

基于 TBM 岩机信息的隧洞断层超前智能感知 加权投票模型研究

李 龙¹, 刘造保¹, 周宏源¹, 齐文彪², 查文华³

(1. 东北大学 资源与土木工程学院, 深部金属矿山安全开采教育部重点实验室, 辽宁 沈阳 110819;

2. 吉林省水利水电勘测设计研究院, 吉林 长春 130021; 3. 东华理工大学 土木与建筑工程学院, 江西 南昌 330013)

摘要: 为实现 TBM 隧洞不良地质的超前判识, 基于吉林引松工程 TBM 隧洞 40.8 亿条施工数据, 采用机器学习方法开展隧洞断层超前智能感知研究。通过分析 TBM 岩-机交互参数(刀盘扭矩和顶护盾压力等)的变化规律, 将断层附近掘进段划分为平稳段、预警段和断层段; 借助皮尔逊相关性分析, 优选出 7 个关键参数作为断层感知的特征属性; 以随机森林和支持向量机为基分类器, 构建隧洞断层智能感知的加权集成投票模型。选取断层附近 771 组 TBM 施工有效掘进循环数据, 以准确率为目标对所建立模型进行训练和测试(539 组训练, 232 组测试); 应用准确率、召回率、F₁-分数评估加权集成投票模型进行断层超前感知的性能, 采用部分依赖图分析关键交互参数在不同预测类别中的敏感度。结果表明, 刀盘扭矩和顶护盾压力等关键交互参数在断层附近显现出不同程度的响应规律, 加权集成投票模型能有效预测断层附近的平稳段、预警段和断层段, 准确率达 89.22%。可为 TBM 隧洞断层前兆特征预警分析和预控措施提供支撑依据。

关键词: 岩石力学; TBM 大数据; 断层智能感知; 机器学习; 隧道智能建造; 加权集成投票模型

中图分类号: TU 45

文献标识码: A

文章编号: 1000-6915(2020)增2-3403-09

Weighted voting model for advanced intelligent perception of tunnel faults based on TBM rock-machine information

LI Long¹, LIU Zaobao¹, ZHOU Hongyuan¹, QI Wenbiao², ZHA Wenhua³

(1. Key Laboratory of Ministry of Education on Safe Mining of Deep Metal Mines, College of Resources and Civil Engineering, Northeastern University, Shenyang, Liaoning 110819, China; 2. Jilin Province Water Resource and Hydropower Consultative Company, Changchun, Jilin 130021, China; 3. School of Civil and Architectural Engineering, East China University of Technology, Nanchang, Jiangxi 330013, China)

Abstract: In order to realize the advance perception of unfavorable geology of TBM tunnel, this paper is devoted to tunnel fault intelligent perception using machine learning methods based on 4.08 billion records of TBM tunnel construction data of Jilin Yinsong project. The variation laws of TBM rock-machine interaction parameters (cutter-head torque and top shield pressure, etc.) are analyzed, and the excavation segments near the fault were divided into stable section, early warning section and fault section. Seven key parameters were selected as the features for fault perception by Pearson correlation analysis. A weighted integrated voting model was constructed for the intelligent perception of tunnel faults using the random forest and support vector machine method as the

收稿日期: 2019-12-18; **修回日期:** 2020-05-21

基金项目: 兴辽英才计划青年拔尖人才项目(XLYC1807094); 国家重点基础研究计划(973 项目)(2015CB058100)

Supported by the Young Talent Program of Liaoning Province(Grant No. XLYC1807094) and National Basic Research Program of China(Grant No. 2015CB058100)

作者简介: 李 龙(1994-), 男, 2015 年毕业于东北大学工业工程专业, 现为博士研究生, 主要从事 TBM 掘进参数优化控制与隧道不良地质智能感知方面的研究工作。E-mail: long_li@hotmail.com。通讯作者: 刘造保(1983-), 男, 现任教授、博士生导师。E-mail: liuzaobao@mail.neu.edu.cn

DOI: 10.13722/j.cnki.jrme.2019.1217

base classifiers. The effective tunneling cycle data of 771 groups of TBM construction near the fault was selected, and the model was trained (539 groups) and tested (232 groups) with accuracy as the target. The accuracy, recall rate and F_1 -score were used to evaluate the performance of the weighted integrated voting model for fault advance perception. Partial dependence plot was used to determine the sensitivity of key interaction parameters in different prediction categories. The results show that the key interaction parameters such as cutter-head torque and top shield pressure exhibit different degrees of responses prone to the fault. The weighted integrated voting model can effectively predict the stable, early warning and fault section with an overall accuracy of 89.22%. The work provides supports for early warning analysis and pre-control measures near the TBM tunnel fault.

Key words: rock mechanics; TBM big data; faults intelligence perception; machine learning; intelligent tunnel construction; weighted integrated voting model

1 引言

建设引水隧洞是解决我国水资源分配不均的有效方式。大洞径、高应力、难支护、大变形是多数深埋引水隧洞的主要特点^[1]。由于 TBM 具有低扰动、开挖速度快以及连续作业等优势^[2], 现已被广泛应用于引水隧洞建设中^[3]。然而, 当 TBM 穿越断层时, 机器的性能往往显著降低, 甚至会发生工程事故, 如: 开挖面失稳、机体振幅增加^[4], 卡机、拱顶坍塌等^[5]。这不仅增加时间与工程成本, 也对工程安全构成威胁^[6]。因此, 断层超前判识有助于司机调控 TBM 控制参数, 对隧洞施工和安全保障具有重要意义。

