s 后静置的第 2.5 min 开始每隔 30 s 采集一张图像, 到第 10 min, 特征值与静置时

间的变化规律如图 16 所示。

从图 16 可以发现在提取区域不变的情况下特征值  $L, C, H$  与不同的细粉含量会出现如下变化规律. 在静置 3 min 时, 从图 16 可以发现特征值  $L$ 、特征值  $C$  和特征值  $H$  在不同的含泥量之间的差异明显无重合现象; 在静置 7-10 min 时间时, 细粉含量为 0.5 和 0.6 的  $L$  几乎相近, 明显有重合现象; 在 4-7 min 时, 细粉含量为 0.5 和 0.6 的  $C$  值相近, 所以不可作为静置时间; 在 2.5 min 时, 细粉含量为 0.8 和 0.7 的  $H$  值有明显的重合现象. 综上所述, 选取 3 min 为静置时间.

### 2) 实际应用中得到输入数据的试验流程

机制砂超声波清洗: 通过筛子筛足够长的时间之后, 再用超声波清洗机进行清洗, 发现仍有质量差. 此结果表明, 机制砂通过筛子筛出细粉无法完全筛出干净, 因为有些细粉是附着在机制砂上的. 所以, 在实际操作中, 直接筛出机制砂中的石粉, 进行称重来计算细粉含量是不现实的. 大量实验表明, 超声波清洗 9 min 可以将机制砂上附着的细粉完全震至水中. 所以采用该方法来实现细粉含量检测的数据输入. 经实验表明, 将石粉通过超声波震入水中, 和通过叶轮搅拌机将石粉放入水中的效果相同, 因为测的细粉含量和机制砂无关, 只和细粉有关. 我们在进行模型训练时, 可以采用通过使用叶轮搅拌机的方法 (约 15 min 左右) 来进行数据采集训练模型, 该方法较超声波清洗较为省时, 超声波清洗之后过滤得到的溶液图像进行预测. 超声波清洗的步骤如下.

- a) 取机制砂样品 200 g 烘干 2 小时.
- b) 将烘干的样本一半左右 (100 g 左右) 放入超声波清洗机, 倒入 300 mL 纯净水, 开始超声波清洗, 边清洗边搅拌, 设置时间为 9 min.
- c) 清洗之后将机制砂溶液全部倒入烧杯中.
- d) 将另一半样本放入超声波清洗机, 加入 300 mL 水, 重复步骤 b) 和步骤 c).
- e) 用 60 mL 水清洗超声波清机, 将清洗之后的溶液倒入烧杯中.
- f) 将烧杯中的溶液通过 0.075 mm 的筛子, 将机制砂筛出, 溶液倒入玻璃杯中.
- g) 将玻璃杯中的溶液静置 3 min, 拍照.

### 3) 确定超声波清洗时间

确定超声波清洗时间过程如图 17 所示.

