

基于模型融合的钻进参数识别岩石类型研究

王亚飞¹, 张占荣¹, 刘华吉¹, 姚震桐^{*2}

(1. 中铁第四勘察设计院集团有限公司, 湖北 武汉 430063; 2. 中国地质大学(武汉)工程学院, 湖北 武汉 430074)

摘要: 岩石类型的识别对于钻进工程的安全和经济效益具有重要意义。钻进过程中的岩石类型实时识别大多是通过随钻测井, 但由于随钻测井成本昂贵, 在地质勘探领域鲜见应用。本文挖掘深部探测 Sinoprobe-05 项目金川科钻的钻进多参数(钻速、钻压、转速、钻头扭矩、泵压力、泵量)数据, 通过一种融合模型算法来识别岩石类型。首先采用 Savitzky-Golay 平滑滤波器降低钻进参数数据的噪声, 然后对数据进行了归一化。最后运用融合模型预测和识别岩石类别。融合模型的初级学习器为支持向量机、人工神经网络和随机森林, 通过次级学习器贝叶斯模型平均算法对每个模型的权重进行计算。结果表明, 多模型融合算法准确率为 0.9686, 比每个单独的算法准确率高。

关键词: 岩石类型; 钻进参数; 支持向量机; 随机森林; 模型融合

中图分类号: P634 文献标识码: A 文章编号: 2096-9686(2023)02-0017-09

Data-driven model for the identification of the rock type by drilling data

WANG Yafei¹, ZHANG Zhanrong¹, LIU Huaji¹, YAO Zhentong^{*2}

(1. China Railway Siyuan Survey and Design Group Co., Ltd., Wuhan Hubei 430063, China;

2. Faculty of Engineering, China University of Geosciences, Wuhan Hubei 430074, China)

Abstract: The identification of rock types is of great significance to the safety and economic benefits of drilling engineering. Real-time rock type during drilling. The identification is mostly through logging while drilling, but it is rarely used in the field of geological exploration due to the high cost of logging while drilling. In this paper, the drilling parameters (ROP, WOB, rotational speed, bit torque, pump pressure, pump volume) of Jinchuan Scientific Drill are used to identify rock types through a fusion model. First, the noise of the drilling parameter data is reduced by the Savitzky-Golay smoothing filter, and then the data is normalized. Finally, the fusion model is used to predict rock types. The primary learners of the fusion model are Support Vector Machines, Artificial Neural Networks and Random Forests, and the weights of each model are calculated by the secondary learner Bayesian model averaging algorithm. The results show that the accuracy of the multi-model fusion algorithm is 0.9686, which is higher than that of each individual algorithm.

Key words: rock type; drilling parameters; support vector machine; random forest; model fusion

0 引言

钻探是地质勘探的主要方法^[1]。岩石类型的识别在钻探领域有着广泛的应用, 如钻进参数优化和

钻头选择^[2]。目前, 大部分岩石类型识别是通过测井数据完成的, 使用前馈神经网络、K-means 聚类算法、KNN 聚类等机器学习算法完成对岩石类型的

收稿日期: 2022-10-20; 修回日期: 2022-12-28 DOI: 10.12143/j.ztgc.2023.02.003

基金项目: 湖北省重点研发计划项目“城市地下空间精细化探测与感知关键技术及装备”(编号: 2021BAA050); 国家重点研发计划项目“城市地下空间开发地下全要素信息精准探测技术与装备”(编号: 2019YFC0605101)

第一作者: 王亚飞, 男, 汉族, 1981年生, 高级工程师, 注册岩土工程师, 主要从事铁路工程地质勘察与路基设计工作, 湖北省武汉市武昌杨园和平大道 745 号, 5951229@qq.com。

通信作者: 姚震桐, 男, 汉族, 1997年生, 硕士在读, 地质工程专业, 主要研究方向为地质钻探装备与机具, 湖北省武汉市洪山区鲁磨路 388 号, yaozt@cug.edu.cn。

引用格式: 王亚飞, 张占荣, 刘华吉, 等. 基于模型融合的钻进参数识别岩石类型研究[J]. 钻探工程, 2023, 50(2): 17-25.

WANG Yafei, ZHANG Zhanrong, LIU Huaji, et al. Data-driven model for the identification of the rock type by drilling data[J]. Drilling Engineering, 2023, 50(2): 17-25.

识别^[3-11]。

随钻测井价格昂贵,在地质勘探领域中并未推广使用。钻机上钻进参数易于获得,许多学者利用钻进参数预测岩石强度或识别岩石类型,在实验室做钻孔实验和收集数据来研究钻孔数据与岩石的关系。LABELLE等^[12]使用神经网络根据顶锚式钻进参数(扭矩、钻压、转速、液压和钻头位置)识别岩石类型。KALANTARI等^[13]使用振动传感器和特殊的声学传感器方法来区分在不同类型岩石中钻孔产生的时频特性。KALANTARI等^[14]提出了基于物理模型来预测岩石强度参数的方法,从钻进数据(钻压、扭矩、钻速和转速)中预测粘聚力、内摩擦角和单轴抗压强度。FLEGNER等^[15]通过钻进过程中的振动声信号进行岩石类型识别。SILVA等^[16]比较了18种实验室生产的砂浆钻孔数据,建立了从其钻孔数据中识别砂浆成分的数学模型。SHANGXIN等^[17]通过可控的钻进参数估计最佳钻进效率和岩石强度。

