

基于鹈鹕算法优化随机森林模型和健康指数的 变压器剩余寿命预测

温爱辉¹, 文 豹¹, 宋坤宇¹, 廖卫平¹, 徐 进², 高凌志^{2*}, 熊韵奇²

(1. 广东电网有限责任公司电力科学研究院, 广东 广州 510080;

2. 华南理工大学, 广东 广州 510641)

摘 要:针对变压器寿命预测模型拟合度低且忽略运行状态影响的问题,提出基于鹈鹕算法(POA)优化随机森林(RF)模型的剩余寿命预测模型(POA-RF 模型)。以预防性试验、在线监测、缺陷报告和台账数据为对象构建由变压器本体、油质和油中溶解气体健康指数等组成的变压器健康指数体系,考虑运行状态对剩余寿命的修正作用,以健康指数体系相关参量为输入,采用鹈鹕算法优化随机森林模型,解决其过拟合问题,提升预测准确性。以收集的变压器实例数据进行实验,验证模型的准确性。结果表明:对于测试样本中的 8 组报废样本,POA-RF 模型预测的剩余寿命平均绝对误差为 1.187 0 年,比 RF 模型的预测值和健康指数理论的计算值分别减小了 20.52% 和 49.02%;对于所有测试样本,和其他常用机器学习算法对比,POA-RF 模型的相关性能指标均为最优,优化效果明显,能够较好地预测变压器的剩余寿命。

关键词:变压器;寿命预测;健康指数;随机森林;鹈鹕算法

Transformer residual life prediction based on pelican algorithm optimized random forest model and health index

WEN Aihui¹, WEN Bao¹, SONG Kunyu¹, LIAO Weiping¹, XU Jin², GAO Lingzhi^{2*}, XIONG Yunqi²

(1. Electric Power Research Institute of Guangdong Power Grid Co., Ltd., Guangzhou 510080, China;

2. South China University of Technology, Guangzhou 510641, China)

Abstract: Aiming at the problem that the transformer life prediction model has low fit and ignores the influence of the operating state, a residual life prediction model (POA-RF model) based on pelican algorithm (POA) optimized random forest (RF) model was proposed. Taking preventive tests, online monitoring, defect reports, and account data as research objects, a transformer health index system consisting of transformer body, oil quality, and dissolved gas in oil health indices was constructed. Considering the corrective effect of operating status on residual life, with relevant parameters of the health index system as inputs, the POA was used to optimize the RF model to solve its overfitting problem and improve prediction accuracy. Experiments are conducted with the collected transformer instance data to verify the accuracy of the model. The results show that for the 8 groups of scrapped samples in the test samples, the average absolute error of the residual life predicted by the POA-RF model is 1.187 0 years, which is 20.52% and 49.02% lower than the predicted value of the RF model and the calculated value of the health index theory, respectively. For all the test samples, compared with other commonly used machine learning algorithms, the relevant performance indicators of the POA-RF model are all the best, and the optimization effect is obvious, which can better predict the residual life of transformer.

Key words: transformer; life prediction; health index; random forest; pelican algorithm

0 引言

变压器具有转换电压、输送电能的作用,其安全稳定运行具有重要意义^[1-3]。据国家能源局发布的 2023 年度全国电力可靠性指标,我国投运时间在

10~20 年的变压器数量最多,达到 11 031 台。随着投运年限的增加,变压器的可用系数呈现下降趋势。因此,对变压器开展健康评估和剩余寿命预测,对运维人员合理安排设备养护措施及电网的安全运行有重要意义^[4-7]。

目前,国内外学者大多围绕变压器健康指数体系进行剩余寿命预测研究。文献[8]建立了变压器

基金项目:南方电网公司科技项目[030000KC23110056(GD KJXM20231264)]。

健康状态三级评估模型,推导出剩余寿命和健康指数计算公式,并通过 500 kV 变电站的变压器进行准确性验证。文献[9]通过变压器主体、绝缘纸、油中溶解碳氢和油质的健康指数关系计算得到最终变压器的健康指数。文献[10]从变压器的绕组温度出发,计算考虑损耗后的变压器绝缘剩余寿命,通过变压器健康指数分级体系得到最终健康指数。当数据量较大时,传统计算方法表现出计算量大、应用较为困难等问题,且较少考虑变压器的运行状态对剩余寿命的影响。