目前, 断层超前判识主要借助超前地质预报, 具体方法主要包括地质分析法、超前钻探法和地球物理物探方法等^[7]。然而, 这些方法常需要增加额外的工序, 多用于钻爆法施工隧道, 而在 TBM 隧道中, 由于作业空间有限而存在较大局限性。通过长期的工程实践, 隧道施工过程中, 发现 TBM 传感器参数可有效反映掘进地层地应力信息和岩体物理力学性质的变化^[8]。已有研究表明, 软硬岩石的共同作用使得断层呈现出极其复杂和不均匀的特性^[9], 但在张应力与挤压力的地质作用下, 断层处及其两侧岩体的密度、孔隙度、硬度等物理性质会发生不同程度变化^[10]。因此, 寻找并建立断层附近岩体地质力学特性变化规律与施工过程 TBM 传感器参数的实时变化之间关系模型, 可望实现断层超前判识与感知。

近年来, 采用机器学习方法进行断层等地质特征智能感知预测方面, 已开展了一些有益的探索。孙振宇等^[10]通过分析构造部位和非构造部位地震属性特征, 利用 4 个地震属性数据采用支持向量机

建立了断层两分类后验识别模型。J. H. Jung 等^[11]基于 TBM 机器参数建立人工神经网络模型, 实现了隧道掘进工作面前方软硬岩三分类预测。Q. L. Zhang 等^[8]基于刀盘扭矩、推进速度、总推力和刀盘转速, 建立了随机森林、支持向量机和 K 最近邻的围岩等级预测模型。但是, 目前对于断层的智能感知研究尚不多见。同时, 一般的研究多基于单一机器学习模型开展地质特征的超前感知, 存在局限性大、准确率低等问题。加权投票模型算法可将 2 种及以上的机器学习方法融合, 合理利用单一模型的差异性和互补性, 提高模型预测精度。

为此, 本文基于吉林引松引水隧洞的 40.8 亿组现场实测数据, 系统分析了含断层掘进段的 TBM 运行数据变化规律, 利用皮尔逊相关系数确定了可反映断层信息的关键 TBM 岩机交互参数, 在此基础上, 借助支持向量机和随机森林机器学习算法, 建立加权投票模型, 实现断层的超前智能识别。

2 工程背景与数据分析

2.1 工程描述

吉林引松工程是中国“十三五”规划的 172 项重大节水供水工程之一, 是吉林省规模最大、技术难度最大的大型跨区域调水工程。本研究主要依托该工程的 TBM3 施工段, 该标段起点为饮马河分水口, 终点为岔路河, 工程始发里程为 71+476, 结束里程为 51+705, TBM 有效掘进长度为 17 488 m, 该标段采用 CREC188 开敞式 TBM 开挖, 隧洞开挖洞径 7.93 m, 埋深为 85~260 m。

TBM 施工段围岩主要包括花岗岩、凝灰质砂岩、灰岩、闪长岩、砂岩夹泥岩, 比例分别为 38.81%, 17.31%, 7.99%, 4.1%, 31.78%。上述围岩可分为四类, 分别是 II 类、III 类、IV 类、V 类, 比例分

别为 10.82%，63.12%，20.99%，5.06%。在 TBM 掘进过程中，桩号范围为 71138~70588，68709~68371，67038~66349，66349~66221 区域内共遭遇 9 个断层。断层统计数据如表 1 所示。

表 1 引水隧洞施工揭露断层统计

Table 1 Construction revealed faults of diversion tunnel			
出露桩号	填充物	出露长度/m	宽度/cm
71133~71123.9	碳质板岩及断层泥	9.1	15~60
71090.5~71079.5	断层泥及断层角砾	11.0	15~28
70995.4~71003.2	碳质板岩及断层泥	7.8	1~7
70911.3~70896	断层泥及碎裂岩	15.3	20~40
70747.5~70742	断层泥及断层角砾	5.5	15~25
70707.5~70703.2	断层泥及断层角砾	4.3	5
66322~66309	断层泥及断层角砾	13.0	10~20
66295~66284	断层泥及断层角砾	11.0	5~10
66239~66226	断层泥及断层角砾	13.0	10~20

2.2 数据特点与特征选择

该隧洞 TBM 共配备 199 个传感器通道，数据的采集频率为 1 Hz，每天单个文本数据可达 86 400，按有效掘进天数 728 d 计算，数据总量约为 40.8 亿。

隧道工程实际中，TBM 掘进数据记录一般以一次掘进循环方式划分。表 1 所示 9 个断层所在的 TBM 掘进段共有 865 个掘进循环，剔除扭矩数值为 0 的掘进循环以及塌方段的掘进循环，最终保留 771 个有效循环。其中，总掘进循环数据中，一共包含 108 个含断层的循环，约占总数据的 14%。

TBM 岩机交互参数为 TBM 与岩体交互时实时变化的参数，包括护盾压力、刀盘扭矩、总推进力等。为实现 TBM 安全平稳掘进，司机往往根据已知的地质信息和岩机交互参数的实时变化进行安全高效调整。因此，掌握 TBM 掘进过程中关键岩机交互参数的变化规律是断层超前感知的核心。

TBM 岩机交互参数中，推进压力的大小可感知围岩的强度情况；刀盘扭矩的大小可推知围岩的完整性情况^[12]；护盾压力可感知围岩收敛对护盾外表面的接触压力^[13]；前盾俯仰角、前盾滚动角能反馈掘进过程中的实时姿态；齿轮密封压力反映了施工过程中润滑油路各控制阀、流量计的指示参数。因此，初步选取刀盘扭矩、顶护盾压力、左侧护盾压力、右侧护盾压力、前盾俯仰角、前盾滚动角、齿轮密封外封压力、推进泵电机电流 8 个关键岩机交互参数，用于断层超前预测。

为降低 2 个参数之间相关性较大对数据造成重

复干扰，提高感知预测准确率，采用皮尔逊相关系数对上述 8 个参数进行两两相关性分析：

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

(1)

式中： r_{xy} 为皮尔逊相关系数，取值范围[−1，1]； $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 为 x 均值； $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ 为 y 均值； n 为样本总数。

