

电潜泵剩余使用寿命预测集成学习算法研究^{*}

郑文培^{1,2} 周少杰^{1,2} 王颖君³ 周涛涛^{1,2}

(1. 中国石油大学(北京)安全与海洋工程学院 北京 102249; 2. 应急管理部油气生产安全与应急技术重点实验室 北京 102249; 3. 中海油研究总院有限责任公司 北京 100027)

摘要:电潜泵采油是目前海上油田最主要的采油方式之一,其故障会影响油井正常生产运行并造成经济损失,因此,进行电潜泵剩余使用寿命预测与故障预防尤为重要。为保障电潜泵优质运行,根据电潜泵的数据特征,提出一种基于集成学习模型的剩余使用寿命预测方法。首先计算各时间点剩余使用寿命作为标签函数,利用随机森林算法筛选高贡献度特征参数输入模型,构建由麻雀搜索算法-卷积神经网络(SSA-CNN)和麻雀搜索算法-长短期记忆(SSA-LSTM)两个基模型经绝对误差加权组成的集成模型。现场数据验证表明,两个基模型算法在不同情况下具备各自的优势和劣势,SSA-CNN在数据波动期更具优势,SSA-LSTM整体预测更为准确,将相同数据代入集成模型中,发现集成模型的预测误差明显小于两个基模型的预测误差兼具两者优势,在整体精度和变化阶段的评估准确率方面均有显著改善。实际算例验证表明,集成模型的预测精度相较基模型提升6.41%,较现有方法有显著提升,具备较强的鲁棒性和稳定性。

关键词:电潜泵; SSA-CNN; SSA-LSTM; 集成学习模型

中图分类号: X937; TN06 **文献标识码:** A **国家标准学科分类代码:** 620.30; 510.40

Prediction of remaining service life of electric submersible pump based on ensemble learning model

Zheng Wenpei^{1,2} Zhou Shaojie^{1,2} Wang Yingjun³ Zhou Taotao^{1,2}

(1. School of Safety and Ocean Engineering, China University of Petroleum(Beijing), Beijing 102249, China;
2. Key Laboratory of Oil and Gas Safety and Emergency Technology, Ministry of Emergency Management,
Beijing 102249, China; 3. CNOOC Research Institute Co. Ltd., Beijing 100027, China)

Abstract: Electric submersible pump oil recovery is currently one of the most important oil recovery methods in offshore oil fields. Its failure can affect the normal production and operation of oil wells and cause economic losses. Therefore, it is particularly important to predict the remaining service life of electric submersible pumps and prevent failures. To ensure the high-quality operation of electric submersible pumps, a remaining service life prediction method based on ensemble learning model is proposed according to the data characteristics of electric submersible pumps. Firstly, the remaining service life at each time point is calculated as the label function, and the random forest algorithm is used to screen the high contribution feature parameters input model. An ensemble model consisting of SSA-CNN and SSA-LSTM base models weighted by absolute error is constructed. On site data verification shows that the two base model algorithms have their own advantages and disadvantages in different situations. SSA-CNN has more advantages during data fluctuation periods, while SSA-LSTM has more accurate overall predictions. When the same data is input into the ensemble model, it is found that the prediction error of the ensemble model is significantly smaller than that of the two base models, combining the advantages of both. There is a significant improvement in overall accuracy and evaluation accuracy during the change stage. The actual calculation verification shows that the prediction accuracy of the integrated model is 6.41% higher than that of the base model, which is significantly improved compared to existing methods and has strong robustness and stability.

Keywords: electric submersible pump; SSA-CNN; SSA-LSTM; ensemble learning model

收稿日期: 2024-09-04 Received Date: 2024-09-04

^{*} 基金项目: 国家重点研发计划(2024YFB3409301)、国家自然科学基金青年项目(72301294)、北京市科协青年人才托举工程(BYESS2023323)、中国石油大学(北京)科研启动基金(2462023BJRC016)项目资助

0 引言

电潜泵作为油田开采中的重要设备,承担着输送原油的任务,发挥着至关重要的作用。然而,由于其长时间运行和恶劣的工作环境,电潜泵常常面临各种故障风险。因此,进行潜油电泵剩余使用寿命预测显得尤为重要。