以上的模型都是基于实验室获得的钻进数据,在实际钻进现场应用中性能未知。对于实际钻进现场,ZHOU等^[18]提出了一种基于数据驱动的自适应无监督方法,通过最小化表征措施的熵梯度来识别岩石类型。HE等^[19-20]提出了一种利用钻孔能量确定岩体岩石质量指数(RQD)的经验方法,利用钻进数据建立了一种用于快速确定岩石强度参数的深度卷积神经网络。OLORUNTOBI^[21]使用基于钻进比能的概念来识别岩石类型。LU等^[22]提出了一种分析模型,该模型使用来自孕镶金刚石钻头的钻孔

数据来预测岩石强度参数。

以上模型基于少数类型的岩石和解析模型,泛化性能比较差。当岩石类型更多时,这些模型将难以识别。为了找到一种可靠的方法来识别实际钻进现场的岩石类型,依托金川科钻配套的钻进参数检测系统采集数据和19种岩石的实物取心,对信息进行融合和对比验证,比较了融合模型和单独的3种机器学习算法(支持向量机、人工神经网络、随机森林)的结果。与随钻测井不同,钻进数据(钻速、钻压、转速、扭矩、泵压和泵流量)通常会在钻进过程中记录,因此无需额外费用,是一种非常经济的识别岩石的方法。

1 数据源与箱线图分析

数据源来自于金川科学钻探孔。甘肃金川科钻项目是大陆科学钻探选址与科学钻探试验项目的重点项目之一,隶属于国家科技部重点研发项目深部探测技术与试验研究项目(SinoProbe-05-01)。金川科钻立项的意义是:解决金川深层资源前景,研究中国镍、铜、铂族金属资源的可持续开发利用。完善创新知识体系,建立具有中国金川特色、具有独创性和全球影响力的成矿理论体系。解决古大陆板块边缘(或内部)构造单元等相关重大基础地质问题^[23]。

图1为金川科钻使用的HXY-9型钻机和金刚石钻头^[24]。LabVIEW编写的钻进数据监测系统可以检测钻速、钻压、转速、扭矩、泵压和泵流量^[25]。该数据库由11147个数据点组成,覆盖钻探进尺1204 m(孔深126~1330 m)。

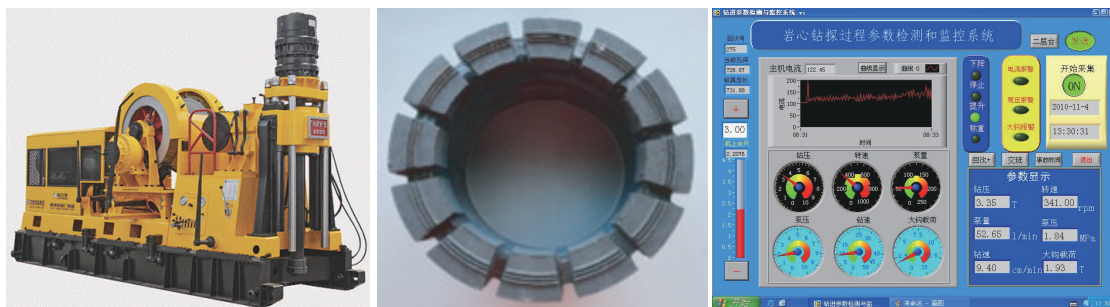


图1 金川科钻钻机、钻头及钻进参数采集系统

Fig.1 JCS D-1 drilling machine, bit and monitoring system

1.1 离群点处理

首先对数据的离群点采用 3σ 方法进行筛分,样本 x 和样本均值 μ 之间的距离为 $Z\text{-score}(x) = (x -$

$\mu)/\sigma$,将不满足条件 $|Z\text{-score}(x)| < 3$ 的样本视为离群值。最终筛分出10177个数据,图2为钻孔深度(126~1330 m)的岩石类型。转速、钻压和泵流量是

人为控制的正变量,扭矩、钻速和泵压是在钻进过程中随岩石类型而变化的被动变量。当操作变量发生变化时,被动变量也会发生变化,因此需要对这些参数进行综合以反映岩石的类型。图 3 显示了每种岩石类型中每个钻孔参数的箱线。从图 3 中可以看出每种岩石的每个钻孔参数的箱线图是不同的,因此利用钻孔参数识别岩石类型是可行的。

柱状	深度/m	岩石类别
	126~425	灰白色蛇纹岩夹杂灰色条带
	425~447	深灰色黑云母角闪岩夹层混合岩带
	447~559	浅灰色中粒混合岩夹斜长角闪岩和大理石镜质体
	559~570	紫红色粗粒混合花岗岩
	570~631	深灰色角闪岩夹紫红色带状混合岩(片麻岩)
	631~707	灰白色带状混合岩与深灰色角闪岩
	707~757	深灰色中细粒黑云母角闪岩
	757~792	灰色角闪岩与浅肉质混合红色混合岩
	792~892	紫红色钾长混合花岗岩夹灰色斜长角闪岩
	892~915	角闪岩夹大理石
	915~956	紫红色中粒混合花岗岩
	956~1002	灰白色大理石与灰黑色中粒角闪岩
	1002~1053	灰色角闪岩
	1053~1150	灰色蛇纹岩
	1150~1183	深灰色角闪岩
	1183~1224	灰白色大理石
	1224~1240	浅灰色钾盐混合岩断裂带
	1240~1271	深灰色黑色云绿泥片岩
	1271~1330	深灰色粗粒角闪岩变粒岩