近年来,针对上述问题,人工智能算法和变压器健康指数体系融合方案被广泛用于变压器剩余寿命预测中。文献[11]以变压器油温、电压和绕组温度为输入量,构建了一种基于深度置信网络(DBN)与健康指数结合的寿命预测模型。文献[12]使用熵权法对变压器健康指数计算公式进行了改进。文献[13]以绝缘油界面张力、油中糠醛含量和含水量为输入量,采用自适应模糊神经网络模型输出剩余寿命。文献[14]考虑了运行年限、糠醛含量、一氧化碳和二氧化碳含量、负荷系数和环境系数等与变压器寿命相关的参数,以混沌序列优化BP神经网络并对变压器剩余寿命进行预测。以上大多数寿命预测模型存在拟合程度不高和泛化能力较差等问题,且考虑的寿命相关参数相对片面。

本文通过构建变压器健康指数体系获得变压器综合健康指数,然后将变压器运行状态和缺陷情况以调节因子形式反馈到变压器的最终健康指数中,随后建立鹈鹕算法(POA)优化随机森林(RF)模型的剩余寿命预测模型(POA-RF模型),解决随机森林算法存在的参数寻优问题,并通过收集到的变压器实例数据进行测试,对比其他常用的预测算法,验证所提寿命预测模型的有效性。

1 健康指数与剩余寿命

1.1 变压器最终健康指数

健康指数是基于专家知识和行业标准对变压器的健康状况进行综合评价的一种方法^[15-18],主要是基于与变压器绝缘材料寿命相关的各种特征参量,对每个特征参量进行量化评估,加权反馈到健康指数中。评估变压器健康状况的相关参数众多,电网运维人员主要通过4类基础途径获取变压器状态参数:预防性试验(基于DL/T 596—2021《电力设备预防性试验规程》标准等,周期性获取的油色谱

和油质试验参数)、在线监测(依托广泛部署的标准化监测装置,实时获取油温、绕组温度和负载率等数据)、缺陷报告(常规运维流程记录的异常现象)及台账数据(基础性数据、记录铭牌等参数信息)。

综合以上分析,本文选择预防性试验报告、在线监测报告、缺陷报告和台账数据作为变压器剩余寿命评估的信息来源。通过设备预防性试验得到油色谱数据和油质试验数据,通过在线监测获取变压器负载率、油温和绕组温度数据,通过缺陷报告统计缺陷次数等,通过台账数据获取变压器的出厂时间、投运时间和运行环境参数等,由此计算变压器主体健康指数 HI_M 、油质健康指数 HI_Q 、油中溶解气体健康指数 HI_D 以及运行状态和缺陷调节系数。变压器最终健康指数(HI)的计算公式如式(1)所示^[6]。

$$HI = K \times HI_{com} \quad (1)$$

式(1)中: HI_{com} 为综合健康指数; K 为运行状态和缺陷调节系数。

1.2 变压器主体健康指数

变压器主体健康指数反映了变压器的整体健康状况,参考英国EA公司提出的反映电力设备健康状况随时间变化的健康指数公式,得到变压器主体健康指数(HI_M)计算公式^[8-9]如式(2)所示。

$$HI_M = HI_0 \times e^{B(T_2 - T_1)} \quad (2)$$

式(2)中: HI_0 为变压器初始健康指数,一般取值为0.5; B 为老化系数; T_1 为投运年份; T_2 为当前运行年份或未来年份。

健康指数的高低反映了变压器本体运行状况的好坏以及可能的预期寿命。健康指数为0~3.0,说明变压器的个别部件轻微老化,状态优秀;健康指数为3.0~4.0,说明变压器的小部分部件老化,状态良好;健康指数为4.0~5.0,说明变压器的部分部件严重老化,状态一般;健康指数为5.0~6.5,说明变压器整体严重老化,状态极差,走向寿命终点。

