

机器学习在材料服役性能预测中的应用

王红珂¹, 刘啸天¹, 林磊¹, 孙海涛², 吕云鹤², 张晏玮¹, 薛飞¹

(1.苏州热工研究院有限公司, 江苏 苏州 215004; 2.生态环境部核与辐射安全中心, 北京 100082)

摘要: 针对材料服役性能预测存在误差大、计算复杂、适用性差等问题, 提出了基于数据挖掘的机器学习预测方法。首先阐述了机器学习的应用流程, 并总结了常用模型原理及其在材料性能预测中的应用。然后采用多种机器学习模型对 RPV 钢的辐照性能进行预测, 并通过 Stacking 集成方法提高了模型的预测精度。结果表明, 机器学习可用于材料服役性能预测, 具有较高的预测精度和可靠性。根据材料服役数据的不同特征选择合适的学习模型, 同时进行模型融合和参数优化, 可有效提高模型的预测精度及运算速度。

关键词: 数据挖掘; 机器学习; 服役性能; 材料工程; 模型预测

中图分类号: TG172

文献标识码: A

文章编号: 1672-9242(2022)01-0011-09

DOI: 10.7643/issn.1672-9242.2022.01.002

Application of Machine Learning in Predicting Service Performance of Materials

WANG Hong-ke¹, LIU Xiao-tian¹, LIN Lei¹, SUN Hai-tao², LYU Yun-he², ZHANG Yan-wei¹, XUE Fei¹

(1. Suzhou Nuclear Power Research Institute, Suzhou 215004, China;

2. Nuclear and Radiation Safety Center, Beijing 100082, China)

ABSTRACT: Aiming at the problems of large error, complex calculation and poor applicability in the prediction service performance of materials, machine learning (ML) based on data mining was proposed. Firstly, the application process of ML is elaborated. Then, the principle of common models and its application in material performance prediction are summarized. Then, various ML models were used to predict the irradiance properties of RPV steel. Furthermore, the prediction accuracy was improved by Stacking integration method. Results show that ML can be used to predict the service performance of materials with high accuracy and reliability. Appropriate models should be selected according to diverse characteristics of materials service data. Model fusion and parameters optimization can improve the prediction accuracy and calculation speed of the ML model effectively.

KEY WORDS: data mining; machine learning; service performance; materials engineering; model prediction

收稿日期: 2021-05-26; 修订日期: 2021-07-25

Received: 2021-05-26; Revised: 2021-07-25

基金项目: 国家重点研发计划 (2017YFB0702200)

Fund: Supported by National Key Research and Development Program of China (2017YFB0702200)

作者简介: 王红珂 (1989—), 男, 硕士, 工程师, 主要研究方向为结构振动测试技术及材料力学评价。

Biography: WANG Hong-ke (1989—), Male, Master, Engineer, Research focus: structural vibration measurement technology and material mechanics evaluation.

通讯作者: 张晏玮 (1984—), 男, 博士, 高级工程师, 主要研究方向为核材料设计及性能评价。

Corresponding author: ZHANG Yan-wei (1984—), Male, Doctor, Senior engineer, Research focus: design and performance evaluation of nuclear materials.

引文格式: 王红珂, 刘啸天, 林磊, 等. 机器学习在材料服役性能预测中的应用[J]. 装备环境工程, 2022, 19(1): 011-019.

WANG Hong-ke, LIU Xiao-tian, LIN Lei, et al. Application of machine learning in predicting service performance of materials[J]. Equipment environmental engineering, 2022, 19(1): 011-019.

材料在服役过程中,由于受光照、热能、机械能、辐照、潮湿等因素的影响,会逐步发生老化,进而导致材料性能下降甚至失效^[1]。材料失效不仅带来巨大的经济损失,造成环境污染和资源浪费,甚至可能酿成安全事故,引发各种社会问题。因此,材料服役性能研究和服役寿命预测一直是材料领域的研究热点之一。

在早期研究中,通常将材料放置在自然环境或人工模拟环境中进行大量性能试验,并在试验过程中监测材料性能的变化情况。然后找出试验条件和材料性能之间的关系,进而预测材料服役性能的变化趋势和服役寿命。但这种方法通常需要投放大量的试样,试验周期漫长,无法真实反映出实际环境中不同因素之间的协同作用和综合效应,在客观性和普适性方面存在不足^[2]。

目前,材料服役性能研究主要分为四个方向:加速模拟实验^[3]、力学性能研究^[4]、数学模型^[5]和数据挖掘^[6]。其中,加速模拟实验是最简便可行的方案,但加速过程会掩盖材料服役的关键细节,难以区分不同条件对材料性能影响的细微差别。力学性能研究通过经典力学、断裂力学、疲劳力学、损伤力学等方法建立材料力学方程,然后进行力学分析和服役性能预测。但对于受力复杂的情况,建立的力学方程非常复杂,并且很难反映材料的真实受力情况。数学模型通过短期试验,推测长期服役条件下材料性能的变化情况,但该方法预测误差较大、计算复杂、普适性有待进一步研究。

数据挖掘通过机器学习,对大量材料服役数据进行学习和规律总结,然后对材料服役性能进行预测,目前已经在医药^[7]、生物信息^[8]、图像识别^[9]、故障诊断^[10]等领域取得了应用成果。文中首先阐述了机器学习在材料服役性能预测中的一般流程及常用机器学习模型原理,并对各种模型在材料服役性能研究应用中的特点进行总结;然后具体研究了机器学习在RPV钢辐照性能预测中的应用。研究结果为材料服役性能预测提供了新的方法和参考。

1 基于机器学习的材料服役性能预测

机器学习在材料服役性能预测的一般流程如图1所示,主要包括两个过程:1)从材料基因数据库中选择要处理的数据并进行预处理、特征选择和样本划分,然后通过机器学习模型进行训练、测试和评估,并获得最优模型及模型参数;2)通过最优模型对新数据进行服役性能预测,获得最终结果。

材料基因数据库为源数据,可源于试验、测试、仿真计算或其他数据库。数据库中数据的来源、格式、完整度等存在差异,需要进行数据清洗使其符合机器学习模型的输入要求。数据清洗包括异常值剔除、缺