对于设备剩余寿命预测的研究方法,大致分为两类,包括基于物理模型的传统预测方法和基于机器学习算法模型预测方法。结合当下设备使用环境与影响因素复杂的特点,机器学习算法模型已成为研究学者在设备预测领域研究的重点。Khabibullin 等^[1]介绍了一种基于现场数据的电潜泵 (electric submersible pump, ESP) 故障预测方法,该方法主要是采用决策树算法建立集成模型,进行不同预测层位的研究。Ghimire 等^[2]提出了混合人工智能模型将卷积神经网络作为特征提取算法与极限学习机相结合进行预测。赵婧宇等^[3]提出了一种利用麻雀搜索算法 (sparrow search algorithm, SSA) 优化长短期记忆网络 (long short-term memory, LSTM) 参数的电力负荷预测模型。韩腾飞^[4]针对单一预测模型性能波动较大和泛化能力弱的问题,提出一种融合多个相异模型的 Stacking 集成学习剩余使用寿命预测模型。赵岩^[5]针对集成学习基于传输理论与轻量级梯度提升机 (light gradient boosting machine, Light GBM) 模型的性能受限于传统特征提取方法和 LeNet-5 网络在面对设备复杂工况时诊断性能较差,提出一种改进 LeNet-5 网络结合 Light GBM 模型的故障诊断方法。江兵等^[6]提出一种蚁群算法结合改进主成分分析法优化反向传播 (back propagation, BP) 神经网络热点温度预测模型。张新生等^[7]提出了一种基于 SSA 优化卷积神经网络 (convolutional neural networks, CNN) 算法的矿浆管道临界流速预测模型。嵇晓燕等^[8]将自回归积分滑动平均模型 (autoregressive integrated moving average model, ARIMA) 和 Prophet 模型结合 BP 神经网络模型,构建了一种针对水质预测的集成学习模型,研究结果表明,这种集成学习模型的预测精度明显优于单一模型。

可以看到,国内外的学者倾向于用优化算法对某模型进行优化,或用几种算法结合建立起集成学习模型进行预测,通过相应结合方式将各个基础算法模型取长补短,以期达到更好的预测效果。但目前,电潜泵现场运行时影响因素众多、监测信息的数量庞大,大量信息可能导致预测模型运算量呈几何级增加,给预测计算带来不小的压力;其次,信息过载可能使得训练的预测模型难以准确捕捉关键特征^[9],降低了故障预测的准确性。

现有研究方法针对影响因素多、监测数据量大条件下的电潜泵剩余寿命预测的研究相对缺少,鉴于此,本

文选择运用随机森林算法筛选高贡献度特征参数,并建立由 SSA-CNN、SSA-LSTM 两个基模型算法,经过绝对误差加权的集成学习预测模型进行电潜泵剩余使用寿命预测,促进解决电潜泵影响因素多、数据量大的问题,同时利用集成模型提升数据训练效果,提高预测精度,缩短预测时间,弥补单一预测方法的不足。

1 集成学习模型基算法原理

1.1 LSTM 网络算法原理

LSTM 网络是一种特殊的循环神经网络。针对循环神经网络难以处理长序列的记忆问题,在传统模型的基础之上增加了输入、输出、忘记控制器对其循环单元进行改进^[10],通过引入“门”结构和“细胞状态”来解决这个问题。门结构可以控制信息的流入和流出,细胞状态则可以长期存储状态。这使得 LSTM 能够在长序列中更好地捕捉依赖关系,序列数据中的趋势、周期性和突发性等特征;挖掘长范围的历史时间数据与未来时间数据的高耦合关联性^[11],因此, LSTM 被广泛的应用于具有时间延续性数据的预测任务中。

LSTM 的核心思想是引入了 3 个门控单元:输入门、遗忘门和输出门,以及 1 个细胞状态,通过这些门控机制来控制信息的流动和记忆的更新。

1) 输入门决定了哪些信息将被更新到细胞状态中。它通过 Sigmoid 激活函数来控制每个元素的输入,其计算公式如式 (1) 所示。

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i) \quad (1)$$

式中: i_t 是输入门的输出; x_t 是当前时间步的输入; h_{t-1} 是上一个时间步的隐藏状态; W_{xi} 和 W_{hi} 是输入和隐藏状态的权重矩阵; b_i 是偏置向量; σ 表示 sigmoid 函数。

2) 遗忘门决定了哪些信息将被保留和遗忘。它通过 sigmoid 激活函数来控制每个元素的遗忘程度,其计算公式如式 (2) 所示。

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f) \quad (2)$$

式中: f_t 是遗忘门的输出;其他符号的含义与输入门类似。

细胞的状态是通过输入门和遗忘门来更新的。首先,通过输入门来更新候选细胞状态,然后通过遗忘门来控制保留和遗忘原有的细胞状态。其计算公式如式 (3) 和 (4) 所示。

$$\tilde{C}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c) \quad (3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

式中: $\tilde{C}_t$ 是候选细胞状态; C_t 是当前时间步的细胞状态。

3) 输出门决定了当前时间步的隐藏状态,以及最终

的输出。它通过 sigmoid 激活函数来控制每个元素的输出,其计算公式如式(5)和(6)所示。

$$o_i = \sigma(\mathbf{W}_{xo} \cdot \mathbf{x}_i + \mathbf{W}_{ho} \cdot \mathbf{h}_{i-1} + \mathbf{b}_o) \quad (5)$$

$$h_i = o_i \cdot \tanh(C_i) \quad (6)$$

式中: o_i 是输出门的输出; h_i 是当前时间步的隐藏状态。

1.2 CNN 算法原理

CNN 作为深度学习的代表算法之一,在分类回归问题上发挥着重要作用^[12]。卷积神经网络具备表征学习的能力^[13],可以对输入信息进行平移不变分类,并且可进行监督学习和非监督学习,其隐含层内的参数共享和层间连接的稀疏性,使其能够以较小的计算量对格点化特征进行处理。尽管 CNN 主要用于处理图像数据,但它同样适用于时间序列预测^[14],特别是一维卷积层在提取时间序列特征方面具有较高的效率。