根据式(1)进行相关性分析，结果如图 1 所示。由图 1 可知，刀盘扭矩与推进泵电机电流相关性为 0.84，2 个岩机交互参数间存在强相关，因而剔除了推进泵电机电流，仅保留 7 个岩机交互参数。

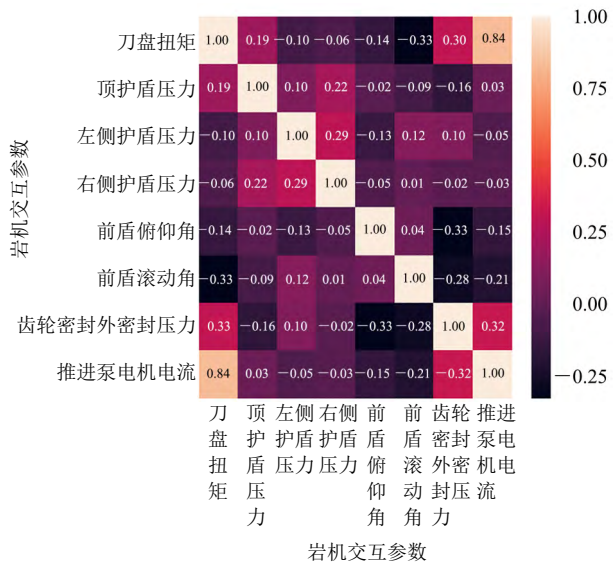


图 1 特征相关性分析
Fig.1 Correlation coefficients of features

由图 1 还可知，除刀盘扭矩与推进泵电机电流外，其余岩机交互参数之间的相关性均小于 0.4，不存在强相关。因此，最终确定刀盘扭矩、顶护盾压力、左侧护盾压力、右侧护盾压力、前盾俯仰角、前盾滚动角、齿轮密封外密封压力共 7 个参数，用于断层超前感知。

2.3 断层附近 TBM 岩机交互参数变化特征

TBM 岩机交互参数中，刀盘扭矩通常被作为判识掌子面前方岩石完整性程度的一个重要参数，而护盾压力也对地层岩性或地应力变化有所响应，因此将以刀盘扭矩(见图 2(a))及顶护盾压力(见图 2(b))为例，阐述断层附近 TBM 岩机交互参数变化规律及分段特征。

如图 2(a)所示，当 TBM 掘进到断层前时，刀

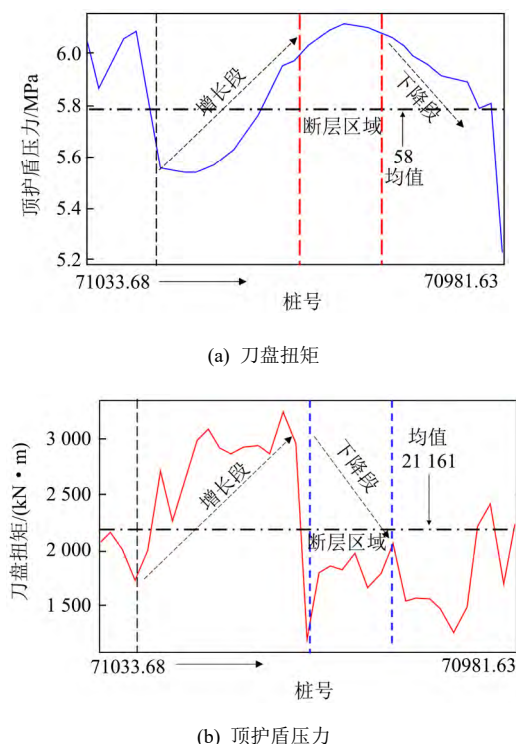


图2 某断层前岩机交互参数变化

Fig.2 Interaction parameters of rock machine change of before a fault

盘扭矩会出现大幅度的增长；掘进到断层时，刀盘扭矩会大幅下降。如图 2(b)所示，顶护盾压力同样出现了相似的规律，即 TBM 越靠近断层，顶护盾压力数值越大。其主要原因在于，断层及其两侧岩体的物理性质发生了变化；另外，随着 TBM 向前掘进，掌子面附近顶部围岩变形增大^[13]。

根据 TBM 掘进参数的数据变化进行断层超前感知(预警)区域的有效划分，是实现断层超前感知的关键前提。如图 2 所示，刀盘扭矩比顶护盾压力具有更早的前兆特征，因而可确定断层区域之前刀盘扭矩增长段为预警区域。

2.4 断层附近区段划分

根据 TBM 掘进掌子面与断层的距离，可将断层附近区段划分为平稳段、预警段和断层段(见图 3)。其中，平稳段被定义为正常掘进段，即前方

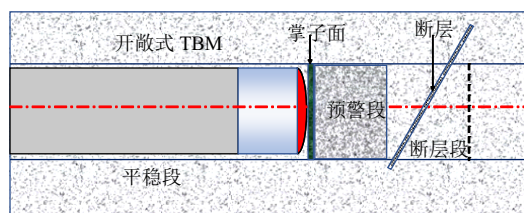


图3 断层附近区段划分及超前感知示意图

Fig.3 Advance warning diagram of TBM fault

断层影响范围之外的掘进区段，在该段掘进过程中，TBM 机器参数对断层无响应；预警段是指前方断层影响范围之内到断层所在位置之前的掘进区段，该段掘进过程中 TBM 机器参数对断层会出现一定响应；断层段是指 TBM 掘进在断层所在位置的区段。

进一步，通过统计断层附近掘进循环数发现，预警段通常出现在断层前的 118 个掘进循环范围内，因而可确定表 1 所示断层附近的平稳段、预警段、断层段 3 个区段划分界限。