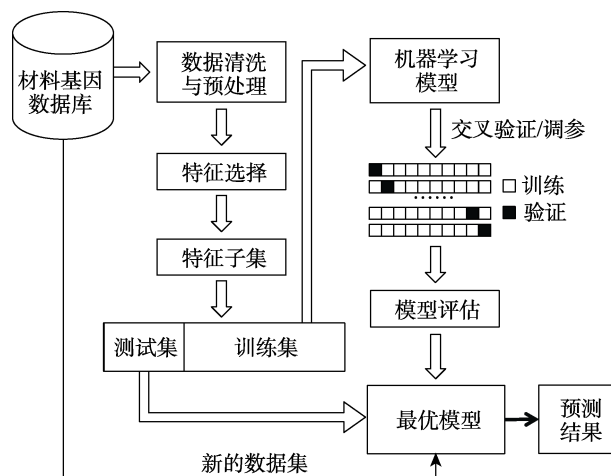


图1 基于机器学习的材料服役性能预测流程

Fig.1 Prediction process of material service performance based on machine learning

失值填充、重复样本删除、数字化等。对于不同量纲或存在数量级差异的特征值,还需要进行归一化或标准化处理,消除量纲带来的偏差。

数据清洗和预处理后需要进行特征选择,找出和目标相关度高的特征进行学习和训练,消除无关特征带来的噪声偏差。常用的特征选择方法有过滤式、包裹式和嵌入式,筛选后的特征子集可提高模型的训练效率和预测精度。

获得特征子集后,通过分层抽样将其划分为训练集和测试集,测试集约占总样本的20%~30%。然后选择合适的模型,并在训练集上进行模型训练和参数调优。模型训练后,根据预测结果进行模型评估,常用的评估指标有平均绝对误差、均方根误差和决定系数等。

假设 y_i 和 $\hat{y}_i$ 分别为第 i 个样本的真实值和预测值, $\bar{y}$ 表示样本的平均值,则各评估指标的计算方法如下。

1) 平均绝对误差 MAE。

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

2) 均方根误差 RMSE。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (2)$$

3) 决定系数 R^2 。

决定系数用来衡量模型对数据集波动的解释程度,最大值为1。 R^2 越接近1,表明模型的拟合越好。

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (3)$$

根据模型评估结果保存最优模型及参数,当有相同场景的新数据输入时,可直接通过最优模型对材料属性进行预测。

2 机器学习模型

根据训练数据是否拥有标记信息, 机器学习可分为监督学习和无监督学习。材料服役性能预测属于监督学习, 常用的学习模型有支持向量机、人工神经网络、决策树和随机森林等。

2.1 支持向量机

支持向量机的目的是获得 N 维空间的最优超平面^[11], 其用于二维空间的分类如图 2 所示。支持向量机可解决三类问题: 1) 样本线性可分, 通过硬间隔最大学习一个分类器; 2) 样本近似线性可分, 通过软间隔最大学习一个分类器; 3) 样本线性不可分, 通过核技法将低维非线性问题转化为高维线性问题, 然后学习一个非线性支持向量机。支持向量机可用于分类也可用于回归, 回归问题的目标函数为:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m l_{\varepsilon}(f(x_i) - y_i) \quad (4)$$

式中: $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ 为超平面的法向量; C 为正则化系数; l_{ε} 为损失函数。

加入核函数后, 式 (4) 可表示为:

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) k(x, x_i) + b \quad (5)$$

式中: $k(x, x_i)$ 为核函数; $\hat{\alpha}_i$ 和 α_i 为第 i 个样本的拉格朗日乘子。常用核函数有高斯核函数、多项式核函数、线性核函数和 Sigmoid 核函数。

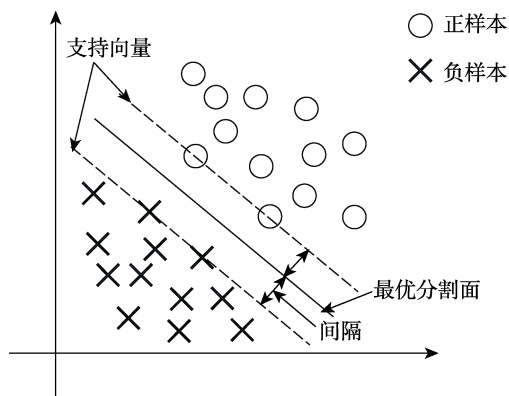


图 2 支持向量机分类示意

Fig.2 Classification schematic diagram of support vector machine (SVM)

2.2 人工神经网络

人工神经网络结构如图 3 所示, 主要包括正向传播和误差反向传递两个过程。其中, 正向传播的输出可表示为:

$$Y_{\text{pred}} = g_2(V^T g_1(W^T X + B_1) + B_2) \quad (6)$$

式中: W 、 B_1 、 g_1 分别为输入层到隐藏层的权重、偏置项和激活函数; V 、 B_2 、 g_2 分别为隐藏层到输出层的权重、偏置项和激活函数, Y_{pred} 为模型输出。

误差反向传递是根据预测值与期望值的偏差逐层回归计算正向传播过程中的误差, 并迭代更新各层神经元的权值, 回归问题的损失函数为:

$$E(\theta) = \frac{1}{N} \sum_{i=1}^N (Y - Y_{\text{pred}})^2 \quad (7)$$

然后根据损失函数和正向传播公式迭代更新各层的权值及偏置项, 并使损失函数越来越小。通过建立多层神经网络结构, 适当增加神经元个数和样本数量, 可提高模型的拟合能力。