CNN 的基本结构大致包括卷积层、激活函数、池化层、输出层等^[15]。

1) 卷积层是深度学习神经网络中的一种重要层级结构,用于提取输入数据中的特征。通过在输入数据上滑动滤波器并执行卷积操作,卷积层可以生成一系列特征图,其中每个特征图对应于一种检测到的特征。这些特征图随后被传递到网络的下一层用于进一步处理和学习。假设输入数据为 X ,卷积核为 K ,那么卷积操作可以表示为:

$$Y(i, j) = (X \times K)(i, j) = \sum_m \sum_n X(m, n) \cdot K(i - m, j - n) \quad (7)$$

式中: $Y(i, j)$ 表示卷积后的输出; $X(m, n)$ 表示输入数据的元素; $K(i - m, j - n)$ 表示卷积核在位置 $(i - m, j - n)$ 的权重。

2) 激活函数在人工神经网络中扮演着关键的角色,其主要作用是引入非线性特性,从而使网络能够学习和表示复杂的非线性函数关系。本文建立的神经网络模型激活函数是 ReLU,它是一种非线性激活函数,可以使得神经网络具有更强的表达能力。ReLU 函数的数学表达式为:

$$\text{ReLU}(x) = \max(0, x) \quad (8)$$

3) 池化层通常紧随在卷积层之后,用于对数据进行维度的降低和压缩。在池化过程中,通过滑动窗口在输入数据上移动,并在每个窗口区域内选择最大值或计算均值作为输出。通过应用池化层可以有效地降低网络的计算压力,进而提升训练效率。最大池化的数学表达式如式(9)所示。

$$M(i, j) = \max_{m, n} I(i + s \cdot m, j + s \cdot n) \quad (9)$$

式中: $M(i, j)$ 是输出的池化结果; I 是输入数据特征; (i, j) 是输出池化结果的坐标; (m, n) 是池化窗口的坐标; s 是池化窗口的大小。

1.3 SSA 原理

CNN 与 LSTM 模型都是近几年炙手可热的计算模型,它们在各自的领域和任务中表现出色^[16]。但是这两种模型同样存在不同的局限性;为了提高基础算法的预测精度,采用 SSA 对其进行优化。

SSA 是近些年提出的新算法,在工业上主要应用于各种参数的预测,例如电力、风电功率等;在其他领域,水利、空气质量方面也有一定的应用^[17],它是一种基于自然界麻雀觅食行为的启发式优化算法,具有较强的全局优化能力^[18]。对于复杂的 CNN 和 LSTM 模型,特征提取对电潜泵剩余使用寿命预测的准确性有重要影响,采用 SSA 进行优化可以更好地搜索参数空间,寻找到更优的模型参数配置。对于 CNN,SSA 主要优化模型中的超参数配置,如卷积层的数量、卷积核的大小、步长等,这些超参数直接影响特征提取的效果和模型的表现能力,同时,可以帮助模型从多尺度上提取信息以应对复杂的电潜泵设备数据,增强模型泛化能力;对于 LSTM,SSA 可以设置一个定量种群大小及迭代次数,能够为 LSTM 在该范围内寻求最佳的参数,如 LSTM 隐藏层的层数、初始学习效率、最大训练周期等,同时,针对电潜泵数据特征,SSA 通过调整 LSTM 模型的时间步长和隐藏状态维度,使其能够更好地捕捉设备状态的长期依赖性,提高设备剩余寿命的预测精度。

此外,SSA 具有一定的随机性,能够帮助算法跳出局部最优点,避免陷入局部最优解,这对于 CNN 和 LSTM 等深度学习模型来说尤为重要,因为这些模型的参数空间通常非常大,存在许多局部最优点;并且 SSA 具有较好的自适应性^[19],能够在搜索过程中动态调整搜索策略,根据当前搜索状态进行调整,有助于更快地收敛到全局最优解;与其他优化算法相比,SSA 算法通常具有较高的搜索效率和收敛速度^[20],这对于优化大规模模型来说同样很重要,可以节省大量的时间和计算资源。

2 基于局部误差的集成学习模型建立

集成学习方法通过将多个基模型的预测形式进行组合,从而实现单一模型局限性的有效补偿^[21]。通过这种方式,不仅能够增强模型对输入数据的适应能力,还能提升其对设备剩余寿命预测的性能和可靠性。

2.1 集成学习模型原理

随机森林算法是以决策树为基学习器的集成学习算法,它利用多个决策树来进行分类或回归任务^[22]。随机森林中的每棵决策树都是由对训练集进行自助采样得到的随机样本构建的,同时每个节点的划分特征也是随机

