

基于机器学习的岩土参数智能预测方法研究

孙佳锐

天津市北洋水运水利勘察设计院有限公司 天津 300460

摘要：岩土参数是岩土工程勘察、设计、施工的基础数据，传统室内试验与原位测试周期长、成本高，且受采样条件限制，难以满足复杂地质快速测算需求。为提高岩土参数预测精度与效率，本文基于机器学习开展智能预测研究。首先梳理岩土参数相关理论，完成数据采集、清洗、特征优化等预处理工作，构建人工神经网络、支持向量回归、随机森林与深度学习预测模型，并进行超参数调试。通过对比各模型预测性能，总结算法优缺点及适用地质条件。研究证明机器学习可有效拟合岩土参数非线性关系，弱化试验条件限制，提升预测稳定性，可为岩土工程智能勘察与地质风险分析提供参考。

关键词：机器学习；岩土参数；智能预测；深度学习

引言：岩土参数精准测定是工程安全施工与优化设计的重要前提，天然岩土体存在空间变异性与非均质性，参数变化规律复杂。当前岩土参数多依靠土工试验与原位勘测获取，操作流程繁琐、检测成本较高，恶劣地质条件下采样难度大，实测数据离散性明显。机器学习具备非线性拟合与数据挖掘优势，现已广泛应用于岩土工程数据分析。本文构建多类机器学习预测模型，优化数据处理流程与模型参数，对比不同算法预测效果，旨在探寻高效、高精度的岩土参数预测方法，解决传统检测方式存在的痛点，推动岩土工程勘测向智能化、数字化方向转型。

1 岩土参数预测的理论基础

1.1 岩土工程常用参数体系

岩土工程参数体系涵盖物理、力学、水理三类核心参数，是描述岩土体基本性质、判定工程稳定性的重要指标。物理参数包含含水率、孔隙比、颗粒密度、液塑限等基础指标，用于直观反映岩土物质组成与结构形态；力学参数以黏聚力、内摩擦角、压缩模量、抗压强度为主，是地基承载力测算、边坡稳定性分析、围岩支护设计的核心依据；水理参数包含渗透系数、毛细上升高度等，多用于地下水作用下岩土形变与渗漏风险研判。不同工程场景参数选取存在差异，土质工程侧重压缩、抗剪参数，岩质工程聚焦强度与裂隙参数。该体系参数相互制约、关联紧密，构成岩土分析基础数据库，为机器学习模型特征筛选、样本训练提供完整指标支撑。

1.2 参数间的内在关联性分析

天然岩土体受沉积环境、地质构造、风化作用影响，各项参数存在复杂非线性关联，并非独立变化。物理参数为基础关联指标，土体含水率升高会降低密度，直接弱化岩土抗剪强度，孔隙比变化同样会改变压缩模量与渗透性能；力学参数之间存在协同变化规律，坚硬岩层抗压强度越高，黏聚力与内摩擦角普遍呈现正向增长趋势；水理参数与物理、力学参数深度耦合，地下水渗透会软化岩土结构，造成强度衰减、形变加剧。传统数理统计方法难以精准拟合复杂关联关系，无法剔除参数冗余干扰。明确参数内在耦合机理，能够优化模型特征组合，减少无效特征输入，提升机器学习模型训练收敛速度与预测精度^[1]。

1.3 机器学习预测的基本原理与适用条件

机器学习依托计算机算法挖掘数据内在规律，通过样本训练构建输入特征与目标参数的映射关系，完成未知数据智能预测。其核心原理为依托海量实测样本，借助算法迭代优化权重参数，不断修正模型误差，拟合多变量非线性函数，规避传统线性拟合的局限性。该技术包含监督学习、无监督学习等类别，岩土参数预测多采用监督学习算法，以实测岩土指标为训练样本，完成定向参数预测。适用条件包含三类，其一需具备足量、高质量实测样本，保障模型训练可靠性；其二适配参数耦合关系复杂、物理机理难以明确的岩土测算场景；其三适用于对预测速度要求高、常规试验难以开展的复杂地质区域。地质极端突变、样本严重缺失场景下，单一机器学习模型预测效果存在局限性。

2 岩土参数智能预测的数据采集与预处理

2.1 数据来源与样本构建

本次研究数据来源于岩土工程勘察报告、室内土工试验、原位现场测试,覆盖软黏土、砂土、硬质岩层等多类常见岩土体,采集指标包含物理、力学、水理多项基础参数。为保障样本全面性,扩充不同埋深、不同地质区域、不同风化程度岩土数据,剔除人工刻意筛选的同质化样本,贴合实际工程复杂地质环境。样本构建遵循科学性原则,明确输入特征与预测标签,将易检测的含水率、密度、颗粒级配等指标设为输入特征,将难以实测的强度、模量参数设为预测标签。同时把控样本容量,合理扩充数据体量,规避小样本造成的模型过拟合问题,为后续模型训练、验证、测试搭建完整、多元、真实的岩土样本数据集。