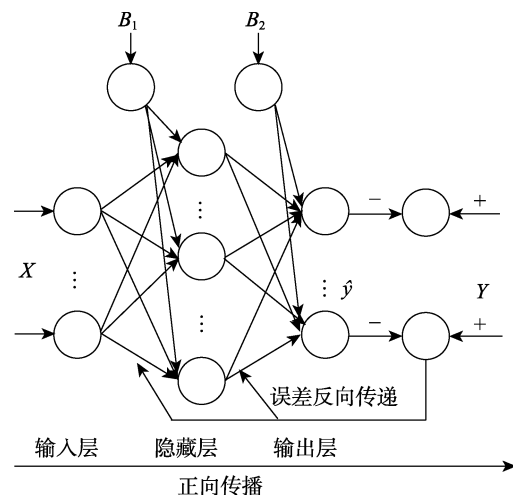


图 3 人工神经网络结构

Fig.3 Structure of artificial neural network

2.3 决策树

决策树是一种基于树结构划分的学习算法, 可分为分类树和回归树, 图 4 为一个深度为 3 层的分类树。决策树主要包括决策树生长和剪枝两个步骤^[12]: 1) 决策树生长采用的属性选择方法有 ID3、C4.5 和 Gini 指数, ID3 和 C4.5 分别采用信息增益和信息增益率进行属性选择; 2) 决策树剪枝是为了防止模型过拟合, 并提高模型的训练速度和识别能力, 通常采用“预剪枝”和“后剪枝”两种策略。

2.4 随机森林

随机森林回归算法的流程如图 5 所示, 其是一种基于决策树模型的 Boosting 集成算法, 可用于分类、回归和特征选择。首先从原始数据中随机抽取若干个样本组成样本子集, 然后基于每个样本子集建立决策树, 并构成树模型集合 $\{h(X, \theta_k), k=1, 2, \dots\}$ 。对于回归问题, $h(X, \theta_k)$ 采用 CART 算法生成回归树, θ_k 决定每棵树的生长过程。

3 机器学习在材料服役性能预测中的应用

材料服役性能与其结构、成分、环境条件等密切

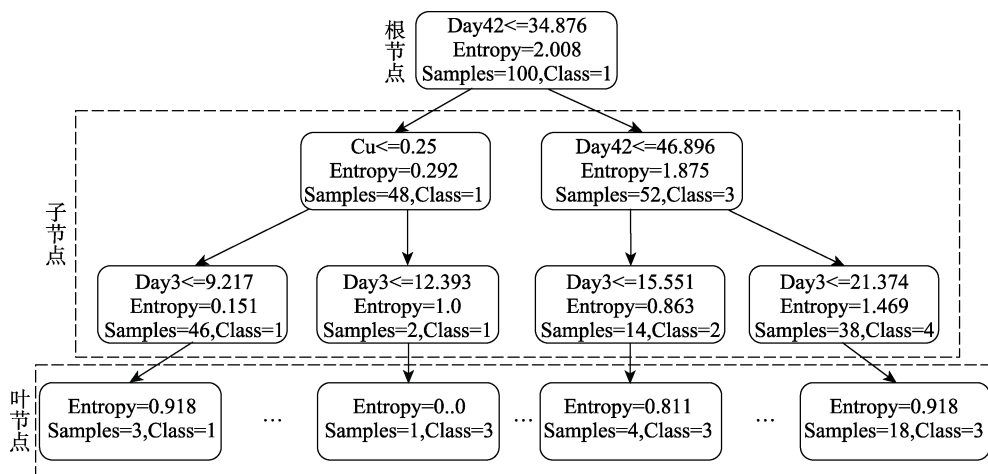


图 4 决策树结构示意图

Fig.4 Schematic diagram of decision tree structure

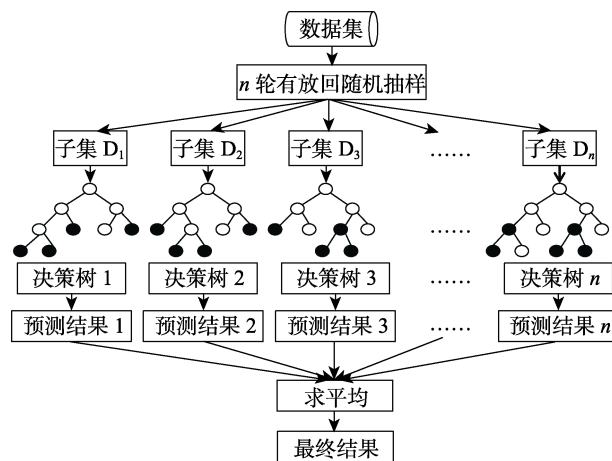


图 5 随机森林回归模型结构

Fig.5 Structure of random forest regression model

相关,不同因素之间的相互作用复杂。机器学习可以从大量服役数据中获得各因素之间的影响规律,并对服役性能进行预测。材料的服役性能主要包括腐蚀、磨损、辐照和疲劳等。

3.1 腐蚀性能预测

影响材料腐蚀的因素包括化学成分、组织结构、加工工艺、服役条件等,并且不同因素又包含多个维度信息,通常采用人工神经网络、支持向量机等拟合能力强、适用纬度高的模型进行腐蚀性能预测。

杜翠微等^[13]采用人工神经网络建立了碳钢和低合金钢的腐蚀速率预测模型,探究了化学成分、环境因素及多因素耦合作用对海水中金属材料腐蚀性能的影响。训练数据为我国典型海水区域长达 16 a 的腐蚀样本,模型输入为合金主要元素成分、pH、温度、溶解氧、盐度、生物附着物等。当仅考虑环境因素时,模型的预测精度达到 75%;当同时考虑环境因素和材料元素含量时,模型的预测精度可提高到 90% 以上。

LY12CZ 铝合金是飞机承重构件的主要材料,刘

延利等^[14]通过盐雾试验获得了铝合金材料的腐蚀试样,并对试样进行疲劳试验和腐蚀深度测量。然后基于腐蚀温度、腐蚀时间、最大点蚀深度及疲劳额定强度,建立了人工神经网络模型,模型对试样最大腐蚀深度的预测误差为 7.24%,对疲劳额定强度的预测误差为 1.63%。

为验证 RE-Ni-Cu 合金铸铁在碱液中的腐蚀性,王玉荣等^[15]通过静态质量损失腐蚀试验获得了 35 组腐蚀样本数据,然后建立了合金成分、腐蚀时间、碱液温度与合金铸铁腐蚀深度的 RBF 神经网络模型。RBF 神经网络对合金铸铁腐蚀深度的预测误差为 8.09%,对样本的耐腐蚀等级和耐腐蚀评价准确率达到了 100%。