图 2 孔深 126~1330 m 岩石类型

Fig.2 The rock type histogram of drilling depth 126~1330m

1.2 降噪

从图 3 中可以明显看出,一些值存在偏差,并且数据集中存在噪声数据。来自实际钻进作业的测量数据总是或多或少地受到噪声的影响。即使在最佳控制条件下,测量数据的误差也约为 5% 或更多。噪声数据会影响模型训练过程,增加训练时间并降低性能^[26]。因为 Savitzky-Golay 技术广泛用于钻进数据预处理^[27],研究时采用 Savitzky-Golay^[28]平滑滤波器,Savitzky-Golay 滤波器可以在不破坏信号原有特性的情况下对信号进行平滑处理,滤波前后对比如图 4 所示。

1.3 数据归一化

数据归一化是机器学习中的一项基本任务。每个钻进数据的数值大小不同,如果数据未进行归一化,训练时间会更长,数据驱动模型的性能会很差。需要对原始钻进数据进行归一化处理,消除对机器学习算法结果的影响。原始数据经过归一化处理后,各项指标处于同一数量级,适合综合比较评价。数据处理时,使用最小最大值归一化来归一化原始钻进数据,如式(1):

$$f(x_i) = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

式中: x_{max} —— 样本数据的最大值; x_{min} —— 样本数据的最小值。

归一化后,所有钻进数据值都落在 [0, 1] 区间内。

2 方法

2.1 融合模型

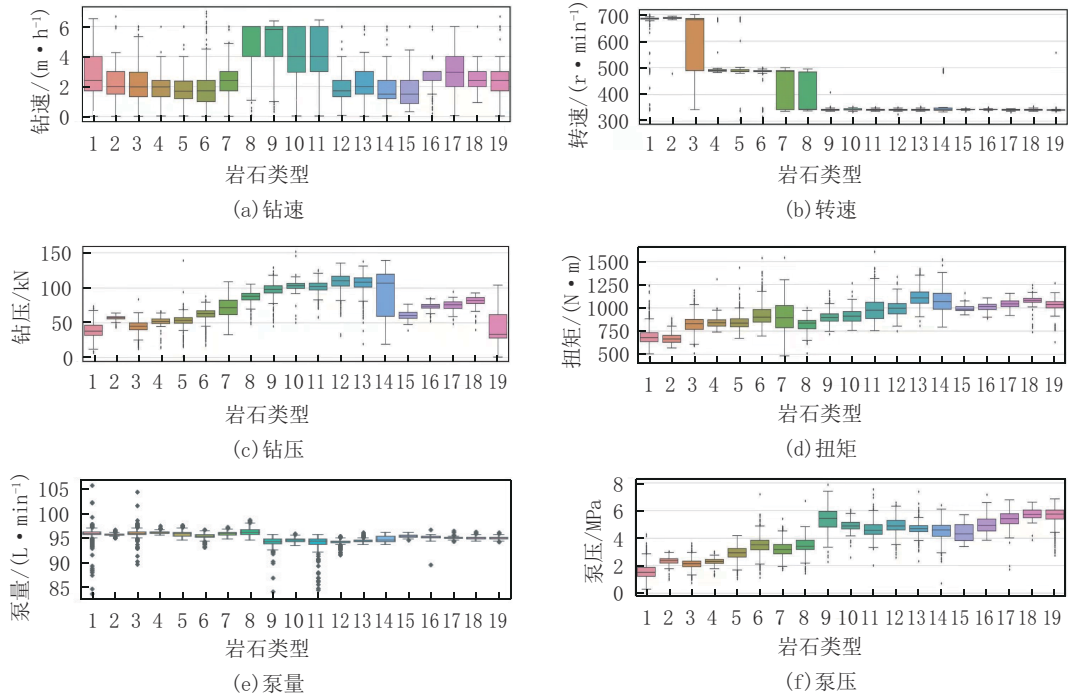
钻进参数之间存在很强的相关性,如图 5 所示,特征会冗余,采用单个模型往往容易陷入局部最优,泛化性能较差。采用模型融合的方法可以提高泛化性能,提高模型精度。

模型融合是指将多个学习器进行融合。模型融合可以减少单个学习器的偏差。模型融合误差由一个取决于单个学习器的误差和模型融合策略算法,定义如下:

$$E = \bar{E} - \bar{A} \quad (2)$$

式中: \bar{E} —— 个体学习者的泛化错误,它取决于个体学习者 E_i 的错误和组合策略算法; \bar{A} —— 个体学习器的差异性,取决于个体学习者 A_i 和融合策略算法。

从式(2)可以明显看出,增加个体学习器多样



1—灰白色蛇纹岩夹杂灰色条带;2—深灰色黑云角闪岩夹层混合岩带;3—浅灰色中粒混合岩夹斜长角闪岩和大理石镜质体;4—紫红色粗粒混合花岗岩;5—深灰色角闪岩夹紫红色带状混合岩(片麻岩);6—灰白色带状混合岩与深灰色角闪岩;7—深灰色中细粒黑云角闪岩;8—灰色角闪岩与浅肉质混合红色混合岩;9—紫红色钾长混合花岗岩夹灰色斜长角闪岩;10—角闪岩夹大理岩;11—紫红色中粒混合花岗岩;12—灰白色大理石与灰黑色中粒角闪岩;13—灰色角闪岩;14—灰色蛇纹岩;15—深灰色角闪岩;16—灰白色大理石;17—浅灰色钾盐混合岩断裂带;18—深灰色黑色云绿泥片岩;19—深灰色粗粒角闪岩变粒岩