选择的,这使得它具有更好的泛化能力和抗过拟合能力;针对电潜泵的数据集特点,随机森林算法能够很好地捕捉数据中的非线性关系,因为每棵决策树都可以自由地对特征空间进行划分,不受线性关系的限制。

基尼指数(Gini index)是衡量分类模型中特征的纯度或不确定性的指标之一。基尼指数越小的特征在决策树的节点分裂过程中具有更高的重要性,因为它们能够更有效地将数据集分割为纯度更高的子集。通过计算各个特征的基尼指数,可以确定哪些特征对于决策树模型的构建是最重要的,进而进行特征选择和模型优化。Gini 指数的计算公式为:

$$Gini(D) = 1 - \sum_{i=1}^n (p_i)^2 \quad (10)$$

式中: D 表示数据集; n 表示数据集中的类别数; p_i 表示数据集中属于第 i 个类别的样本占总样本的比例。

2.2 集成学习模型建立原理

集成学习是一种通过组合多个单一模型来构建更强大的模型的机器学习技术,即通过结合多个弱学习器的预测结果,最终得到一个更准确、更稳定的整体预测^[23]。集成学习的核心思想是结合多个模型,通过某种策略将多个基学习器的预测结果进行结合,从而得到最终的整体预测^[24]。通过结合多个模型的预测结果,可以在保持准确性的同时提高模型的稳定性和泛化能力,从而更好地适应未知数据。

局部误差在数值方法中具有重要意义,其涵盖了模型在每个特定数据点或样本上的预测准确度^[25]。在处理时序数据等预测任务中,局部误差的分析有助于识别模型的局部缺陷,指导模型的优化方向。因此,对于提高预测模型的性能至关重要。本文选取计算局部误差作为集成学习模型的组合指标,其中绝对误差被采用作为局部误差的衡量标准。绝对误差的计算简单明了,它代表了预测值与实际值之间的接近程度,因此较小的绝对误差意味着预测的准确性较高:

$$e = |y_i - \hat{y}_i| \quad (11)$$

式中: e 是绝对误差; y_i 是实际值; $\hat{y}_i$ 是模型预测值。

则模型 i 的权重 w_i 可以计算为:

$$w_i = \frac{1}{e_i} \quad (12)$$

基于此,得到集成模型的预测结果,计算公式为:

$$Y = \sum_{i=1}^N Y_i \quad (13)$$

式中: N 表示基模型的数量; Y_i 表示各个基模型的预测结果。

3 集成学习模型数据验证及分析

3.1 集成学习模型参数计算

研究数据来源于现场油田实际开发生产数据,根据对数据初步整理汇总分析可以发现,电潜泵的数据量不仅庞大,维度高,各个参数的相关性也很高;在发生故障的时候,各个参数都会发生一定的变化,由单一的参数作为标签参数建立故障预警的模型是不合适的。将电潜泵剩余使用寿命看作使用寿命的一次函数;电潜泵的监测为每 20 min 采一个点,已知每个泵的启泵与停泵的时间,即运行天数;通过这几个量就可以求出泵剩余使用寿命的数量变化规律。以 A 井为例,画出其故障前一个月的剩余使用寿命图,如图 1 所示,横坐标为每 20 min 一个的采样时间点。

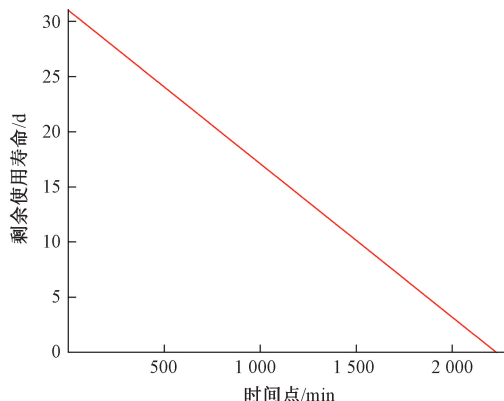


图 1 A 井电潜泵剩余使用寿命与使用寿命一次函数
Fig. 1 Remaining service life and linear function of service life of A well electric submersible pump

根据建立的预测模型,选取 A 井故障前 2 个月的数据输入随机森林贡献度算法模型中进行预测模型输入参数选择,得到结果如图 2 所示。根据贡献度大小选取功率、频率、进口压力和振动参数作为输入参数。

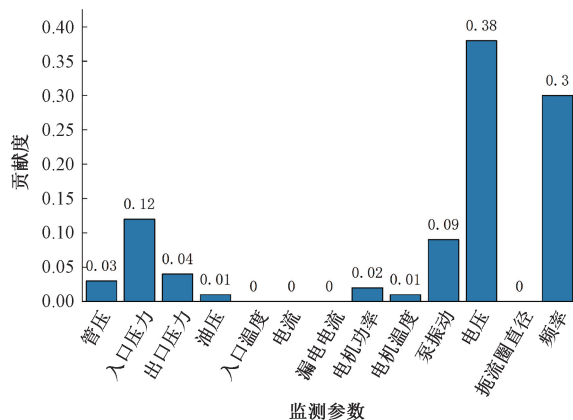


图 2 随机森林各参数导致故障贡献度计算结果
Fig. 2 Calculation results of fault contribution caused by various parameters of random forest