3 断层超前智能感知模型

人工辨识断层的平稳段、预警段、断层段具有反应不及时、主观判断存在偶然性等不足，且在数据量大的时候，往往还存在判断困难等问题，而机器学习方法具备自动学习、处理和感知的优势。为此，有必要借助机器学习方法建立断层超前智能感知模型，利用断层附近区段 TBM 掘进的 7 个关键岩机交互参数，实现断层附近平稳段、预警段、断层段 3 个区段的智能感知。

每一种机器学习方法均具有不同适用条件、优势及不足。例如，反向传播神经网络模型具有强大的非线性拟合能力，但其泛化能力往往需要有足够的样本保障，这在实际工程中往往难以满足；随机森林方法能够较好地表述所研究对象的随机性，且能够提取出所预测对象的各种特征及其重要性排序，但其泛化能力也存在一定局限性；支持向量机模型具有良好的小样本分类能力，但在特征识别上存在局限性。为了充分利用不同分类算法之间的差异性和互补性，选择随机森林和支持向量机作为基分类器，基于加权投票策略建立断层的智能感知模型。

3.1 数据集划分与标准化

由节 2 可知，断层附近共有 771 组有效掘进循环数据，其中平稳段、预警段、断层段数据分布分别为 545, 118, 108 组。本文将 70%数据用于训练，30%数据用于测试。训练集为 539 组，用来优化感知模型的参数；测试集为 232 组，用来验证模型的泛化感知能力。

为去除 TBM 掘进数据中不同岩机交互参数的单位限制，并消除不同量纲之间数值的影响，对数据进行了归一化，从而使不同量级的 TBM 掘进数据能够进行比较与加权计算。本文采用标准分数归一化对 7 个岩机交互参数进行缩放，即均值为 0，

方差为 1, 见下式:

$$x' = \frac{x - \mu}{\sigma} \quad (2)$$

式中: x 为原始数据输入; x' 为归一化输出; μ , σ 分别为原始数据的均值和标准差。

3.2 方法简述

(1) 随机森林

随机森林是 Bagging 集成算法中最具代表性的一个算法, 具有特征重要性评估以及良好的鲁棒性, 被广泛应用于分类问题^[14]。

随机森林以决策树为基分类器, 通过随机重抽样方法从训练数据集中抽取数据组成新的训练集, 并利用随机特征选取的方法来构建多个决策树, 这部分数据约占整个训练数据的 2/3 左右, 剩余的 1/3 数据被称为袋外数据, 被用来估计每一棵决策树的误差, 从而获得所有决策树的平均估计误差^[15]。最终分类问题的目标输出是由构建的所有单个决策树投票决定。

(2) 支持向量机

支持向量机针对小样本、高维度、非线性分类问题具有良好的鲁棒性与泛化能力, 其机制是寻找一个满足分类要求的最优分类超平面, 在保证预测目标分类精度的同时, 使平面两侧距平面最近的两类样本之间的距离最大化^[16]。

以两分类数据为例, 训练样本集为 (x_i, y_i) , $i = 1, 2, \dots, n$, $x \in R^n$, $y \in \{+1\}$, 线性判别函数 $g(x) = w \cdot x + b$, 可利用超平面:

$$w \cdot x + b = 0 \quad (3)$$

式中: w 为权向量, b 为分类阈值。为使所有样本被正确分类, 需满足如下约束条件:

$$y_i(w \cdot x + b) \geq 1 \quad (i = 1, 2, \dots, n) \quad (4)$$

当满足如上约束条件后可获得最优分类面, 其中分类间隔为 $2/\|w\|$ 。

因此, 最优分类面问题可被表示在如下约束条件下:

$$\min \varphi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w) \quad (5)$$

最优分类函数可以被表示为

$$f(x) = \text{sgn}(w \cdot x + b) = \text{sgn}\left(\sum_{i=1}^n l_i y_i (x_i \cdot x) + b\right) \quad (x \in R^n) \quad (6)$$

式中: l_i 为拉格朗日因子。

松弛变量可以被用于约束条件下线性不可分的情况, 通过在目标函数中引入惩罚函数来进一步求解最优分类超平面问题, 公式如下:

$$\min \varphi(w, \zeta) = \frac{1}{2} (w \cdot w) + C \left(\sum_{i=1}^n \zeta_i \right) \quad (7)$$

式中: ζ 为松弛变量; C 为惩罚因子, 控制着模型的泛化能力。

如果对于线性不可分的问题, 支持向量机常引入核函数 $H(x_i, x)$, 将输入映射到一个高维的特征向量空间, 判别函数如下:

$$f(x) = \text{sgn}\left(\sum_{i=1}^n l_i y_i H(x_i, x) + b\right) \quad (8)$$

径向基核函数在非线性分类中表现出了较好的分类性能, 公式如下所示:

$$H(x_i, x_j) = \exp(-g \|x_i - x_j\|^2) \quad (g > 0) \quad (9)$$

式中: g 为影响径向基核函数在分类中的效果的参数。

由此不难看出, 惩罚参数 C 与核函数参数 g 是影响支持向量机分类性能的关键参数。

3.3 超参数优化

最优超参数组合下的模型能显著提高预测精度。因此有必要优化机器学习模型的超参数。网格搜索算法被广泛应用于机器学习模型的超参数优化。

网格搜索算法又称为“穷举法”, 通过将超参数排列组合, 并将各个超参数按照一定的间距形成网格中的点, 然后搜索网格中的每一个交点, 找出对应目标函数最优值, 从而得到最优超参数组合^[17]。该方法较好地避免了由于超参数之间耦合导致的多解性问题, 防止陷入局部最优值。

为获取模型的最优超参数, 通过 10 折交叉验证与网格搜索法对随机森林、支持向量机进行超参数优化。测试集的预测结果为 3 阶混淆方阵, 记 $3M$:

$$3M = \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{bmatrix} \quad (10)$$