海底管道腐蚀速率的有效预测可减少管道失效风险,李响等^[16]基于 Q235 钢材海洋挂片的腐蚀数据建立了环境温度、含氧量、含盐量、pH 和腐蚀速率之间的 SVM 模型。首先采用遗传算法对模型参数进行优化,优化后的模型对试样腐蚀速率的预测误差小于 3%,优于 BP 神经网络模型。

针对长输管道腐蚀速率的预测,马钢等^[17]建立了 PSO-SVM 模型,模型输入为 CO₂ 含量、硫化物含量、输送温度、输送压力和流速,输出为管道腐蚀速率。然后分别采用粒子群算法 (PSO)、遗传算法 (GA)、交叉验证 (CV)、最小二乘法 (LS) 和果蝇算法 (FOA) 对模型参数进行优化。预测结果见表 1,其中 PSO-SVM 模型的平均绝对误差和均方根误差最小,分别为 0.58% 和 6.18×10^{-4} ,但该模型的训练时间较长,仍需进一步优化。

支持向量机 (SVM) 可有效避免过拟合和局部极小值问题,适用于数据样本较小、特征维度中等的数据集。但是 SVM 模型参数对预测结果影响较大,通常结合主成分分析法、遗传算法、粒子群优化算法、交叉验证和最小二乘法等对模型参数进行优化,并采用优化后的参数进行模型训练和预测,从而提高模型效率和预测精度。

表 1 不同模型的时间及预测误差^[17]

Tab.1 Prediction time and error of different models^[17]

模型	训练时间/s	平均绝对误差/%	均方根误差/ 10^{-3}
PSO-SVM	11.26	0.58	0.618
LS-SVM	8.09	3.28	2.682
GA-SVM	6.98	10.72	9.271
CV-SVM	9.75	7.70	8.618
FOA-SVM	10.33	2.54	2.11

3.2 磨损性能预测

磨损与材料硬度、受力状态、摩擦系数等因素有关,但这些参数通常难以全部准确获得,容易造成数据缺失,通常采用对缺失值容忍度好的随机森林模型进行预测。由于磨损也是设备机械故障的一个重要原因,但磨损故障样本较小,容易造成样本不均衡,因此可选择适用于小样本的支持向量机模型进行预测。

随机森林具有可解释性好、对缺失值及异常值容忍度高等特点,可用于降维和重要特征筛选。赵帅等^[18]提出了一种基于随机森林和主成分分析法相结合的刀具磨损状态评估模型,首先通过降维获得刀具磨损的特征主分量,然后对刀具磨损状态进行评估,模型对刀具磨损状态的评估准确率达到 94.3%。降维和重要特征筛选提高了模型的效率和评估精度,降低了噪声对评估结果的影响,增强了模型的鲁棒性。

铁谱分析法是航空发动机磨损故障诊断的常用方法,但该方法依赖于专家经验和人工判断,诊断误差较大且效率很低。张建等^[19]提出一种基于免疫算法优化的支持向量机模型,首先通过免疫算法对模型的惩罚因子、松弛变量等进行优化,然后对航空发动机故障类型进行预测,预测准确率可达 98%,添加了 2%的噪声后,预测准确率仍大于 95%。

人工神经网络具有自适应性好、非线性建模能力强等优点,适用于特征维度高、规律复杂的场景。针对聚酰胺复合材料(PA6)磨损性能的预测,吕若云等^[20]提出一种基于 BP 神经网络和 RBF 神经网络的混合神经网络,结构如图 6 所示。模型输入为材料成分、载荷和转速,输出为摩擦因数和磨损率。混合神

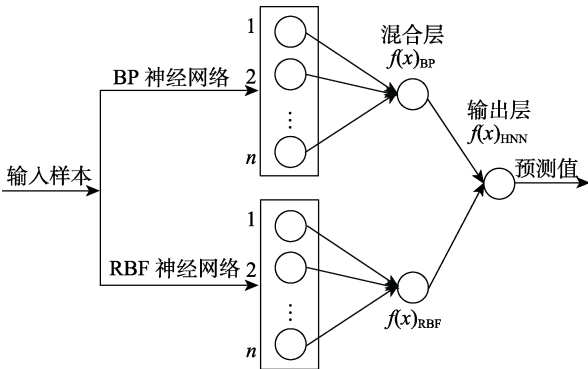


图 6 混合神经网络结构^[20]

Fig.6 Structure of hybrid neural network^[20]

经网络对材料的摩擦因数、磨损率的预测误差分别为 3.01%和 0.32%,为材料磨损性能的研究节省更多时间和成本。

磨损量和切削力通常采用解析法或经验公式进行预测,但存在预测精度低、数学模型复杂等局限性。李鑫等^[21]将改进后的神经网络用于合金车削刀具磨损量和切削力的预测,预测结果见表 2,神经网络的平均预测误差为 2.4%,小于经验公式 4.9%的预测误差。但是神经网络的预测精度高度依赖于训练样本,更倾向于训练样本范围内的数据,并且缺乏对切削加工过程中物理本质的描述,可解释性差。

表 2 神经网络模型与经验公式对刀具铣削力的预测结果^[21]

Tab.2 Prediction results of tool milling force by neural network model and empirical formula^[21]

编号	实验值/N	神经网络预测值/N	经验公式值/kN	神经网络预测误差/%	经验公式误差/%
1	238	240.68	229.40	1.12	-3.61
2	348	367.59	373.45	5.63	7.31
3	355	347.34	372.02	-2.16	4.80
4	351	341.07	374.35	-2.83	6.65
5	438	435.85	428.92	-0.49	-2.07
6	243	248.27	255.41	2.17	5.11

3.3 辐照性能预测

压力容器的辐照性能与材料成分、辐照条件、微观结构等密切相关,Mathew 等^[22]提出一种基于人工神经网络的压力容器辐照脆化预测模型,模型输入为化学成分、辐照温度、中子注量和注量率,输出为材料的转变温度。人工神经网络可以较好地反映出材料成分、辐照条件与辐照转变温度之间的关系,预测结果与试验结果基本一致。