3.2 集成学习模型数据验证

根据 A 井故障前 25 d 的参数变化趋势,得到对比图如图 3 所示。可以观察到一系列关键参数呈现出了异常的变化;特别是泵进口压力参数发生了明显的直线上升,电压等其他参数出现了不同程度的异常波动。

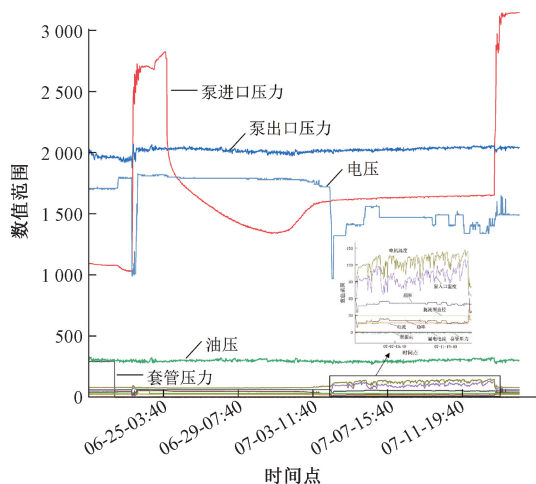


图3 A井电潜泵故障前25天各参数变化对比

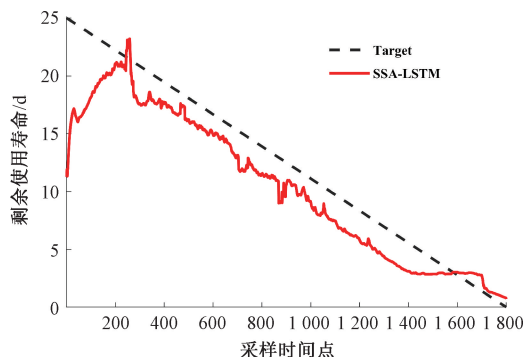
Fig. 3 Comparison chart of parameter changes in the 25 days before the failure of the electric submersible pump in Well A

将选定的参数首先输入两个基模型中,得到结果如图4所示。在整体表现上,SSA-LSTM算法的预测效果优于SSA-CNN算法;然而,在数据出现较大波动的前期阶段、故障停机前一个阶段,SSA-CNN算法的预测误差明显小于SSA-LSTM算法。在同样出现数据波动的后期,两种算法都发生了预测时间高于真实时间的情况,推测为由于后期数据变化的方式、数量与前期相似,算法学习过了前期的变化,以至于后期会出现高于实际情况的预测值。

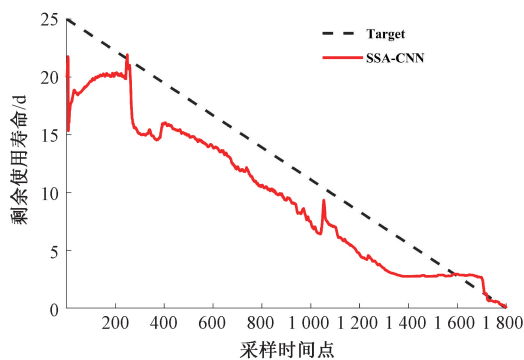
SSA-CNN和SSA-LSTM算法在不同情况下都有各自的优势和劣势。SSA-CNN在数据波动较大等阶段可能表现更优,如在第200~1000个采样点之间数据波动较大,SSA-CNN该时间区间的预测曲线相较于SSA-LSTM波动更为平缓稳定。但在整体预测中,SSA-LSTM结果更准确,为表述数据稳定性情况,采用数据间平均绝对误差(mean absolute error, MAE)指标进行评估,SSA-CNN、SSA-LSTM的预测数据与真实时间值的MAE分别为2.8928、2.3041,可知SSA-LSTM在整体预测中更为准确。

两种算法预测出的结果普遍偏低,尽管在某种程度上这种情况有利于提前进行维护工作;预测的剩余使用寿命较短时,可以提醒维修团队及时介入,减少潜在的生产损失。然而,电潜泵作为复杂的设备,其维护和更换过程同样需要相当高的成本投入。所以,预测结果应该力

求更接近真实值,通过更精准地预测剩余使用寿命,可以更好地规划维修计划,避免过早或过晚地进行维护,从而最大程度地降低维护成本,确保电潜泵的高效稳定运行。基于此,提出这两种算法的集成学习模型,以获得更好的预测性能。