图3 不同岩石种类的钻进参数箱线

Fig.3 Boxplot of each drilling parameters in each type of rock

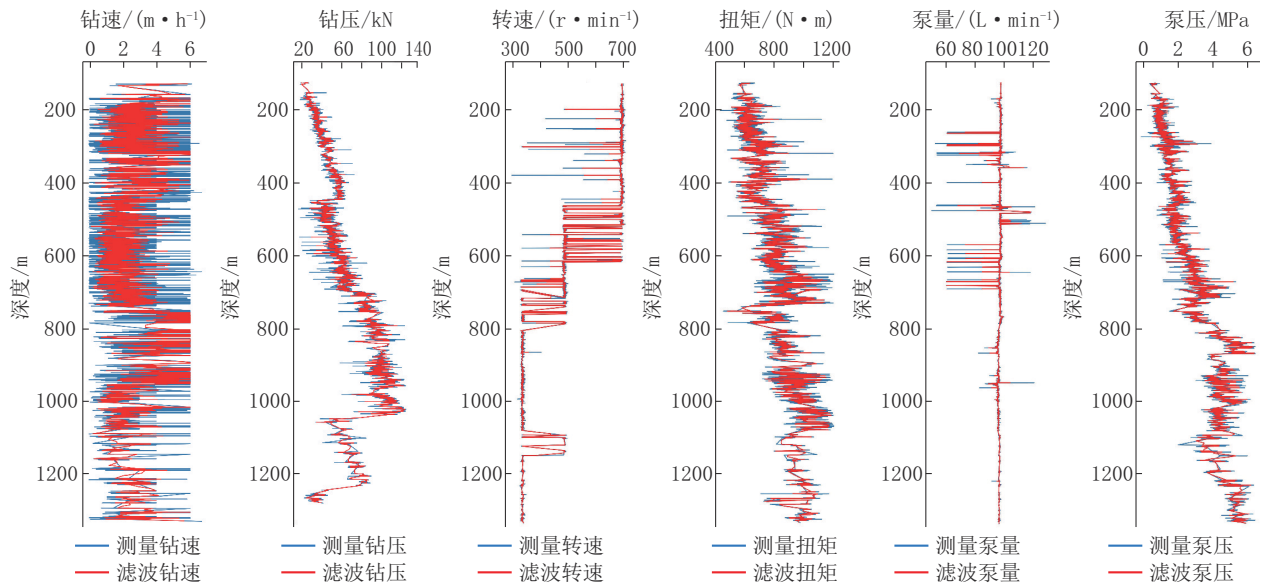


图4 滤波前后对比

Fig.4 Comparison of original and denoised data

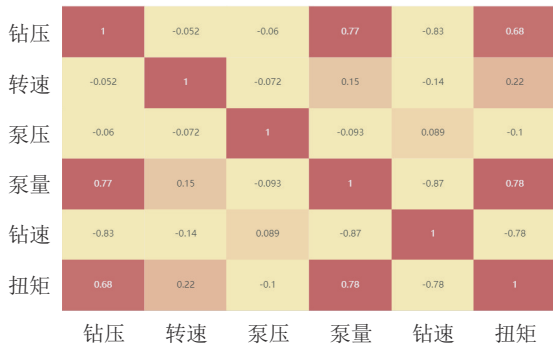


图 5 钻进参数的 Pearson 相关性

Fig.5 Pearson correlation of drilling parameters

性和减少单个学习器误差将改善总体误差。如果使用一个好的算法来融合不同的个体学习器,以最小化个体学习器的泛化误差,最大化歧义,那么我们可以获得最佳的融合模型。目前常用的人工智能模型有支持向量机、人工神经网络、随机森林、极端森林、梯度提升机、LightGBM、AdaBoost 等。因为随机森林、极端森林、梯度提升机、LightGBM、AdaBoos 均为集成学习算法,差异度较小,所以只选用常用的随机森林作为一个算法。因此,本文选取了支持向量机、人工神经网络和随机森林这 3 种差异大的模型采用贝叶斯模型平均方法进行融合。贝叶斯模型平均方法适用于大信息集合,能够最大限度地运用和反映岩石类别信息。由于事前对所有可能的模型进行加权平均,所以贝叶斯模型平均方法能够避免人为主观遴选解释变量造成的信息损失。该方法以贝叶斯统计理论为基础严格推演出模型后验概率分

布,并以此作为模型权重,有效地解决了不同模型的权重问题。图 6 展示了模型融合的伪代码。

```

Algorithm 1: Stacked
Input: Training dataset
       D = {(x1, y1), (x2, y2), ..., (xm, ym)};
       Learning algorithm L1, L2, ..., LT;
       Learning algorithm L.
1: for t=1,2,...,T do
2:   ht = Lt(D);
3: end for
4: D' = ∅
5: for i=1,2,...,T do
6:   for t=1,2,...,T do
7:     zit = ht(xi);
8:   end for
9:   D' = D' ∪ ((zi1, zi2, ..., ziT), yi);
10: end for
11: h' = L1(D')
Output: H(x) = h'(h1(x), h2(x), ..., hT(x))
    
```