针对辐照环境下材料硬度的预测,Kemp 等^[23]提出了一种基于贝叶斯框架的人工神经网络模型,并分析了温度、辐照剂量、化学成分对辐照后马氏体钢屈服应力的影响。神经网络模型可以很好地捕捉到屈服应力与化学成分及辐照条件之间的非线性关系,预测结果有助于材料性能试验设计及缺失信息的补充。

Castin 等^[24]提出了两种不同的人工神经网络模型,并用于压力容器钢辐照后屈服应力的预测。其中贝叶斯训练神经网络的预测结果更接近于试验值,预测精度要好于 BP 神经网络。通过对现有数据的训练和模型优化,实现了高中子注量条件下材料的辐照性能预测。

高功率微波辐照条件下电子元器件会发生损伤失效,金焱等^[25]采用支持向量机对电子元器件的损伤概率进行预测,同时采用模糊神经网络在相同条件下对算例进行了预测分析。支持向量机和模糊神经网络

都较好地获得了预测结果,但在小样本条件下支持向量机的预测精度更高,预测结果也更为稳定。

3.4 疲劳性能预测

疲劳是指交变载荷作用下结构产生裂纹或断裂的一种现象,通常用疲劳寿命或疲劳强度表示^[26]。材料疲劳失效的因素有很多,如材料成分、加工工艺、表面状态、受力情况等。不同因素的细微差别都可能造成疲劳性能的较大差异,因此需要根据材料疲劳失效因素的维度、样本数量、数据完整度等选择合适的学习模型。

Agrawal 等^[27]基于 NIMS 提供的钢材料疲劳数据集,对材料疲劳性能影响因素进行重要性筛选,然后采用数十种模型对材料疲劳寿命进行预测。其中人工神经网络、决策树、多项式回归模型的预测精度都达到了预期效果,模型均方根误差最小为 26.44。

为预测 316 不锈钢的低周疲劳寿命,Mathew 等^[28]

采用神经网络对不同温度和氮含量条件下各因素对材料疲劳性能的影响进行研究。其中,人工神经网络可准确预测 316 不锈钢在任何温度条件下的疲劳寿命,预测误差为 5.4%,具有较高的精度和鲁棒性。

尽管人工神经网络已经在复杂材料疲劳性能预测中发挥出良好的作用,但在材料疲劳样本较少时容易出现过拟合和局部极小值,模型预测精度大大降低。吴峰崎等^[29]建立了基于支持向量机的材料疲劳寿命预测模型,模型预测误差为 7.91%,可以在较少样本条件下实现材料疲劳寿命预测。

3.5 模型对比分析

通过对各种机器学习模型在材料服役性能预测中的应用分析,不同服役数据的特征不同、数据分布及特征复杂度存在差异。因此需要根据具体的应用场景及数据特点选择合适的模型进行训练和性能预测,常用机器学习模型的特点及使用场景见表 3。

表 3 常用机器学习模型性能对比
Tab.3 Performance comparison of common machine learning models

机器学习模型	优点	不足
随机森林 (Random Forest)	➢ 对异常值和缺失值容忍度好 ➢ 不易出现过拟合 ➢ 可解释性好 ➢ 可进行特征选择和降维	• 倾向于取值多的属性 • 噪声敏感 • 决策树个数较多时,模型训练开销大
支持向量机 (Support Vector Machine)	➢ 适用于小样本训练 ➢ 泛化能力强 ➢ 可有效避免过拟合和局部极小值 ➢ 适用于高维度样本训练	• 训练效率低 • 对缺失值比较敏感 • 模型参数和核函数选择比较敏感
人工神经网络 (Artificial Neural Network)	➢ 自适应学习能力强 ➢ 抗干扰能力强 ➢ 非线性映射能力强 ➢ 数据量大,预测精度高	• 需要大量训练样本 • 复杂度高、训练耗时 • 误差传播带来损失 • 可解释性差
决策树 (Decision Tree)	➢ 树结构可以可视化,便于理解 ➢ 不受数据缩放影响 ➢ 可处理有缺失值的样本	• 容易过拟合,泛化能力差 • 模型不稳定,噪声敏感

4 基于模型融合的 RPV 钢辐照性能预测

4.1 辐照数据

文中所用的 RPV 钢辐照数据来源于文献数据^[30]及试验数据,共有 390 条有效样本,数据分布如图 7 所示。首先剔除与辐照性能相关性较小的特征,最终选择中子注量、注量率、辐照温度、Cu 含量、Ni 含量、Mn 含量、Si 含量、P 含量作为模型输入,辐照后的转变温度增量 ΔRT_{NDT} 作为模型输出。

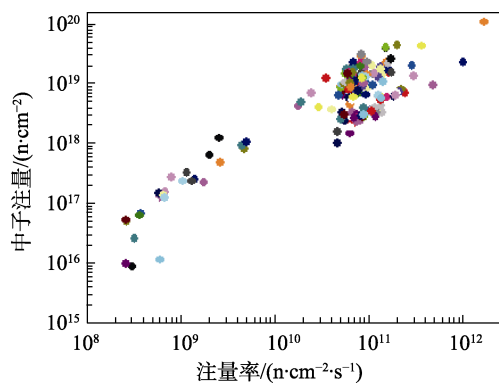


图 7 辐照数据分布

Fig. 7 Distribution of radiation data

4.2 数据划分

由于 RPV 钢辐照数据样本较小且离散性大，因此需要保证数据划分后仍保持原始数据的分布规律，避免数据不均衡造成预测偏差。首先对相关性较高的 Cu 含量进行区间划分并标记为 4 个类别，然后按类别进行分层抽样，抽样结果见表 4。其中，分层抽样后的样本分布与原始数据基本相同，Cu 含量类别“4”的样本占比误差最大，为-4.76%；随机抽样后的样本分布与原始样本差异较大，最大的占比误差为 8.57%。分层抽样后，将 80%的数据用作训练，20%用作测试。