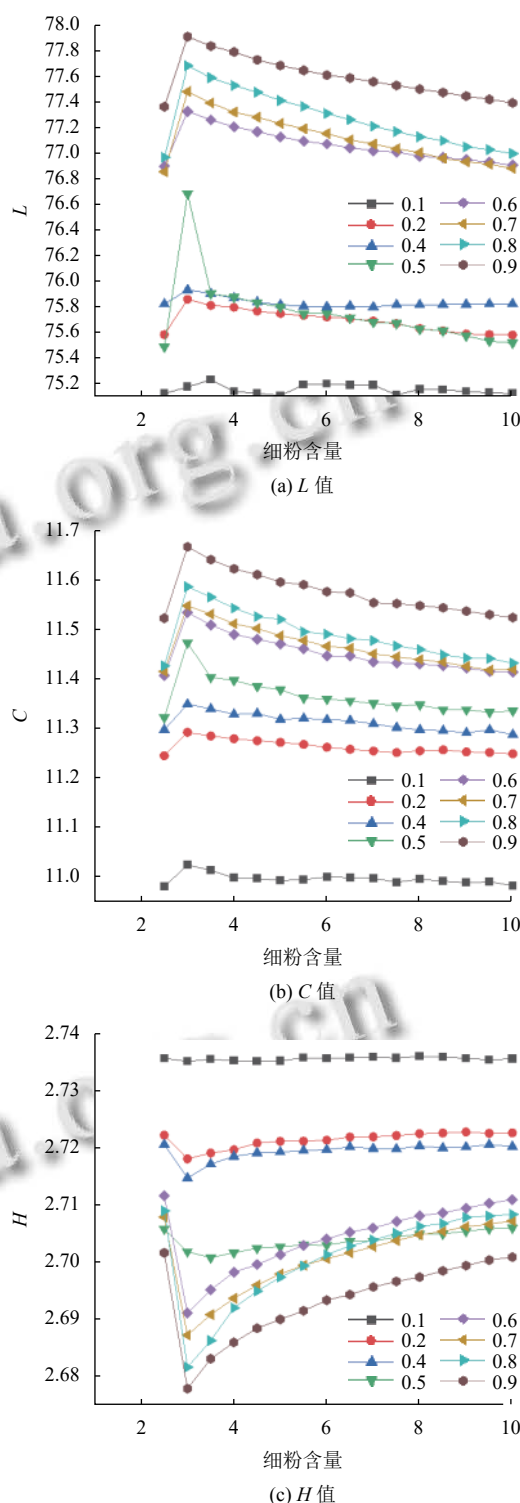


图 16 不同细粉含量的 LCH 值随静置时间的变化图

首先随机确定一个时间 7 min, 进行超声波清洗步骤 a)–f), 对筛出的机制砂烘干, 测量质量, 当设置超声波清洗时间为 8 min 时, 若烘干质量小于设置为 7 min 的烘干质量时, 继续增加分钟数, 继续试验, 直到找到

使机制砂质量稳定的最小分钟数;若设置 8 min 的烘干质量等于 7 min 的烘干质量时,应逐渐减少清洗分钟数,直到找到数量增加的分钟数,来确定机制砂质量稳定的最小分钟数,通过多次试验,得机制砂超声波清洗能将机制砂上附着得细粉全部震下来的时间为 9 min.表 1 为清洗时间和剩余质量的关系.

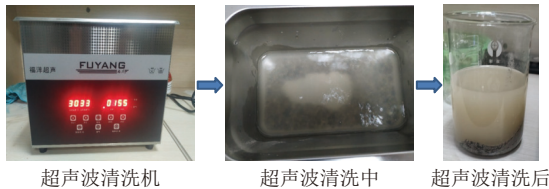


图 17 机制砂超声波清洗示意图

表 1 清洗时间和剩余质量关系

清洗时间 (min)	200 g 机制砂剩余质量 (g)
7	194.24
8	193.56
9	193.01
10	192.98

### 3 实验结果与分析

#### 3.1 模型预测

图片大小为 1628×1236,采集之后首先进行预处理,预处理之后的数据需要划分为训练集和测试集.

在本实验中共进行了 41 组数据的实验,其中 32 组为训练集,9 组为测试集.

#### 3.2 实验结果

##### 1) 模型预测结果

表 2 中实测值和预测值都为归一化之后的结果,表中实测值是指直接称重的测量值归一化之后的结果,预测值为模型训练验证集的结果.

表 2 实测值和预测值对比

溶液图像	实测值	预测值
1	0.6153846	0.5480589
2	0.2051282	0.2043766
3	0.5128205	0.5480589
4	0.4102564	0.3887776
5	0.9230769	0.9072618
6	0.5641026	0.5480589
7	0.8461538	0.9072618
8	0.4358974	0.3887766
9	0.7179487	0.73571

表 3 为本文算法与多元线性回归模型比较.  $MAE$  (平均绝对误差),  $MSE$  (均方误差),  $r2\_score$  (决定

系数) 这 3 个评价指标计算方式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (12)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (13)$$

$$r2\_score = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

表 3 本文算法与多元线性回归模型比较

模型	MAE	MSE	r2_score
多元线性回归	0.224622	0.061907	0.896144
BP神经网络	0.065945	0.007824	0.914598
XGBoost网络	0.156159	0.029404	0.950670
改进XGBoost	0.031404	0.001445	0.967762

$MAE$  表示平均绝对误差,该值越小,模型越好;  $MSE$  表示均方误差,该值越小,模型越好;  $r2\_score$  表示预测值和真实值的接近程度,该值越接近 1,表示预测值和近似值越接近,模型越好.