图 6 模型融合算法伪代码

Fig.6 Pseudocode of staked model algorithm

本文的算法框架如图 7 所示,首先使用训练集数据来训练初级学习器(支持向量机、人工神经网络、随机森林)。为了避免过拟合,使用 5 折交叉验证对初级学习器进行训练。得出训练好的初级学习器后,将训练集和测试集输入到初级学习器中。训练集在初级学习器的输出被用作新的训练数据,测试集在初级学习器的输出的平均值作为新的测试数据。之后新的数据集输入到贝叶斯模型平均算法中,然后得到最终的预测值。

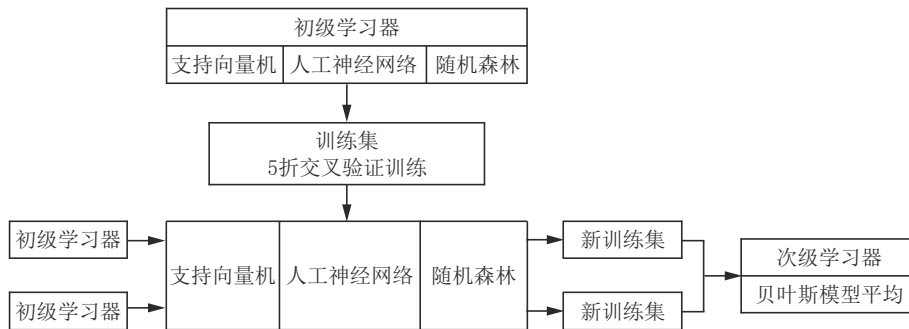


图 7 本文模型融合结构

Fig.7 Structure of stacked model

2.2 支持向量机

支持向量机方法是由 Vanpik 和贝尔实验室合作开发的^[29]。支持向量机原理基于统计学习理论

和结构最小化,与许多机器学习技术使用的传统经验风险最小化相比,它已显示出更高的性能^[30]。

分类任务通常涉及将数据分成训练集和测试

集。训练集中的每个实例都包含一个“目标值”和几个“属性”。支持向量机的目标是建立一个模型,该模型仅在给定测试数据属性的情况下预测测试数据的目标值。给定实例标签对 (x_i, y_i) 的训练集, $i=1, \dots, l$ 。其中 $x_i \in R^n, y_i \in \{1, -1\}^l$, 支持向量机需要解决以下优化问题:

$$\min_{w, b, \xi} \frac{1}{2} W^T W + C \sum_{i=1}^l \xi_i \quad (2)$$

约束:

$$\begin{aligned} y_i(w^T \phi(x_i) + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0 \end{aligned} \quad (3)$$

将低维空间难以处理的非线性分类问题转化为高维空间,得到最佳分类超平面。为了将低维非线性问题映射到高维规划线性问题,开发了核函数。支持向量机具有不同的内核(线性内核、多项式内核、径向基函数内核等)来执行高维特征空间中的任务。

2.3 人工神经网络

人工神经网络已被证明在解决许多超出经典数学和传统程序的计算能力的复杂工程问题方面提供了高水平的能力。人工神经网络模型由称为神经元的基本处理单元组成。人工神经网络由学习算法、传递函数和网络结构3部分组成。它至少由3层组成:输入层、隐藏层和输出层。人工神经网络建模的第一步是训练网络。数据通过输入层到达隐藏层,然后到达输出层。在输出层,将数据与实际数据进行比较。实际数据和预测数据之间的差异被传回模型,以更新每个连接之间的权重和每层的偏差。这样,所有数据集都会继续训练,直到平均误差降低到某个定义的限制^[31]。

人工神经网络在处理模糊数据、随机数据和非线性数据方面优势明显。特别适用于规模大、结构复杂、信息不清晰的系统。

2.4 随机森林

随机森林是Bagging的扩展变体。基于Bootstrap Sample的Bagging是并行集成学习方法最著名的代表^[32]。Bagging是Bootstrap和聚合的组合。对 N 个样本的数据集进行Bootstrap采样,得到数据集 D_i 。训练模型在 D_i 上重复 M 次,得到 M 个模型,然后通过平均来降低模型的方差。

随机森林的基础学习器是决策树。在决策树的训练过程中进一步增加了随机属性的选择。具体来

说,传统决策树在选择分区属性时会在当前节点的属性集中选择一个最优属性。在随机森林中,对于决策树中的每个节点,从节点的属性集中随机选择一个包含 k 个属性的子集,然后从子集中选择一个最优属性进行划分。随机森林简单,易于实现,计算开销低,在许多任务中具有强大的性能。被誉为“代表综合学习技术的方法”。随机森林中基本学习器的多样性不仅来自样本干扰,还来自属性干扰,通过增加每个学习器之间的差异程度来提高泛化性能^[33]。