式中: n_{11} , n_{22} , n_{33} 分别为某一分类器正确识别平稳段、预警段、断层段类别的数量; n_{12} , n_{13} 表示将平稳段误判为预警段、断层段; n_{21} , n_{23} 表示将预警段误判为平稳段、断层段, n_{31} , n_{32} 表示将断层段误判为平稳段、预警段。

并用下式计算准确率 Auc 以评估模型性能:

$$Auc = \frac{n_{11} + n_{22} + n_{33}}{N_{total}} \quad (11)$$

式中: N_{total} 为总类别数。

同时, 以召回率 R 和 F_1 分数评估模型在测试集中的预测性能, 精确率、召回率、 F_1 分数计算如下所示:

$$P = \frac{n_{11}}{n_{11} + n_{21} + n_{31}} \quad (12)$$

$$R = \frac{n_{11}}{n_{11} + n_{12} + n_{13}} \quad (13)$$

$$F_1 = \frac{2PR}{P + R} \quad (14)$$

其中, F_1 分数是模型精确率 P 和召回率的加权平均。

对随机森林算法的决策树数量、决策树最大深度和叶节点上的最小样本、节点可分的最小样本数 4 个超参数进行优化; 支持向量机中采用了径向基核函数, 对惩罚参数与核函数参数进行优化。各超参数取值范围和所得的最优参数如表 2 所示。

表 2 超参数搜索范围和最优值

Table 2 Hyperparameters search range and optimal values

分类器	参数	取值范围	取值间距	最优值
随机森林	决策树数量	100~700	50	275
	决策树最大深度	1~20	1	19
	叶节点上的最小样本	1~5	1	1
	节点可分的最小样本数	2~5	1	2
支持向量机	核函数	径向基	-	径向基
	惩罚系数 C	1~20	1	15
	核函数参数 g	0.1~1	0.1	0.5

3.4 加权集成投票模型

确定随机森林、支持向量机的超参数优化后, 将二者作为基分类器, 构建加权集成投票分类模型。

加权集成投票分类模型中, 每个基分类器分配不同的权值, 其中随机森林、支持向量机的权重分别为 w_1 , w_2 ; 然后, 权值乘上每个基分类器预测三个类别的概率值, 从而形成一个矩阵, 记为 CM ; 最终, 平均概率最高的类别作为最终输出结果, 结果记为 res 。 CM 及 res 计算公式分别如下:

$$CM = [w_1 \quad w_2] \begin{bmatrix} r_1 & r_2 & r_3 \\ s_1 & s_2 & s_3 \end{bmatrix} \quad (15)$$

$$res = \max \begin{bmatrix} w_1 r_1 + w_2 s_1 / w_1 + w_2 \\ w_1 r_2 + w_2 s_2 / w_1 + w_2 \\ w_1 r_3 + w_2 s_3 / w_1 + w_2 \end{bmatrix} \quad (16)$$

式中: r_1 , r_2 , r_3 分别为随机森林中三类别预测概率; s_1 , s_2 , s_3 分别为支持向量机中三类别预测概率。其中 $r_1 + r_2 + r_3 = 1$, $s_1 + s_2 + s_3 = 1$ 。

本文所建立的断层智能感知模型实施流程如图 4 所示。通过对 2 个基分类器权值进行优化, 确定 w_1 , w_2 的权值分别为 0.1, 0.9。

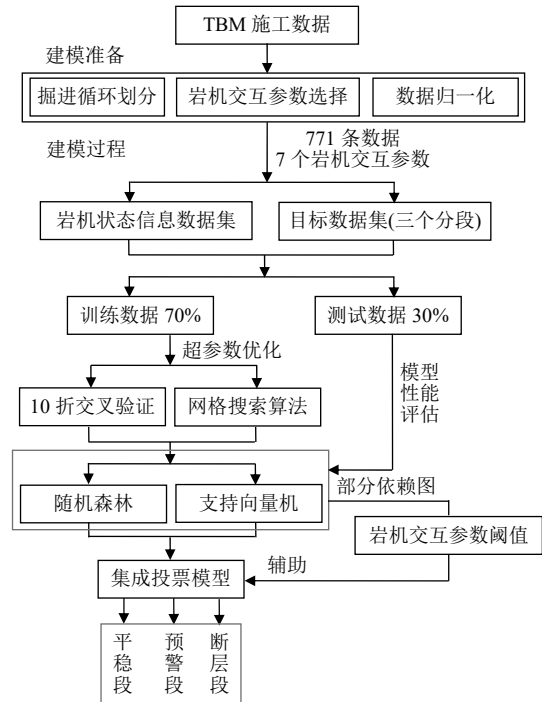


图 4 断层智能感知流程

Fig.4 Framework of fault perception

4 结果与讨论

4.1 断层感知结果

最优超参数下, 随机森林、支持向量机和集成投票感知模型在训练集中的预测结果如图 5 所示。图 5 中, 数字含义与 3 阶混淆方阵中各元素含义相同。由图 5 可知, 超参数最优的 3 个模型成功建立了 7 个 TBM 岩机交互参数与断层之间的非线性关系。

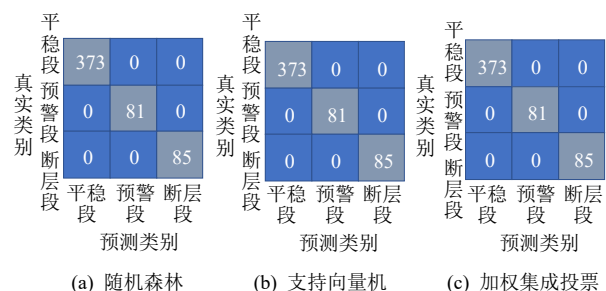


图 5 不同模型在训练集上的感知预测结果矩阵

Fig.5 Perceptual prediction result matrix of different models on training set

为检验 3 个模型在平稳段、预警段、断层段的分类性能以及泛化能力，采用测试集对 3 个模型进行测试。表 3 分别统计了 3 个模型在平稳段、预警段、断层段中的召回率和 F_1 。