##### 2) 本文方法与传统方法的比较

表 4 为本文方法与传统方法的比较.由表 4 可知,相比于传统的砂当量试验和亚甲蓝试验,本文的方法较为省时,所需试剂较少.相对于砂当量试验中求絮凝物和物质总高度之比和亚甲蓝试验中通过亚甲蓝试剂进行试纸色晕,本文的方法都较为简单,耗费人力较少.

表 4 本文方法与传统方法相比较

对比项目	本文方法	砂当量试验	亚甲蓝试验
耗时长度	30 min 左右	4 小时左右	5 小时左右
所用试剂	无	氯化钙、甘油、甲醛	亚甲蓝

#### 3.3 本文算法 LCH 重要性

本文算法 LCH 重要性分布图如图 18.由图中可以看出,  $L$ ,  $C$ ,  $H$  这 3 个特征分量中,亮度  $L$  的重要性最大,饱和度  $C$  次之,色相  $H$  最小.因为溶液的细粉含量的多少直观地影响到溶液的明暗程度,是细粉含量越大,溶液越暗.饱和度是指图片色彩的纯洁性,色彩越纯洁,饱和度越低,在本试验中,溶液的细粉含量对于饱和度的影响较弱,说明即使溶液有细粉含量区别,但是色彩纯洁性区别不大.色相是包括红橙黄绿青蓝紫 8 种色彩,溶液的细粉含量对该特征的影响最小.



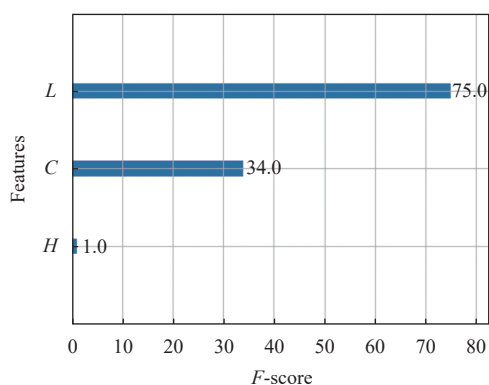


图 18 LCH 重要性分布图

#### 4 结论与展望

本文提出了一种基于机器学习 XGBoost 的机制砂溶液细粉含量检测系统, 通过对机制砂筛出小于 0.075 mm 的石粉和泥的量来表示细粉含量的多少. 实验过程要经历筛出、制成溶液、拍摄、预处理、模型预测等步骤, 实验步骤仅需十几分钟, 可以较快完成机制砂细粉含量定量预测分析. 本文通过贝叶斯优化对 XGBoost 参数进行多次迭代, 可以提高 XGBoost 的  $r2\_score$ , 同时可以节省计算参数结果的时间. 通过实验表明, 该模型预测的  $r2\_score$  可以达到 0.967 762, 相比于传统的多元线性回归模型、BP 神经网络模型、XGBoost 模型的  $r2\_score$  有所提高, 同时, 该模型的预测精度可以达到 0.1, 在实际的生产应用中, 可以缩短细粉含量计算时间, 简化细粉含量计算实验流程, 有非常重要的价值. 在之后的机制砂研究中, 可以采用该实验方法进行细粉含量的预测, 该方法对于进行机制砂测量智能化高效化具有重要的意义.