2.5 贝叶斯模型平均算法

贝叶斯模型平均是基于贝叶斯理论将模型自身不确定性考虑在内的统计分析方法。其由一个模型空间 $M = \{M_1, M_2, \dots, M_n\}$ 来生成:

$$P(\theta|D) = \sum_{j=1}^K P(M_j|D) P(\theta|M_j, D) \quad (4)$$

式中: M_j ——单个模型; θ ——待估参数向量; K ——模型空间中含有模型的个数。

模型后验概率为:

$$P(M_j|D) = \frac{P(D|M_j) P(M_j)}{\sum_{h=1}^K P(D|M_h) P(M_h)} \quad (5)$$

$$P(D|M_j) = \int P(D|\theta_j, M_j) P(\theta_j|M_j) d\theta_j \quad (6)$$

式(6)表示模型 M_j 所对应的似然函数积分。其中, θ_j 表示模型 M_j 所对应的参数向量, $P(\theta_j|M_j)$ 表示模型 M_j 所对应的参数先验概率分布, $P(D|\theta_j, M_j)$ 表示模型 M_j 所对应的似然函数。 $P(M_j)$ 表示模型 M_j 的先验分布。

$$E[\theta|D] = \sum_{j=0}^K E[\theta|D, M_j] P(M_j|D) \quad (7)$$

$$\begin{aligned} \text{Var}[\theta|D] = & \sum_{i=0}^K \left(\text{Var}[\theta|D, M_j] + \left(E[\theta|D, M_j] \right)^2 \right) - E[\theta|D]^2 \end{aligned} \quad (8)$$

式(7)和(8)表明,基于贝叶斯模型平均方法得到的参数向量 θ 的后验均值也是加权平均形式。贝叶斯模型平均建模方法假定未知参数向量 θ 不再是固定常数,而是与模型 M_j 一样服从某一特定的统计分布。贝叶斯模型平均方法以后验概率为权重对可能的单项模型进行加权平均,并以后验概率作为选择解释变量的客观标准,有效地处理了建模过程中普遍存在的模型不确定性问题。

3 结果

验证时,采用留一法来验证模型的性能。整个数据集的 80% (8141 点) 作为训练数据集, 20% (2036 点) 作为测试数据集。准确性用于验证模型的性能。对于样本数据集 D , 它是正确分类的样本在样本总数中的比例, 定义为式(9):

$$acc(f; D) = \frac{1}{m} \sum_{i=1}^m \mathbb{I}(f(x_i) = y_i) \quad (9)$$

机器学习算法是使用 scikit-learn 库实现的。每种算法的性能和超参数如表 1 所示, 各模型训练集与测试集结果见图 8, 融合模型在测试集上的准确率为 0.9686, 高于单个模型, 模型的泛化性能强。

表 1 模型性能对比

Table 1 Performance and hyperparameters of SVC, ANN, RF

	超参数寻优范围	最优超参数	训练精度	测试精度
支持向量机	kernel: {'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'} C: {1, 10, 100, 1000, 10000}	kernel: rbf C: 10000	0.9529	0.9209
神经网络	hidden_layer_sizes: {(64, 64,), (128, 128,), (256, 256,), (512, 512,), (1024, 1024,), (2048, 2048,)}	hidden_layer_sizes: (1024, 1024,)	0.9031	0.8826
随机森林	n_estimators: [10, 1000] max_depth: [3, 40]	n_estimators: 800 max_depth: 22	1.0	0.9322
模型融合			0.9982	0.9686