表 3 不同感知模型的性能指标比较
Table 3 Comparison of performance indicators of different perception models

分类器	类别	召回率	F_1 分数
随机森林	平稳段	0.912 8	0.894 6
	预警段	0.567 6	0.666 7
	断层段	0.695 7	0.640 0
支持向量机	平稳段	0.924 4	0.938 1
	预警段	0.756 8	0.767 1
	断层段	0.826 1	0.730 8
加权集成投票	平稳段	0.930 2	0.938 4
	预警段	0.756 8	0.777 8
	断层段	0.826 1	0.745 1

由表 3 可知，3 个模型中，平稳段的预测召回率、 F_1 分数最高，表明了平稳段相较于其他 2 个类别，有较好的区分性。预警段中召回率最高仅为 75.68%， F_1 分数最高为 77.78%，其原因可能是平稳段与预警段存在数据上的交叉。断层段中召回率最高为 82.61%， F_1 分数最高为 74.51%，预测结果较差的原因很大程度是由于在不同断层下，其宽度、产状、充填物对 TBM 施工影响程度不同，造成部分断层中的岩机交互参数存在较小的区分性。同时，断层数据集中三类别的不均衡，也是造成预警段与断层段的预测结果较差的一个原因。

对比 3 个模型在测试集中的预测结果可知，在统计学理论和结构风险最小原理基础上建立的支持向量机，比随机森林在断层地超前感知中具有更好的稳定性。加权集成投票模型充分利用了基分类器之间的差异性和互补性，其预测性能优于基分类器。3 个模型在测试集中的预测结果如图 6 所示。

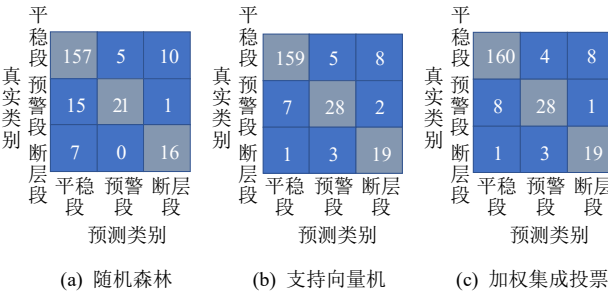


图 6 不同模型在测试集上的感知预测结果矩阵

Fig.6 Perceptual prediction result matrix of different models on test set

采用 5 折交叉验证方法以及宏平均和准确率，对加权集成投票模型的预测性能以及在训练集中的稳定性进行了进一步检验，宏平均和准确率的关系式如下：

$$m - p = \frac{1}{n} \sum_{i=1}^n P_i \tag{17}$$

式中： $m - p$ 为宏平均， P 为精确率。

检验结果如表 4 所示，由表 4 可知，加权集成投票模型的宏平均和准确率均值分别为 0.811 3 和 0.883 1，高于随机森林和支持向量机模型的预测结果。结果表明，加权集成投票模型的性能要优于单一的随机森林和支持向量机模型，也说明随机划分数据集对加权集成投票模型的影响小于单一的随机森林和支持向量机模型。

表 4 不同感知模型总体评价结果
Table 4 Overall evaluation results of all classifiers

模型	宏平均	准确率
随机森林	0.753 1	0.855 2
支持向量机	0.802 8	0.877 5
加权集成投票	0.811 3	0.883 1

综合表 3、4 的结果可知，加权集成投票模型的性能优于随机森林和支持向量机模型，说明加权集成投票模型更适合断层超前预测。为此，下文重点分析 7 个岩机交互参数对加权集成投票模型性能的影响。

4.2 讨论

所建立的断层超前预测模型，优选了 7 个岩机交互参数，分别是齿轮密封外密封压力、左侧护盾压力、右侧护盾压力、顶护盾压力、刀盘扭矩、前盾俯仰角、前盾滚动角。实际上，这 7 个岩机交互参数对断层的感知贡献有所差异。为探究选用的岩机交互参数在预测断层中的重要性，分析其在模型中的敏感度。敏感度算法流程可分为如下 4 个步骤^[18]：

- (1) 将断层数据集划分为训练集 T_n 和 T_t ，利用 2 个基分类器(随机森林、支持向量机)在训练集 T_n 上对其进行学习，获得 2 个分类器 f_{rf} ， f_{svm} ，并通过加权组合形成最终集成投票模型 F ；
- (2) 通过加权集成投票模型对测试集 T_t 进行预测，获得准确率 S ；
- (3) 对测试集 T_t 某一列岩机交互参数的数据进行 5 次随机排列，其他 6 个参数的数据排列保持不动，利用 F 分别重新预测排列后的测试集 $T'_i(i =$

1, 2, ..., 5), 获得 5 个准确率 S_i , 并得出 5 次 $S - S_i$ 的差值范围, 以此作为一个岩机交互参数敏感度;

(4) 将步骤(3)中打乱的数据复原, 换下一个岩机交互参数并重复步骤(3), 直到所有的参数的敏感度都计算完成。最终对 7 个岩机交互参数的敏感进行排名。

本文分析了所述 7 个岩机交互参数在加权集成投票模型预测断层时的敏感度, 结果如图 7 所示。由图 7 可知, 7 个岩机交互参数的敏感度均高于 0.06, 说明这些岩机交互参数能够反映平稳段、预警段、断层段的特征, 可作为断层超前感知的关键参数。