(a) SSA-LSTM algorithm prediction result chart



(b) SSA-CNN algorithm prediction result chart

图4 两种基模型算法剩余寿命预测结果

Fig. 4 Remaining life prediction results of two basic model algorithms

将同样的数据代入到建立的集成模型中,得到的结果如图5所示。集成模型的预测误差明显小于两种基模型的预测误差,这表明集成模型能够有效地结合两种模型的优势,从而提高预测的准确性和稳定性。

具体来看,观察到在数据变化的前期阶段,首个采样时间点位置,集成模型预测的剩余使用寿命为17.8 d,SSA-CNN基模型预测的剩余使用寿命为20.1 d,SSA-LSTM基模型预测的剩余使用寿命为11.3 d,集成模型更倾向于赋予SSA-CNN模型更多的权重。在数据波动较大的情况下,SSA-CNN模型能够更准确地捕捉到数据的变化趋势,误差值比SSA-LSTM值小,因此集成模型更倾向于采用其预测结果,以降低整体的预测误差。

在预测中期阶段,第1000个采样时间点位置,集成模型预测的剩余使用寿命为8.5 d,SSA-CNN基模型预测的剩余使用寿命为7.4 d,SSA-LSTM基模型预测的剩余

使用寿命为 9.1 d,由于 SSA-LSTM 算法在整体表现上更为优秀,因此集成模型更多地采用了 SSA-LSTM 算法的预测结果,使预测结果更加平稳,并且更接近真实值。

在算法后期,即故障停机前的一段时间内,集成模型的预测结果与真实情况更为吻合;两个基模型中存在的高于真实值的部分,长度也变短了。整体来看,集成模型相较于两个基模型来说是更为准确的。

经实际算例对比验证表明,集成模型的预测精度提升了 6.41%,有效提高了预测的准确性和稳定性。

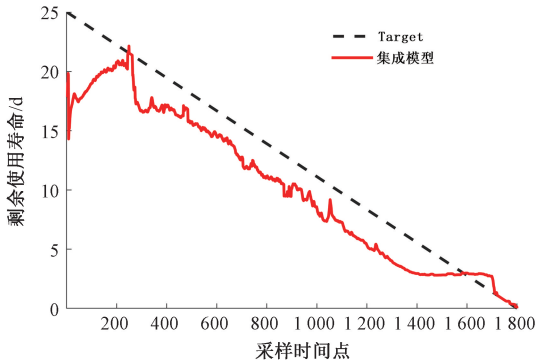


图 5 集成学习模型算法剩余寿命预测结果

Fig. 5 Remaining life prediction results of ensemble learning model algorithm

4 模型性能对比评估

4.1 集成学习模型性能对比评估

为了说明集成模型在电潜泵剩余寿命预测的有效与优势,经文献调研与探究,利用针对电潜泵剩余寿命预测较先进的预测模型 PSO-ELM、GA-BP 对剩余寿命进行预测并进行结果对比,情况如表 1 所示,由表 1 可以看出,集成模型的平均绝对百分比误差 (mean absolute percentage error, MAPE) 更低,从其他 3 个评价指标也可以得出,集成模型的预测结果更好。

表 1 模型评价指标值

Table 1 Model evaluation index value				
预测模型	MAPE	RMSE	MAE	R ²
集成模型	0.327 9	2.111 6	1.652 0	0.894 3
PSO-ELM 模型	0.525 0	2.721 5	1.861 0	0.858 0
GA-BP 模型	3.029 4	2.906 3	1.699 8	0.838 1

4.2 集成学习模型鲁棒性评估

在不同工况或数据噪声等干扰因素存在的情况下,模型的预测稳定性十分重要^[26],本文通过数据噪声测试以观测模型在干扰因素存在的情况下是否稳定,从而验证集成模型的鲁棒性。

首先,向输入数据添加不同幅度的高斯噪声(噪声标准差分别为数据均值的 1%、5%、10%),模拟数据采集过程中的随机扰动。然后,将噪声数据输入 SSA-LSTM、SSA-CNN 两个基模型以及集成模型,分别计算 3 种模型相较于原始数据的预测误差。结果得出,经 1%、5%、10% 3 种幅度对数据噪声化后,集成模型相较于原始数据预测误差分别为 0.51、1.85、1.69, CNN 基模型相较于原始数据预测误差分别为 1.20、2.31、2.31, LSTM 基模型相较于原始数据的预测误差分别为 0.92、2.51、1.82,可以看出集成模型在各级数据干扰条件下的误差变化幅度显著明显小于两个基模型,说明在干扰因素存在的情况下集成模型性能稳定,具有较好鲁棒性。