表 4 数据抽样结果及误差
Tab.4 Results and errors of data sampling

Cu 含量类别	原始数据		随机抽样		分层抽样	
	样本数	占比	样本数	占比误差	样本数	占比误差
1	125	0.32	115	-8.00%	125	0%
2	175	0.45	190	8.57%	175	0%
3	69	0.18	65	-5.80%	70	1.45%
4	21	0.05	20	-4.76%	20	-4.76%

4.3 模型预测

数据划分后采用单一模型进行训练和预测，并对各种模型的预测结果进行评估，结果如图 8 所示。其中，极端梯度增强算法模型（Extreme Gradient Boosting, XGBoost）的均方根误差最小为 10.16， R^2 达到 0.88。基于树模型的集成算法，如随机森林（Random Forest, RF）、梯度提升决策树（Gradient Boosting Decision Tree, GBDT）、自适应提升算法（Adaptive Boosting, AdaBoost）、XGBoost 和轻量级梯度提升算法（Light Gradient Boosting Machine,

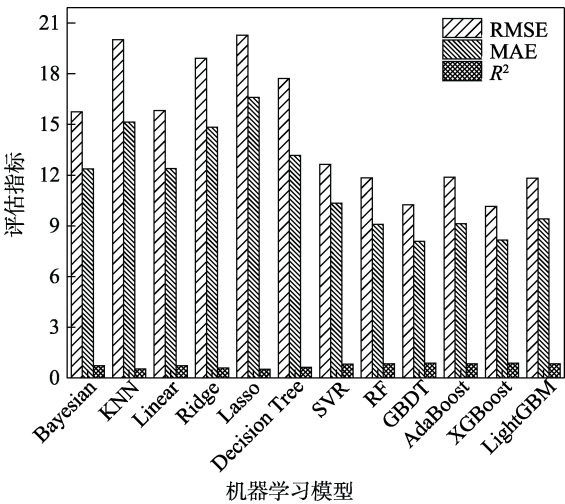


图 8 各种模型评估结果
Fig.8 Evaluation results of various models

LightGBM）等模型的预测平均绝对误差均小于 10，表明树模型对本数据集的预测精度相对较高。

为提高辐照性能的预测精度，采用 Stacking 方法在单一模型预测结果的基础上再次进行模型训练和预测。方法流程如图 9 所示，具体步骤如下：

1) 将训练集均分为 5 个样本子集，4 个用于训练，1 个用于预测；

2) 采用 RF、GBDT、AdaBoost、XGBoost 和 LightGBM 模型分别在 4 个训练子集上进行训练，并对剩余 1 个样本子集进行预测。五个模型的预测结果重组为新训练集，大小与训练集相同；

3) 采用上述五个模型分别对测试集进行预测，并将预测结果求平均后作为新测试集；

4) 将新训练集和新测试集组成一个新数据集，然后采用支持向量机回归（Support Vector Regression, SVR）模型在新数据集上进行训练。

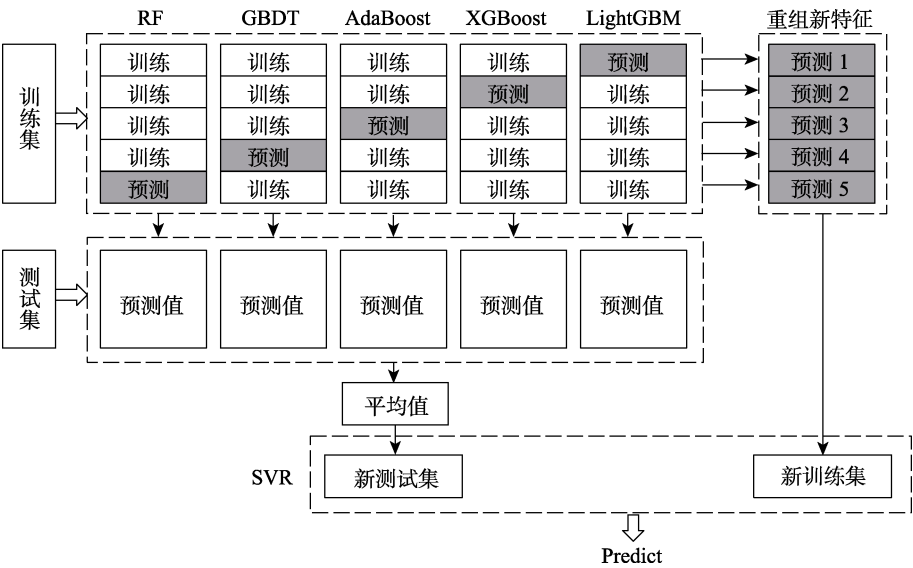


图 9 基于 Stacking 集成方法的预测流程
Fig.9 Prediction process based on stacking integration method

基于 Stacking 方法的预测结果如图 10 所示, 模型预测值与试验值基本一致, 全部落在 45° 线附近区域。模型预测结果与试验值的均方根误差 RMSE 为 9.94, 平均绝对值误差 MAE 为 8.01, R^2 达到 0.89。相比于单一模型, 基于 Stacking 方法的集成模型预测性能得到提升, 对 RPV 钢辐照性能预测具有较高的准确度和可靠性。

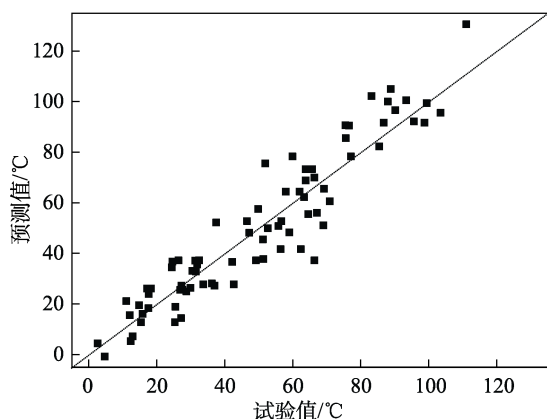


图 10 集成模型预测结果

Fig.10 Prediction results of the integrated model

5 总结与展望

从机器学习流程出发, 首先概括了常用机器学习模型的原理及其在材料服役性能预测中的应用; 然后采用多种机器学习模型对 RPV 钢的辐照性能进行预测, 并采用 Stacking 方法提高了模型的预测精度。主要结论如下:

- 1) 机器学习可用于材料服役性能预测, 具有较高的预测精度和可靠性;
- 2) 模型选择应考虑材料服役数据的质量、模型特点及样本数量;
- 3) 重要特征提取、模型融合和参数优化算法可提高模型的精度和运算效率。

目前, 机器学习在材料服役性能预测中的应用仍处于初级阶段, 大量工作仍需进一步开展。如丰富完善已有的材料基因数据库, 开展更高精度机器学习算法研究, 优化机器学习模型参数, 提高运算效率, 结合物理模型及相关理论, 增强机器学习模型的解释性。

参考文献:

- [1] 王强, 季洋, 徐大可. 医用金属材料腐蚀疲劳性能研究进展[J]. 表面技术, 2019, 48(7): 193-199, 210.
WANG Qiang, JI Yang, XU Da-ke. Research Progress on the Corrosion Fatigue of Biomedical Metallic Alloys[J]. Surface Technology, 2019, 48(7): 193-199, 210.
- [2] 梁义, 王博, 魏世丞, 等. 海洋飞溅区 Zn 涂层加速腐蚀

实验模拟研究[J]. 材料导报, 2017, 31(10): 51-55.

- LIANG Yi, WANG Bo, WEI Shi-cheng, et al. Simulation Study on the Accelerated Corrosion of Zn Coating in Splash Zone[J]. Materials Review, 2017, 31(10): 51-55.
- [3] 卞贵学, 张杨广, 张勇, 等. 飞机某模拟涂装试样加速腐蚀与自然暴露的对比研究[J]. 装备环境工程, 2020, 17(5): 25-32.
BIAN Gui-xue, ZHANG Yang-guang, ZHANG Yong, et al. Comparative Study on Accelerated Corrosion and Natural Exposure of Simulated Coated Specimen of Aircraft Structures[J]. Equipment Environmental Engineering, 2020, 17(5): 25-32.
- [4] 龚美琦, 张春涛, 王汝恒. 碳纤维布加固腐蚀钢材力学性能试验研究[J]. 重庆理工大学学报(自然科学), 2021, 35(3): 106-111.
GONG Mei-qi, ZHANG Chun-tao, WANG Ru-heng. Experimental Study on Mechanical Properties of Corroded Steel Strengthened with CFRP[J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(3): 106-111.
- [5] 偶国富, 刘旭, 金浩哲, 等. 液固两相流对 A182F347 不锈钢材料的磨损特性[J]. 煤炭学报, 2016, 41(11): 2883-2888.
OU Guo-fu, LIU Xu, JIN Hao-zhe, et al. Erosion Resistance of A182F347 Stainless Steel in Liquid-Solid Two Phase Flow System[J]. Journal of China Coal Society, 2016, 41(11): 2883-2888.
- [6] 胡建军, 曹卓, 但雅波, 等. 基于特征选择和机器学习的材料弹性性能预测[J]. 华南理工大学学报(自然科学版), 2019, 47(5): 48-55.
HU Jian-jun, CAO Zhuo, DAN Ya-bo, et al. Elastic Property Prediction of Materials Based on Machine Learning and Feature Selection[J]. Journal of South China University of Technology (Natural Science Edition), 2019, 47(5): 48-55.
- [7] 谢志勇, 周翔. 基于机器学习的医学影像分析在药物研发和精准医疗方面的应用[J]. 中国生物工程杂志, 2019, 39(2): 90-100.
XIE Zhi-yong, ZHOU Xiang. Machine Learning in Medical Imaging: The Applications in Drug Discovery and Precision Medicine[J]. China Biotechnology, 2019, 39(2): 90-100.
- [8] 张晓龙, 杨艳霞. 机器学习在生物信息学中的应用[J]. 武汉科技大学学报(自然科学版), 2005, 28(2): 201-204.
ZHANG Xiao-long, YANG Yan-xia. Application of Machine Learning in Bioinformatics[J]. Journal of Wuhan Yejin University of Science and Technology, 2005, 28(2): 201-204.
- [9] 刘飞, 张俊然, 杨豪. 基于深度学习的医学图像识别研究进展[J]. 中国生物医学工程学报, 2018, 37(1): 86-94.
LIU Fei, ZHANG Jun-ran, YANG Hao. Research Progress of Medical Image Recognition Based on Deep Learning[J]. Chinese Journal of Biomedical Engineering, 2018, 37(1): 86-94.
- [10] 李红贤, 韩延, 吴敬涛, 等. 基于 ICA 包络增强 MEMD 的滚动轴承故障诊断[J]. 航空动力学报, 2021, 36(2): 405-412.
LI Hong-xian, HAN Yan, WU Jing-tao, et al. Rolling Bearing Fault Diagnosis Based on MEMD with ICA En-