#### 参考文献

- 盛余飞. 石粉含量对机制砂砂浆孔结构特性的影响研究. 四川建材, 2020, 46(5): 11–12. [doi: 10.3969/j.issn.1672-4011.2020.05.007]
- 胡胜荣. 细集料砂当量对沥青混合料路用性能的影响分析. 运输经理世界, 2021, (2): 131–132. [doi: 10.3969/j.issn.1673-3681.2021.02.065]
- 陆宇, 杜骋, 金光来, 等. 细集料砂当量对沥青混合料路用性能的影响研究. 山西建筑, 2021, 47(3): 101–102, 152. [doi: 10.3969/j.issn.1009-6825.2021.03.038]
- 马兴, 郭贵勇. 砂当量试验仪的校准方法及其结果分析与研究. 科技创新导报, 2020, 17(1): 52, 55. [doi: 10.16660/j.cnki.1674-098X.2020.01.052]
- 朱立无, 朱岩. 砂当量试验仪校准结果不确定度评定. 计量与测试技术, 2019, 46(5): 107–110. [doi: 10.15988/j.cnki.

- 1004-6941.2019.5.034]
- 6 郭景宝, 马旭梁, 李大勇, 等. 一种快速检测原砂含泥量的自动装置. 铸造, 2019, 68(2): 181–185. [doi: 10.3969/j.issn.1001-4977.2019.02.015]
- 7 刘泽宇. 浊度法原砂含泥量检测方法及其装置 [硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2019.
- 8 郭景宝. 原砂含泥量快速检测装置的研制 [硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2019.
- 9 马旭梁, 刘泽宇, 李大勇. 浊度法检测原砂含泥量. 哈尔滨理工大学学报, 2020, 25(6): 126–130. [doi: 10.15938/j.jhust.2020.06.018]
- 10 王凤雅. 原砂含泥量的图像检测方法研究 [硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2021.
- 11 Zhu X, Chu J, Wang KD, *et al.* Prediction of rockhead using a hybrid N-XGBoost machine learning framework. Journal of Rock Mechanics and Geotechnical Engineering, 2021, 13(6): 1231–1245. [doi: 10.1016/j.jrmge.2021.06.012]
- 12 Chelgani SC, Nasiri H, Alidokht M. Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A “conscious-lab” development. International Journal of Mining Science and Technology, 2021, 31(6): 1135–1144. [doi: 10.1016/j.ijmst.2021.10.006]
- 13 Sheng S, Li R, Wang X, *et al.* Predicting the 7<sup>th</sup> day efficacy of acupoint application of Chinese herbs (*Xiao Zhong Zhi Tong Tie*) in patients with diarrhea—A machine-learning model based on XGBoost algorithm. World Journal of Traditional Chinese Medicine, 2022, 8(2): 257–261. [doi: 10.4103/wjtcn.WJTCM\_60\_21]
- 14 蒋妮, 程港, 贺思敏, 等. 基于决策树与多元线性回归模型的出生体重影响因素分析. 中国卫生统计, 2022, 39(2): 202–206.
- 15 高虹雷, 门昌骞, 王文剑. 多核贝叶斯优化的模型决策树算法. 国防科技大学学报, 2022, 44(3): 67–76. [doi: 10.11887/j.cn.202203009]
- 16 Zhao SJ, Shan YL, Gultepe I. Prediction of visibility in the Arctic based on dynamic Bayesian network analysis. Acta Oceanologica Sinica, 2022, 41(4): 57–67. [doi: 10.1007/s13131-021-1826-z]
- 17 何军, 林广东, 申小军, 等. 基于贝叶斯优化 XGBoost 的隧道沉降量预测. 计算机系统应用, 2022, 31(7): 379–385. [doi: 10.3969/j.issn.1003-3254.2022.07.050]
- 18 杨灿, 刘磊磊, 张遗立, 等. 基于贝叶斯优化机器学习超参数的滑坡易发性评价. 地质科技通报, 2022, 41(2): 228–238.
- 19 孙斌, 储芳芳, 陈小惠. 基于贝叶斯优化 XGBoost 的无创血压预测方法. 电子测量技术, 2022, 45(7): 68–74. [doi: 10.19651/j.cnki.emt.2108646]
- 20 刘青, 鲁成, 马天祥, 等. 基于贝叶斯优化 LightGBM 算法的主动式搜索时间调整方法. 华北电力大学学报 (自然科学版). <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJLAST&filename=HBDL20220716000&uniplatform=NZKPT&language=gb>. [2022-07-19].

(校对责编: 孙君艳)