5 结 论

本文针对电潜泵现场运行历史数据,分析计算了每个时间点的剩余使用寿命作为预测标签,采用随机森林算法选择贡献度较高的参数作为预测模型输入特征;结合 SSA-CNN 和 SSA-LSTM 算法建立了集成学习模型;数据对比表明,集成算法兼具两者优势,预测精度提升了 6.41%,有效提高预测准确性和稳定性,整体精度和变化阶段预测准确率都有显著改善。为影响因素多、监测数据量庞大条件下的电潜泵剩余寿命预测的研究提供有效方案。此外,经过与现有较先进剩余寿命预测模型结果对比,得出集成模型的预测结果更好;经过鲁棒性验证,得出集成模型在干扰因素下仍能性能稳定,说明了集成模型在电潜泵剩余寿命预测的有效性。后续研究将实现电潜泵剩余寿命预测的现场应用验证,并针对其他复杂作业环境拓展本文方案的适用性研究,以实现更广泛的运用价值。

参考文献

[1] Khabibullin R A, Shabonas A R, Gurbatov N S, et al. Prediction of ESPs failure using ML at western siberia oilfields with large number of wells [C]. Proceedings of the SPE Russian Petroleum Technology Conference, 2020.

[2] Ghimire S, Deo R C, Casillas-Pérez D, et al. Probabilistic-based electricity demand forecasting with hybrid convolutional neural network-extreme learning machine model[J]. Engineering Applications of Artificial Intelligence, 2024, 132, DOI: 10.1016/j.engappai.2024.107918.

[3] 赵婧宇, 池越, 周亚同. 基于 SSA-LSTM 模型的短期电力负荷预测[J]. 电工电能新技术, 2022, 41(6): 71-79.

ZHAO J Y, CHI Y, ZHOU Y T. Short term power load

- forecasting based on SSA-LSTM model [J]. Advanced Technology of Electrical Engineering and Energy, 2022, 41(6): 71-79.
- [4] 韩腾飞. 基于集成学习的机械设备剩余使用寿命预测方法研究[D]. 南京: 南京林业大学, 2023.
- HAN T F. Research on remaining service life prediction method of mechanical equipment based on ensemble learning[D]. Nanjing: Nanjing Forestry University, 2023.
- [5] 赵岩. 基于集成学习的设备故障识别方法研究[D]. 秦皇岛: 燕山大学, 2023.
- ZHAO Y. Research on equipment fault identification method based on ensemble learning[D]. Qinghuangdao: Yanshan University, 2023.
- [6] 江兵, 杨春, 杨雨亭, 等. 基于 ACO 优化 BP 神经网络的变压器热点温度预测[J]. 电子测量与仪器学报, 2022, 36(10): 235-242.
- JIANG B, YANG CH, YANG Y T, et al. Transformer hotspot temperature prediction based on ACO optimized BP neural network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36 (10): 235-242.
- [7] 张新生, 贺凯璐. 基于 SSA-CNN 的长距离矿浆管道临界流速预测[J]. 安全与环境学报, 2022, 22(5): 2524-2531.
- ZHANG X SH, HE K L. Prediction of critical flow velocity in long-distance slurry pipelines based on SSA-CNN [J]. Journal of Safety and Environment, 2022, 22(5): 2524-2531.
- [8] 嵇晓燕, 杨凯, 陈亚男. 基于 ARIMA 和 Prophet 的水质预测集成学习模型[J]. 水资源保护, 2022, 38(6): 111-115.
- JI X Y, YANG K, CHEN Y N. Integrated learning model for water quality prediction based on ARIMA and Prophet[J]. Water Resources Protection, 2022, 38(6): 111-115.
- [9] 王朋飞, 刘长良, 徐健, 等. 基于图注意力和时间卷积网络的风电齿轮箱故障预警方法[J]. 电子测量与仪器学报, 2023, 37(8): 204-213.
- WANG P F, LIU CH L, XU J, et al. Fault warning method for wind turbine gearbox based on graph attention and time convolutional network[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(8): 204-213.
- [10] GAO W, GAO J, YANG L, et al. A novel modeling strategy of weighted mean temperature in China using RNN and LSTM[J]. Remote Sensing, 2021, 13(15): 3004.
- [11] 李文静, 王潇潇. 基于简化型 LSTM 神经网络的时间序列预测方法[J]. 北京工业大学学报, 2021, 47(5): 480-488.
- LI W J, WANG X X. Time series prediction method based on simplified LSTM neural network[J]. Journal of Beijing University of Technology, 2021, 47(5): 480-488.
- [12] CARPENTER C. Machine-Learning upscales realistic discrete fracture simulations [J]. Journal of Petroleum Technology, 2021, 73(11): 65-66.
- [13] 雷春丽, 夏奔锋, 薛林林, 等. 小样本下自校正卷积神经网络的滚动轴承故障识别方法[J]. 仪器仪表学报, 2022, 43(9): 122-130.
- LEI CH L, XIA B F, XUE L L, et al. Rolling bearing fault recognition method using self correcting convolutional neural network under small sample size [J]. Chinese Journal of Scientific Instrument, 2022, 43(9): 122-130.
- [14] 崔岩, 方春华, 文中, 等. 基于时频谱图和自适应动态权重 PSO-CNN 的外破振动信号识别[J]. 国外电子测量技术, 2023, 42(1): 144-152.
- CUI Y, FANG CH H, WEN ZH, et al. External vibration signal recognition based on time-frequency spectrum and adaptive dynamic weight PSO-CNN [J]. Foreign Electronic Measurement Technology, 2023, 42(1): 144-152.
- [15] 苏斌, 侯思祖, 郭威. 基于图像融合和双通道卷积神经网络的配电网故障选线方法研究[J]. 电子测量与仪器学报, 2024, 38(9): 54-66.
- SU B, HOU S Z, GUO W. Research on fault line selection method for distribution network based on image fusion and dual channel convolutional neural network[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 54-66.
- [16] 蒋富康, 陆金桂, 刘明昊, 等. 基于 CEEMDAN 和 CNN-LSTM 的滚动轴承故障诊断[J]. 电子测量技术, 2023, 46(5): 72-77.
- JIANG F K, LU J G, LIU M H, et al. Fault diagnosis of rolling bearings based on CEEMDAN and CNN-LSTM[J]. Electronic Measurement Technology, 2023, 46 (5): 72-77.
- [17] 张文帅, 王占刚. 基于改进麻雀算法优化深度极限学习机的缺失数据预测[J]. 电子测量技术, 2022, 45(15): 63-67.
- ZHANG W SH, WANG ZH G. Optimization of missing data prediction for deep extreme learning machine based on improved sparrow algorithm [J]. Electronic Measurement Technology, 2022, 45(15): 63-67.
- [18] LI F, YU Z, GAO J, et al. Brake discs surface defect detection using the IGD-IHT algorithm and the PIQEDS-ISSA-NESN algorithm [J]. Instrumentation, 2024,