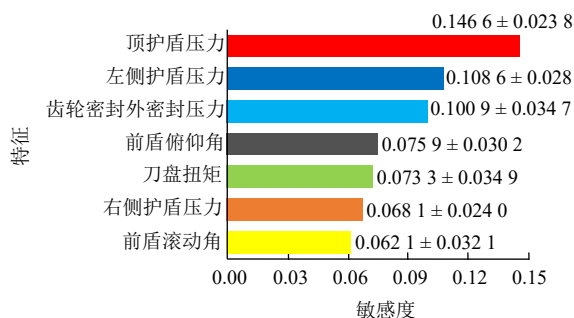


图 7 加权集成投票模型特征敏感度排名

Fig.7 Features sensitivity ranking of RF-SVM model

同时, 分析被动岩机参数在平稳段、预警段、断层段 3 个类别下的预测概率是至关重要的, 这有助于发现岩机交互参数在模型中的变化规律。部分依赖图^[18]显示了岩机交互参数对先前拟合模型预测结果的边际效应, 可很好地表示岩机交互参数在 3 个预测类别中的相关关系。同时, 顶护盾压力、左侧护盾压力、齿轮密封外密封压力、前盾俯仰角、刀盘扭矩 5 个岩机交互参数除了具有较高的敏感度外, 还在 3 个类别中具有较明显的响应规律, 如图 8 所示。

由图 8(a)可知, 当顶护盾压力达到 4 MPa 后, 在预警段与断层段的预测概率增大; 由图 8(b)可知, 左侧护盾压力达到 3 MPa 后, 在预警段与断层段的预测概率增大; 由图 8(c)可知, 齿轮密封外密封压力数值区间范围在 0~0.002 5 MPa 时, 其在预警段与断层段的预测概率较大, >0.002 5 MPa 时, 在平稳段的预测概率显著提高; 由图 8(d)可知, 前盾俯仰角数值大于 0.35° 时, 其在平稳段的预测概率显著增加; 由图 8(e)可知, 随着刀盘扭矩数值的增大, 特别是当刀盘扭矩增长到超过 2 500 kN·m 后, 在预警段的预测概率增大。因此, 顶护盾压力、左侧护盾压力、齿轮密封外密封压力等 5 个关键交互参

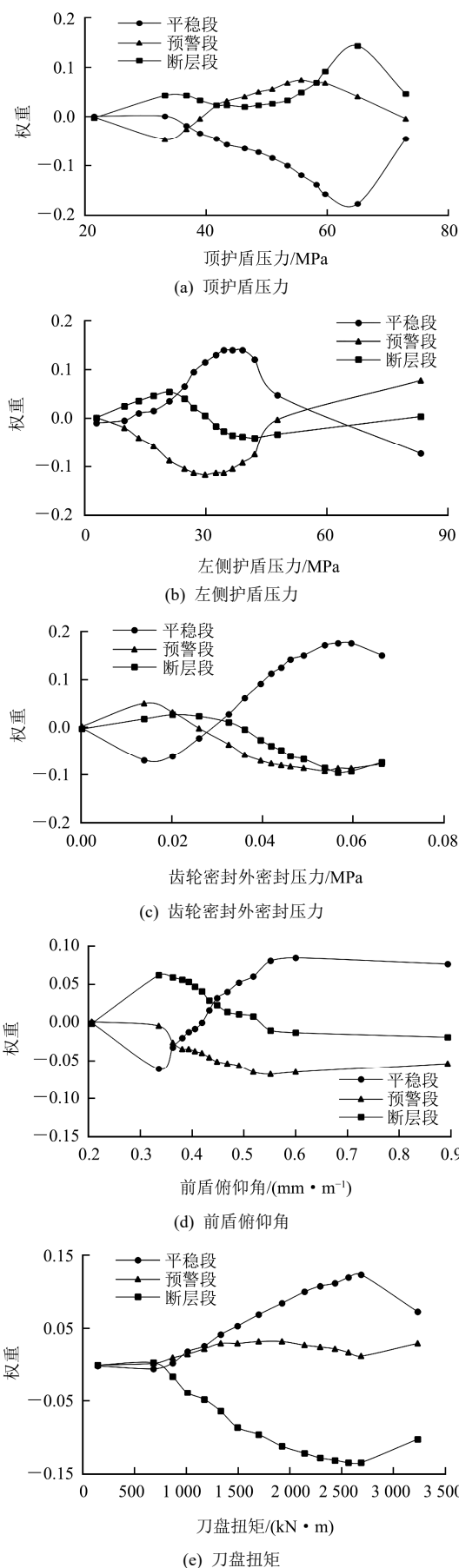


图 8 不同岩机交互参数对加权集成投票模型预测结果的影响
Fig.8 The prediction probability of five interaction parameters of rock machine under RF-SVM model

数的数值区间, 可作为断层超前感知的依据。

5 结 论

基于断层附近岩体物理力学性质会发生变化从而导致 TBM 岩机交互参数发生响应, 建立了以随机森林和支持向量机为基分类器的断层超前感知加权投票模型, 并对断层附近区段开展了感知预测和评价, 得到如下主要结论:

(1) TBM 岩机交互参数在断层附近出现了不同程度响应, 如刀盘扭矩和顶护盾压力等在临近断层前数值升高, 而在断层处刀盘扭矩等参数则出现快速下降, 因此可以将断层区段划分为平稳段、预警段和断层段。

(2) 基于加权集成投票方法建立的断层超前感知模型, 可以有效实现断层的平稳段、预警段和断层段三阶段超前感知, 在测试集中的预测准确率达到 89.22%, 其模型性能比随机森林和支持向量机等单一模型更好。

(3) 在加权集成投票模型中, 不同岩机交互参数具有不同的贡献权重, 贡献权重大小依次为顶护盾压力、左侧护盾压力、齿轮密封外密封压力、前盾俯仰角、刀盘扭矩、右侧护盾压力、前盾滚动角。

(4) 通过部分依赖图得到了岩机交互参数在平稳段、预警段和断层段中的相关性, 顶护盾压力、左侧护盾压力与刀盘扭矩等 5 个关键参数的数值区间范围也可辅助作为断层超前感知的依据。

致谢 特别致谢 TBM 掘进数据提供方——973 项目(2015CB058100)“TBM 安全高效掘进全过程信息化智能控制与支撑软件基础研究”提供数据。笔者在本文研究过程中使用了部分开源网络资源 <https://scikit-learn.org>, eli5 库和 pdpbox 库, 在此一并表示感谢!