- velop Enhancement[J]. Journal of Aerospace Power, 2021, 36(2): 405-412.
- [11] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报, 2011, 40(1): 2-10.
DING Shi-fei, QI Bing-juan, TAN Hong-yan. An Overview on Theory and Algorithm of Support Vector Machines[J]. Journal of University of Electronic Science and Technology of China, 2011, 40(1): 2-10.
- [12] 邵旻晖. 决策树典型算法研究综述[J]. 电脑知识与技术, 2018, 14(8): 175-177.
SHAO Min-hui. A Review of Typical Decision Tree Algorithms[J]. Computer Knowledge and Technology, 2018, 14(8): 175-177.
- [13] 杜翠微, 赵研研, 李磊, 等. 基于碳钢和低合金钢16年海水暴露腐蚀数据的神经网络预测模型[J]. 中国建材科技, 2010, 19(S1): 118-126.
DU Cui-wei, ZHAO Yan-yan, LI Lei, et al. The ANN Predictive Models for Seawater Corrosion of Carbon Steel and Low-Alloy Steel Exposed for 16 Years[J]. China Building Materials Science & Technology, 2010, 19(S1): 118-126.
- [14] 刘延利, 钟群鹏, 张峥. 基于人工神经网络的预腐蚀铝合金疲劳性能预测[J]. 航空学报, 2001, 22(2): 135-139.
LIU Yan-li, ZHONG Qun-peng, ZHANG Zheng. Predictive Model Based on Artificial Neural Net for Fatigue Performances of prior-Corroded Aluminum Alloys[J]. Acta Aeronautica et Astronautica Sinica, 2001, 22(2): 135-139.
- [15] 王玉荣, 乌日根. 基于RBF神经网络RE-Ni-Cu合金铸铁静态腐蚀性能预测[J]. 兵器材料科学与工程, 2014, 37(6): 66-68.
WANG Yu-rong, WU Ri-gen. Static Corrosion Prediction of RE-Ni-Cu Alloy Cast Iron Based on RBF Neural Network[J]. Ordnance Material Science and Engineering, 2014, 37(6): 66-68.
- [16] 李响, 余建星, 苗春生, 等. 基于遗传算法 SVM 的海洋环境腐蚀速率预测[J]. 中国海洋平台, 2018, 33(5): 56-62.
LI Xiang, YU Jian-xing, MIAO Chun-sheng, et al. Prediction of Corrosion Rate in Marine Environment Based on Support Vector Machine Combining with Genetic Algorithm[J]. China Offshore Platform, 2018, 33(5): 56-62.
- [17] 马钢, 李俊飞, 白瑞, 等. 基于 PSO-SVM 模型的油气管道内腐蚀速率预测[J]. 表面技术, 2019, 48(5): 43-48.
MA Gang, LI Jun-fei, BAI Rui, et al. Prediction of Corrosion Rate in Oil and Gas Pipelines Based on PSO-SVM Model[J]. Surface Technology, 2019, 48(5): 43-48.
- [18] 赵帅, 黄亦翔, 王浩任, 等. 基于随机森林与主成分分析的刀具磨损评估[J]. 机械工程学报, 2017, 53(21): 181-189.
ZHAO Shuai, HUANG Yi-xiang, WANG Hao-ren, et al. Random Forest and Principle Components Analysis Based on Health Assessment Methodology for Tool Wear[J]. Journal of Mechanical Engineering, 2017, 53(21): 181-189.
- [19] 张建, 李艳军, 曹愈远, 等. 免疫支持向量机用于航空发动机磨损故障诊断[J]. 北京航空航天大学学报, 2017, 43(7): 1419-1425.
ZHANG Jian, LI Yan-jun, CAO Yu-yuan, et al. Immune SVM Used in Wear Fault Diagnosis of Aircraft Engine[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(7): 1419-1425.
- [20] 吕若云, 李笃信. 基于混合神经网络的 PA6 摩擦磨损性能预测[J]. 塑料, 2016, 45(1): 86-89.
LYU Ruo-yun, LI Du-xin. Prediction of Tribological Properties for PA6 Based on Hybrid Neural Network[J]. Plastics, 2016, 45(1): 86-89.
- [21] 李鑫, 史振宇, 蒋森河, 等. 人工神经网络预测刀具磨损和切削力[J]. 控制理论与应用, 2018, 35(12): 1731-1737.
LI Xin, SHI Zhen-yu, JIANG Sen-he, et al. Artificial Neural Network Predicts Tool Wear and Cutting Force[J]. Control Theory & Applications, 2018, 35(12): 1731-1737.
- [22] MATHEW J, PARFITT D, WILFORD K, et al. Reactor Pressure Vessel Embrittlement: Insights from Neural Network Modelling[J]. Journal of Nuclear Materials, 2018, 502: 311-322.
- [23] KEMP R, COTTRELL G A, BHADOSHIA H K D H, et al. Neural-Network Analysis of Irradiation Hardening in Low-Activation Steels[J]. Journal of Nuclear Materials, 2006, 348(3): 311-328.
- [24] CASTIN N, MALERBA L, CHAOUADI R. Prediction of Radiation Induced Hardening of Reactor Pressure Vessel Steels Using Artificial Neural Networks[J]. Journal of Nuclear Materials, 2011, 408(1): 30-39.
- [25] 金焱, 胡云安, 黄隽, 等. 支持向量机回归在电子器件易损性评估中的应用[J]. 强激光与粒子束, 2012, 24(9): 2145-2150.
JIN Yan, HU Yu-nan, HUANG Jun, et al. Application of Support Vector Regression to Vulnerability Assessment of Electronic Devices Illuminated or Injected by High Power Microwave[J]. High Power Laser and Particle Beams, 2012, 24(9): 2145-2150.
- [26] DESHPANDE P D, GAUTHAM B P, CECEN A, et al. Application of Statistical and Machine Learning Techniques for Correlating Properties to Composition and Manufacturing Processes of Steels Proceedings of the 2nd World Congress on Integrated Computational Materials Engineering (ICME). Springer International Publishing, 2013: 155-160.
- [27] AGRAWAL A, DESHPANDE P D, CECEN A, et al. Exploration of Data Science Techniques to Predict Fatigue Strength of Steel from Composition and Processing Parameters[J]. Integrating Materials and Manufacturing Innovation, 2014, 3(1): 1-19.
- [28] MATHEW M D, KIM D W, RYU W S. A Neural Network Model to Predict Low Cycle Fatigue Life of Nitrogen-Alloyed 316L Stainless Steel[J]. Materials Science and Engineering: A, 2008, 474(1-2): 247-253.
- [29] 吴峰崎, 刘龙. 基于支持向量机回归算法的疲劳寿命预测研究[J]. 起重运输机械, 2015(2): 5-8.
WU Feng-qi, LIU Long. Study on Fatigue Life Evaluation Based on Regression Algorithm of Support Vector Machine[J]. Hoisting and Conveying Machinery, 2015(2): 5-8.
- [30] ELLIOT B J, HACKETT E M, LEE A D. Reactor Pressure Vessel. Status Report[R]. Office of Scientific and Technical Information (OSTI), 1996.