- 11(3): 62-73.
- [19] 郭建帅, 崔双喜, 郭建斌, 等. 基于 VMD-SSA-HKELM 的超短期负荷预测[J]. 国外电子测量技术, 2022, 41(6): 105-111.
- GUO J SH, CUI SH X, GUO J B, et al. Ultra short term load forecasting based on VMD-SSA-HKELM[J]. Foreign Electronic Measurement Technology, 2022, 41(6): 105-111.
- [20] 杨鹏飞. 优化卷积神经网络诊断核辐射探测器电路故障[J]. 核电子学与探测技术, 2023, 43(4): 772-777.
- YANG P F. Optimizing convolutional neural networks for diagnosing circuit faults in nuclear radiation detectors[J]. Nuclear Electronics & Detection Technology, 2023, 43(4): 772-777.
- [21] HAO H T, WANG K. Fault detection and diagnosis of pneumatic control valve based on a hybrid deep learning model[J]. Instrumentation, 2023, 10(4): 12-26.
- [22] KATTERBAUER K, SHEHRI A A, QASIM A, et al. Hydrogen storage volume assessment and uncertainty quantification utilizing random forest ensemble learning [C]. Proceedings of the SPE Annual Technical Conference and Exhibition, 2023.
- [23] 胡倩, 陈红坤, 孙志达. 风电场短期风速的集成学习预测模型[J]. 电力系统及其自动化学报, 2016, 28(1): 1-4.
- HU Q, CHEN H K, SUN ZH D. Integrated learning prediction model for short-term wind speed in wind farms[J]. Proceedings of the CSU-EPSA, 2016, 28(1): 1-4.
- [24] 魏玮, 吕游, 齐欣宇, 等. 基于 CNN-LSTM-AM 动态集成模型的电站风机状态预测方法[J]. 仪器仪表学报, 2023, 44(4): 19-27.
- WEI W, LYU Y, QI X Y, et al. State prediction method for power plant wind turbines based on CNN-LSTM-AM dynamic integration model [J]. Chinese Journal of Scientific Instrument, 2023, 44(4): 19-27.
- [25] 仇佳佳. 基于混合误差准则的组合代理模型方法及其在有限元模型修正中的应用[D]. 镇江: 江苏大学, 2019.
- QIU J J. Combination surrogate model method based on mixed error criterion and its application in finite element model correction[D]. Zhenjiang: Jiangsu University, 2019.
- [26] 邹筱瑜, 胡亮, 王福利, 等. 基于信号分解深度网络的轴承剩余寿命预测[J]. 仪器仪表学报, 2024, 45(8): 45-57.
- ZOU X Y, HU L, WANG F L, et al. Remaining life prediction of bearings based on signal decomposition deep network[J]. Chinese Journal of Scientific Instrument, 2024, 45(8): 45-57.

作者简介



郑文培(通信作者), 2010 年于中国石油大学(北京)获得博士学位, 现为中国石油大学(北京)副教授, 主要研究方向为油气装备可靠性评估、结构损伤检测、智能识别算法。

E-mail: zhengwp@cup.edu.cn

Zheng Wenpei (Corresponding author) received his Ph. D. degree from China University of Petroleum (Beijing) in 2010. Now he is an associate professor at China University of Petroleum (Beijing). His main research interests include reliability assessment of oil and gas equipment, structural damage detection, and intelligent recognition algorithms.