2.2 数据清洗与异常值处理

原始岩土数据受检测仪器误差、人工操作失误、地质突发异变影响,存在缺失值、异常值,直接使用会降低模型预测精度。数据清洗阶段优先筛选完整有效样本,采用插值法补充微量缺失数据,大面积缺失的无效样本直接剔除,规避数据残缺造成的训练偏差。异常值处理结合数理统计与工程经验,依托箱线图判定离散异常数据,区分测量误差异常值与地质特殊突变数据,剔除人为误差产生的错误样本,保留特殊地质真实极值数据^[2]。完成清洗后核查数据分布状态,保障样本数据贴合正态分布规律,弱化极端数据对模型训练的干扰。纯净的数据集能够简化算法拟合难度,提升模型收敛稳定性,为特征工程开展奠定数据基础。

2.3 特征工程

特征工程是优化机器学习模型性能的核心环节,主要包含特征筛选、特征变换、特征衍生三项流程,用于剔除冗余干扰指标,强化有效特征关联度。结合岩土参数内在关联性,采用相关性分析算法筛选特征,剔除高共线性、低关联度的冗余指标,保留对预测目标影响显著的核心特征;通过对数变换、归一化变换优化原始数据分布,弱化量纲差异带来的运算偏差;依托基础参数衍生复合型特征,构建岩土密实程度、风化程度衍生指标,丰富特征维度。合理的特征工程能够简化模型结构,降低算法运算负荷,避免特征冗余引发的过拟合问题。贴合岩土机理的特征优化处理,可精准挖掘数据隐藏规律,大幅提升后续预测模型的拟合精度。

2.4 数据集划分与标准化

为客观评判模型泛化能力,将预处理后的岩土数据集按比例划分为训练集、验证集与测试集,划分过程采

用随机打乱方式,规避数据排序规律对训练结果的干扰。训练集用于模型算法学习、权重迭代优化,验证集用于调试超参数、修正模型结构,测试集用于模拟工程实测场景,客观评定模型预测性能。各类数据集保持特征分布一致性,保障样本多元性。同时开展数据标准化处理,采用极值标准化方法统一指标量纲,将不同量级岩土参数转化至同等数值区间,消除物理、力学参数数值跨度差异带来的算法运算偏差。标准化处理可加快模型梯度下降收敛速度,提升训练效率,适配多类机器学习算法运算要求。

3 基于机器学习的岩土参数预测模型构建

3.1 预测任务定义

结合岩土工程实际应用需求,本次研究定义岩土参数回归预测任务,以常规易检测岩土指标为输入变量,精准预测岩土力学核心参数。任务明确以黏聚力、内摩擦角、压缩模量为主要预测目标,适配地基勘测、边坡防护、隧道施工主流工程测算需求。任务设定遵循实用性原则,兼顾不同岩土体适配性,覆盖土质、岩质两类地质样本。同时明确模型评价指标,选用决定系数、平均绝对误差、均方误差判定预测效果,决定系数越趋近于1、误差数值越小,代表模型拟合精度越高。任务划分训练、验证、测试全流程,严控数据泄露问题,保障模型训练严谨性。清晰的任务定义能够规范模型搭建流程,为多类算法对比分析、性能优化提供统一判定标准^[3]。

3.2 基于 ANN 的岩土参数预测模型

人工神经网络(ANN)是模拟人脑神经元信息传递机制构建的智能算法,凭借优秀的非线性映射与自适应拟合能力,能够有效处理岩土参数之间复杂、隐蔽的耦合关系,十分适配岩土工程多因素、非线性的参数预测场景。本文构建经典三层前馈神经网络结构,整体包含输入层、单隐藏层与输出层,结构简洁且贴合岩土样本的数据特征。其中输入层导入经过特征筛选、标准化处理后的岩土基础指标;隐藏层引入非线性激活函数,突破线性运算局限,实现高维特征变换与数据挖掘;输出层直接输出黏聚力、压缩模量等目标力学参数。模型采用反向传播算法不断迭代更新网络权重与阈值,以最小化损失函数为优化目标,持续缩小预测数值与真实试验数值之间的误差。模型训练阶段对学习率、最大迭代次数、隐藏层神经元数量等关键超参数进行调试,依托验证集反复优化结构,有效规避过拟合、欠拟合等不良训练现象。该模型无需明确物理机理公式,自适应学习能力突出,但其迭代运算耗时偏长,在小样本数据集上