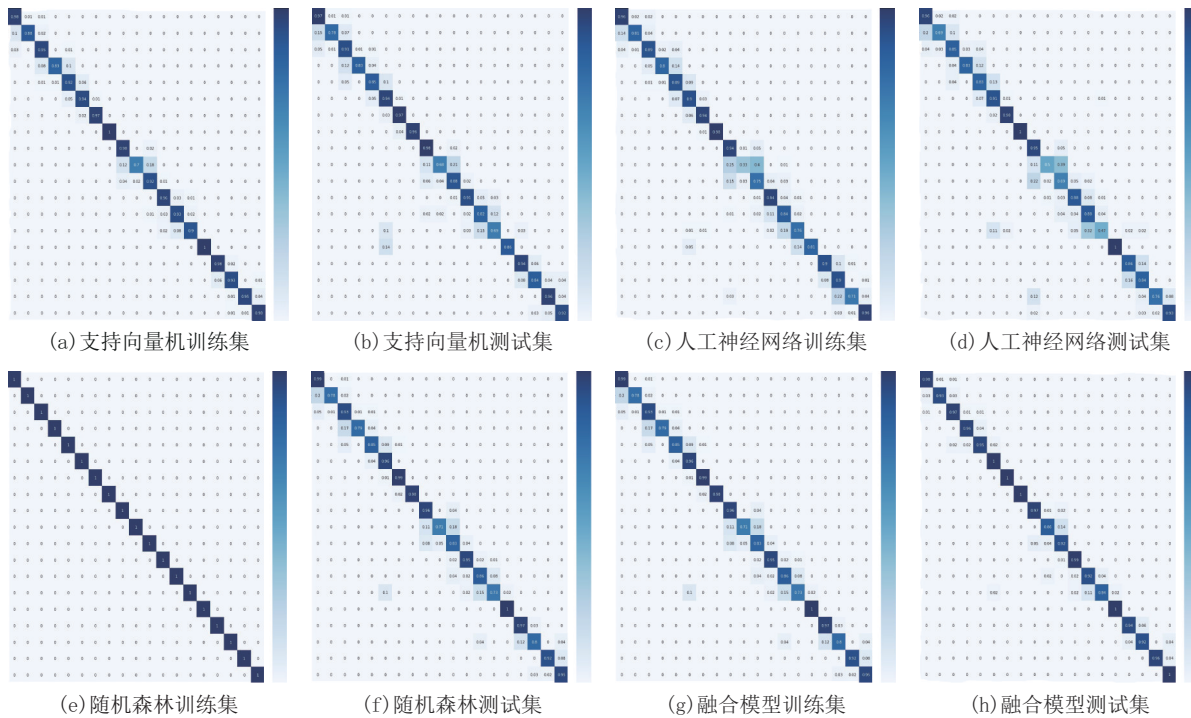


图 8 各模型训练集与测试集结果

Fig.8 Results of each model training set and testing set

4 结论

本文提出了一种融合模型预测岩石类别。模型融合算法的初级学习器为支持向量机、神经网络和随机森林, 通过次级学习器贝叶斯模型平均算

法对每个模型的权重进行计算。结果表明, 多模型融合算法准确率为 0.9686, 比每个独立的算法准确率高。支持向量机、神经网络和随机森林 3 个模型的差异大, 适合对其进行融合以提高模型精

度。贝叶斯模型平均方法适用于大信息集合,能够最大限度地运用和反映岩石类别信息。由于事前对所有可能的模型进行加权平均,所以贝叶斯模型平均方法能够避免人为主观遴选解释变量造成的信息损失。该方法以贝叶斯统计理论为基础严格推演出模型后验概率分布,并以此作为模型权重,有效地解决了不同模型的权重问题。

与随钻录井相比,该方法快速、经济,钻进数据监测系统可以将数据快速传输到计算机。通过钻进数据识别岩石类型在实际工程现场具有很好的应用价值。它可以帮助工程师控制钻进质量,优化钻进参数以获得最佳的钻速并判断是否到达目的层。基于同一区域的全部钻进数据,可以建立具有强泛化性能的模型。这样,通过高精度地识别岩石类型,可以高效、快速地完成钻进工程。

参考文献(References):

- [1] LI Y, SHE L, WEN L, et al. Sensitivity analysis of drilling parameters in rock rotary drilling process based on orthogonal test method[J]. *Engineering Geology*, 2020,270:105576.
- [2] QIN M, WANG K, PAN K, et al. Analysis of signal characteristics from rock drilling based on vibration and acoustic sensor approaches[J]. *Applied Acoustics*, 2018,140:275-282.
- [3] 周长春,姜杰,李谦,等.基于融合特征选择算法的钻速预测模型研究[J]. *钻探工程*,2022,49(4):31-40.
ZHOU Changchun, JIANG Jie, LI Qian, et al. Research on drilling rate prediction model based on fusion feature selection algorithm[J]. *Drilling Engineering*, 2022,49(4):31-40.
- [4] 张欣,田英英,韩泽龙,等.基于机器学习算法的井漏预测与诊断理论模型研究[J]. *钻探工程*,2022,49(2):58-66.
ZHANG Xin, TIAN Yingying, HAN Zelong, et al. Research on lost circulation prediction and diagnosis theoretical model based on machine learning algorithm [J]. *Drilling Engineering*, 2022,49(2):58-66.
- [5] 程万,孙家应,尹德战,等.深层泥页岩井壁失稳机理与预测模型研究进展[J]. *钻探工程*,2021,48(10):21-28.
CHENG Wan, SUN Jiaying, YIN Dezhan, et al. Research status of the wellbore failure mechanism and predicting model in deep mudstone and shale [J]. *Drilling Engineering*, 2021, 48 (10):21-28.
- [6] 李谦,曹彦伟,朱海燕.基于人工智能的钻速预测模型数据有效性下限分析[J]. *钻探工程*,2021,48(3):21-30.
LI Qian, CAO Yanwei, ZHU Haiyan. Discussion on the lower limit of data validity for ROP prediction based on artificial intelligence[J]. *Drilling Engineering*, 2021,48(3):21-30.
- [7] 陈晓君,陈小根,宋刚,等.基于人工神经网络模型的岩石特性预测[J]. *探矿工程(岩土钻掘工程)*,2019,46(1):34-38.
CHEN Xiaojun, CHEN Xiaogen, SONG Gang, et al. Prediction of rock characteristics based on artificial neural network model[J]. *Exploration Engineering (Rock & Soil Drilling and Tunneling)*, 2019,46(1):34-38.
- [8] 董青青,梁小丛.基于优化的BP神经网络地层可钻性预测模型[J]. *探矿工程(岩土钻掘工程)*,2012,39(11):26-28.
DONG Qingqing, LIANG Xiaocong. A model for predicting formation drillability based on optimized BP neural network[J]. *Exploration Engineering (Rock & Soil Drilling and Tunneling)*, 2012,39(11):26-28.
- [9] 熊虎林,李谦.基于地层成分和钻进参数的钻速预测模型[J]. *探矿工程(岩土钻掘工程)*,2018,45(10):195-201.
XIONG Hulin, LI Qian. ROP prediction model based on formation composition and drilling parameters [J]. *Exploration Engineering (Rock & Soil Drilling and Tunneling)*, 2018, 45(10): 195-201.
- [10] AO Y, ZHU L, GUO S, et al. Probabilistic logging lithology characterization with random forest probability estimation [J]. *Computers & Geosciences*, 2020,144:104556.
- [11] SABAH M, TALEBKEIKHAH M, WOOD D A, et al. A machine learning approach to predict drilling rate using petrophysical and mud logging data [J]. *Earth Science Informatics*, 2019,12(3):319-339.
- [12] LABELLE D, BARES J, NOURBAKHSI I. Material classification by drilling [C]//*Proceedings of the 17th International Symposium on Automation and Robotics in Construction. (International Association for Automation and Robotics in Construction)*:10. Citeseer, 2000.
- [13] KALANTARI S, HASHEMOLHOSSEINI H, BAGHBANAN A. Estimating rock strength parameters using drilling data [J]. *International Journal of Rock Mechanics and Mining Sciences*, 2018,104:45-52.
- [14] KALANTARI S, BAGHBANAN A, HASHEMALHOSSEINI H. An analytical model for estimating rock strength parameters from small-scale drilling data [J]. *Journal of Rock Mechanics and Geotechnical Engineering*, 2019,11(1):135-145.
- [15] FLEGNER P, KAČUR J, DURDÁN M, et al. Processing a measured vibroacoustic signal for rock type recognition in rotary drilling technology [J]. *Measurement: Journal of the International Measurement Confederation*, 2019,134:451-467.
- [16] SILVA A, NOGUEIRA R, SILVA A. Classification model of low-strength mortars from drilling data [J]. *Construction and Building Materials*, 2020,246:118484.
- [17] SHANGXIN F, YUJIE W, GUOLAI Z, et al. Estimation of optimal drilling efficiency and rock strength by using controllable drilling parameters in rotary non-percussive drilling [J]. *Journal of Petroleum Science and Engineering*, 2020,193:107376.
- [18] ZHOU H, HATHERLY P, RAMOS F, et al. An adaptive