参考文献(References):

- [1] 周建军, 杨振兴. 深埋长隧道 TBM 施工关键问题探讨[J]. 岩土力学, 2014, 35(增2): 299 - 305.(ZHOU Jianjun, YANG Zhenxing. Discussion on key issues of TBM construction in deep buried tunnel[J]. Rock and Soil Mechanics, 2014, 35(Supp.2): 299 - 305.(in Chinese))
- [2] 王梦恕. 中国盾构和掘进机隧道技术现状、存在的问题及发展思路[J]. 隧道建设, 2014, 34(3): 179 - 187.(WANG Mengshu. Current status, existing problems and development ideas of Chinese shield and tunneling machine tunnel technology[J]. Tunnel Construction, 2014, 34(3): 179 - 187.(in Chinese))
- [3] 张镜剑, 傅冰骏. 隧道掘进机在我国应用的进展[J]. 岩石力学与工程学报, 2007, 26(2): 226 - 238.(ZHANG Jingjian, FU Bingjun. Advances in the application of tunnel boring machine in China[J]. Chinese Journal of Rock Mechanics and Engineering, 2007, 26(2):

- 226 - 238.(in Chinese))
- [4] RAMONI M, ANAGNOSTOU G. Tunnel boring machines under squeezing conditions[J]. Tunnelling and Underground Space Technology, 2010, 25(2): 139 - 157.
- [5] GONG Q M, YIN L J, MA H S, et al. TBM tunnelling under adverse geological conditions: an overview[J]. Tunnelling and Underground Space Technology, 2016, 57: 4 - 17.
- [6] LI S C, NIE L C, LIU B. The practice of forward prospecting of adverse geology applied to hard rock TBM tunnel construction: The case of the songhua river water conveyance project in the middle of Jilin Province[J]. Engineering, 2018, 4(1): 131 - 137.
- [7] 李术才, 刘 斌, 孙怀凤, 等. 隧道施工超前地质预报研究现状及发展趋势[J]. 岩石力学与工程学报, 2014, 33(6): 1 090 - 1 113.(LI Shucai, LIU Bin, SUN Huaifeng, et al. State of art and trends of advanced geological prediction in tunnel construction[J]. Chinese Journal of Rock Mechanics and Engineering, 2014, 33(6): 1 090 - 1 113.(in Chinese))
- [8] ZHANG Q L, LIU Z Y, TAN J R. Prediction of geological conditions for a tunnel boring machine using big operational data[J]. Automation in Construction, 2019, 100: 73 - 83.
- [9] PALTRINIERI E, SANDRONE F, DUDT J P, et al. Probabilistic simulations of TBM tunnelling in highly fractured and faulted rocks[J]. Tunnelling and Underground Space Technology, 2016, 57: 183 - 194.
- [10] 孙振宇, 彭苏萍, 邹冠贵. 基于 SVM 算法的地震小断层自动识别[J]. 煤炭学报, 2017, 42(11): 2 945 - 2 952.(SUN Zhenyu, PENG Suping, ZOU Guangui. Automatic identification of small faults based on SVM algorithm[J]. Journal Of China Coal Society, 2017, 42(11): 2 945 - 2 952.(in Chinese))
- [11] JUNG J H, CHUNG H, KWON Y S, et al. An ANN to predict ground condition ahead of tunnel face using TBM operational data[J]. KSCE Journal of Civil Engineering, 2019, 23(7): 3 200 - 3 206.
- [12] 王梦恕, 王占山. TBM 通过断层破碎带的施工技术[J]. 隧道建设, 2001, (3): 1 - 4.(WANG Mengshu, WANG Zhanshan. Construction technology of TBM through fault fracture zone[J]. Tunnel Construction, 2001, (3): 1 - 4.(in Chinese))
- [13] 程建龙, 杨圣奇, 李学华, 等. 位移释放率对双护盾 TBM 护盾压力的影响研究[J]. 岩土力学, 2016, 37(5): 1 399 - 1 407.(CHENG Jianlong, YANG Shengqi, LI Xuehua, et al. Study on the influence of displacement release rate on the pressure of double shield TBM shield[J]. Rock and Soil Mechanics, 2016, 37(5): 1 399 - 1 407.(in Chinese))
- [14] SVETNIK V, LIAW A, TONG C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling[J]. Journal of Chemical Information and Computer Sciences, 2003, 43(6): 1 947 - 1 958.
- [15] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5 - 32.
- [16] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273 - 297.
- [17] 李清毅, 周 昊, 林阿平, 等. 基于网格搜索和支持向量机的灰熔点预测[J]. 浙江大学学报: 工学版, 2011, 45(12): 2 181 - 2 187.(LI Qingyi, ZHOU Hao, LIN Aping, et al. Prediction of ash fusion temperature based on grid search and support vector machine[J]. Journal of Zhejiang University: Engineering Science, 2011, 45(12): 2 181 - 2 187.(in Chinese))
- [18] ALTMANN A, TOLOSI L, SANDER O, et al. Permutation importance: a corrected feature importance measure[J]. Bioinformatics, 2010, 26(10): 1 340 - 1 347.
- [19] FRIEDMAN J, HASTIE T, TIBSHIRANI R. The elements of statistical learning[M]. New York: Springer Series In Statistics New York, 2001: 299 - 346.