易出现泛化偏差,更适用于拥有海量勘测数据的工程项目,完成高精度连续岩土参数预测。

3.3 基于 SVR 的岩土参数预测模型

支持向量回归 (SVR) 以间隔最大化原理为核心,通过核函数将低维非线性数据映射至高维空间,完成线性回归拟合,适配中等容量岩土样本预测。模型选用径向基核函数处理岩土数据,精准捕捉参数隐性关联规律,设置惩罚系数、核函数带宽管控模型复杂度,平衡拟合精度与泛化能力。相较于神经网络,SVR 模型结构简单、运算速度快,具备优秀的抗干扰能力,可弱化岩土异常数据的负面影响。在岩土参数预测中,该模型对离散度偏高的地质样本适配性更强,稳定性突出,但其复杂高维特征拟合能力有限,多参数耦合关联极强场景下预测精度略低于神经网络,多用于中小型工程简易参数快速测算。

3.4 基于随机森林的岩土参数预测模型

随机森林属于集成学习算法,由多棵决策树并行构成,依托投票机制整合预测结果,具备抗过拟合、容错性强的优势。模型采用 bootstrap 抽样方式随机选取训练样本与特征指标,构建差异化决策树,规避单一决策树算法局限性。训练过程中调控决策树数量、树深度、分支节点数量,平衡运算效率与预测精度。该算法可输出特征重要性排序,直观判定含水率、孔隙比等指标对岩土力学参数的影响权重,兼具预测与数据分析功能。随机森林对岩土不均匀样本适配性良好,能够处理离散与连续混合数据,缺陷为极端数值预测灵敏度不足,适用于常规地质条件下岩土参数批量预测分析。

3.5 基于深度学习的岩土参数预测模型

深度学习依托多层神经网络结构深化特征挖掘能力,相较于传统机器学习,可自动提取岩土深层隐性特征,适配复杂地质、大样本数据集预测场景。本次搭建浅层深度神经网络,增设批量归一化层优化数据分布,搭配 dropout 层抑制过拟合问题,提升模型泛化性能。模型通过多层非线性变换拆解参数耦合关系,精准捕捉岩土体隐蔽变化规律,适配高维、海量岩土勘测数据。深度学习模型预测精度高、扩展性强,可适配多地域、多类型岩土体综合预测,但其结构复杂、超参数数量多,训练硬件要求高,小样本地质数据场景下易出现过拟合现象,更适配大型岩土工程海量数据智能化测算^[4]。

3.6 模型超参数优化

超参数直接决定机器学习模型结构与运算逻辑,为提升各类模型预测性能,采用网格搜索结合交叉验证方式完成参数优化。针对 ANN 模型调试隐藏层神经元数量、学习率、迭代次数;优化 SVR 模型核函数参数、惩罚系数;调整随机森林决策树数目、最大深度;修正深度学习网络层数、正则化系数。优化过程以验证集误差为判定依据,锁定误差最低组合作为最优超参数,同时严控模型复杂度,避免过度优化引发的过拟合问题。优化结果显示,调参后各类模型误差明显降低,泛化能力显著提升。科学的超参数优化能够适配岩土数据特性,充分发挥不同算法优势,为后续模型对比、工程应用提供最优模型样本。

结束语

本文系统化开展机器学习岩土参数智能预测研究,梳理岩土参数基础理论,完成数据采集、清洗、特征优化等预处理工作,搭建四类主流机器学习预测模型,并实施超参数优化调试,全面分析各类算法适配特性与预测优劣。机器学习可精准拟合岩土参数非线性关联,突破传统试验检测局限,其中深度学习拟合精度最优,随机森林稳定性突出,SVR 适配中小样本,ANN 综合性能均衡。为进一步推动技术落地,后续可融合多源勘测数据优化样本体系,引入混合算法改良单一模型缺陷,搭建岩土参数智能化预测平台。依托人工智能赋能岩土勘测行业,能够降低试验成本、提升测算效率,为地质勘察、工程设计、灾害防控提供可靠技术支撑。

参考文献

- [1] 黄涵,赵宇宸,武颖利,等. 基于小样本机器学习的土石坝堆石料流变参数反演方法 [J]. 水电能源科学,2026,44(3):99-104.
- [2] 刘学伟,刘云豪,刘滨,等. 考虑分区劣化的围岩力学参数精细化反演机器学习模型研究 [J]. 岩土工程学报,2025,47(11):2305-2315.
- [3] 汪姚文,杨伟峰,郑鑫源,等. 基于机器学习的宁波淤泥质黏土参数取值优化模型 [J]. 工程地质学报,2025,33(1):38-46.
- [4] 徐全善,祝毓萍,沈正. 基于 CPTU 测试岩土参数机器学习应用研究 [J]. 电脑知识与技术,2023,19(33):94-99,102.