- data driven model for characterizing rock properties from drilling data[C]//IEEE, 2011:1909-1915.
- [19] HE M, ZHANG Z, REN J, et al. Deep convolutional neural network for fast determination of the rock strength parameters using drilling data[J]. *International Journal of Rock Mechanics and Mining Sciences*, 2019,123:104084.
- [20] HE M, LI N, ZHU J, et al. Advanced prediction for field strength parameters of rock using drilling operational data from impregnated diamond bit[J]. *Journal of Petroleum Science and Engineering*, 2020,187:106847.
- [21] OLORUNTOBI O, BUTT S. Application of specific energy for lithology identification[J]. *Journal of Petroleum Science and Engineering*, 2020,184:106402.
- [22] LU Y, LI C, HE Z, et al. Variations in the physical and mechanical properties of rocks from different depths in the Songliao Basin under uniaxial compression conditions[J/OL]. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 2020,6(3):43[2021-10-05].
- [23] 杨经绥,许志琴,汤中立,等.大陆科学钻探选址与钻探实验[J].*地球学报*,2011,32(S1):84-112.
YANG Jingsui, XU Zhiqin, TANG Zhongli, et al. Continental Scientific Drilling: Site selection and pilot holes[J]. *Acta Geoscientica Sinica*, 2011,32(S1):84-112.
- [24] 董海燕,欧阳志勇,吴海霞,等.深部探测金川预导孔深孔钻探钻头的应用与分析[J].*探矿工程(岩土钻掘工程)*,2013,40(9):41-46.
DONG Haiyan, OUYANG Zhiyong, WU Haixia, et al. Application of the bit for pre-pilot hole deep drilling in Jinchuan deep exploration[J]. *Exploration Engineering (Rock & Soil Drilling and Tunneling)*, 2013,40(9):41-46.
- [25] 罗光强,胡郁乐.科学深钻DPI-1智能化多功能钻参仪的研制与应用研究[J].*地质与勘探*,2014,50(4):777-782.
LUO Guangqiang, HU Yule. Design and application of the DPI-1 intelligent drilling parameter instrument for scientific drilling[J]. *Geology and Exploration*, 2014,50(4):777-782.
- [26] GARCIA L P F, CARVALHO A C P L F De, LORENA A C. Neurocomputing effect of label noise in the complexity of classification problems [J]. *Neurocomputing*, 2015, 160: 108-119.
- [27] ASHRAFI S B, ANEMANGELY M, SABAH M, et al. Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field[J]. *Journal of Petroleum Science and Engineering*, 2019, 175: 604-623.
- [28] SAVITZKY A, GOLAY M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. *Analytical Chemistry*, 1964,36(8):1627-1639.
- [29] VAPNIK V N, CHERVONENKIS A. A note on one class of perceptrons [J]. *Automation and Remote Control*, 1964, 25 (1):821-837.
- [30] BELLO O, TEODORIU C, YAQOUB T, et al. Application of artificial intelligence techniques in drilling system design and operations: A state of the art review and future research pathways[C]//Society of Petroleum Engineers—SPE Nigeria Annual International Conference and Exhibition, 2016.
- [31] Agwu O E, Akpabio J U, Alabi S B, et al. Artificial intelligence techniques and their applications in drilling fluid engineering: A review[J]. *Journal of Petroleum Science and Engineering*, 2018,167:300-315.
- [32] Efron B, Tibshirani R J. *An Introduction to the Bootstrap*[M]. New York: CHAPMAN & HALL, 1993.
- [33] Pavlov Y L. Random forests[J]. *Machine Learning*, 2001, 45: 5-32.

(编辑 荐华)