由式(2)得到老化系数 B 的计算公式为式(3)。

$$B = \frac{\ln HI_M - \ln HI_0}{T_2 - T_1} \quad (3)$$

设 $T_{exp} = T_2 - T_1$, T_{exp} 为考虑变压器绝缘在负荷条件下逐渐老化的预期剩余寿命,即当变压器再使用 T_{exp} 后,整体老化严重,状态极差,面临报废,此时,变压器的健康指数 HI 等于6.5,则可得到式(4)。

$$B = \frac{\ln 6.5 - \ln 0.5}{T_{exp}} \quad (4)$$

变压器预期剩余寿命 T_{exp} 和其负荷情况、运行环境、投运时间以及运行时的油温和绕组温度有关。因此,通过负荷系数(f_L)、环境系数(f_E)、储存系数(f_S)、变压器平均油温系数(f_O)和绕组温度系数(f_W)对厂家设定的变压器预期剩余寿命进行修正,即可得到修正后的预期剩余寿命 T'_{exp} ^[16],如式(5)所示。

$$T'_{\text{exp}} = \frac{T_{\text{exp}}}{f_L f_E f_S f_O f_W} \tag{5}$$

负荷系数反映了变压器实际使用情况和负荷能力的关系,一般来说,变压器的负荷率标准范围应该为50%~80%,负荷系数和负荷率的关系如表1所示^[8]。

表1 变压器负荷系数

Table 1 Transformer load factor

变压器平均负荷率/%	负荷系数 f_L
0~40	1.00
40~60	1.05
60~70	1.10
70~80	1.25
80~150	1.60

温度、湿度和海拔高度等均会不同程度地影响变压器的绝缘老化程度。为简化环境系数的计算,本文考虑以变压器的安装位置和所处位置的最高温度为依据确定环境系数,如表2所示^[18]。由于变压器出厂后并不一定立刻投入使用,而是可能作为备用变压器进行储存,此时变压器的绝缘状态可能会出现一定老化。因此,需要考虑储存时间对变压器预期剩余寿命的影响。变压器的储存系数见表3^[18]。

表2 变压器环境系数

Table 2 Transformer environmental factor

变压器安装位置及最高温度	环境系数 f_E
室内	0.95
室外,年最高温度≤40℃	1.00
室外,年最高温度>40℃	1.10

表3 变压器储存系数

Table 3 Transformer storage factor

变压器出厂至投运时间间隔/d	储存系数 f_S
0~365	1.00
365~730	1.02
730~1460	1.05
>1460	1.10

温度对变压器绝缘老化有较大影响,根据蒙辛格规则推导出的热老化6℃法则,在80~140℃的温度范围内,温度每升高6℃,绝缘寿命减少一半。本研究参考文献[16]得到变压器的平均油温和平均绕组温度系数分别见表4和表5。

表4 变压器平均油温系数

Table 4 Transformer average oil temperature factor

变压器平均油温/℃	变压器平均油温系数 f_O
0~45	0.90
45~60	0.95
60~75	1.00
75~160	1.10

表5 变压器平均绕组温度系数

Table 5 Transformer average winding temperature factor

变压器平均绕组温度/℃	变压器平均绕组温度系数 f_W
0~45	0.90
45~70	0.95
70~80	1.00
80~160	1.10

1.3 油质健康指数

油质健康指数主要针对绝缘油的微水含量、击穿电压、介质损耗因数($\tan\delta$)和体积电阻率等参数进行评估,计算公式如式6所示^[8-9]。参考文献[8,18],相关评分和系数分别见表6和表7。

$$HI_Q = \sum_{i=4}^4 S_i \times W_i \tag{6}$$

式(6)中: HI_Q 是油质健康指数; i 是各油质参数的编号; S_i 是各油质参数对应的评分; W_i 是各油质参数对应的权重。

表6 油质参数评分表

Table 6 Scoring table for oil quality parameters

评分	0分	2分	6分	8分	10分
微水含量/(mg/L)	<5	5~10	10~15	>15	—
介质损耗因数/%	<0.5	0.5~1.0	1.0~1.5	1.5~2.0	>2.0
击穿电压/kV	>50	40~50	30~40	<30	—
体积电阻率/(×10 ¹⁰ Ω·m)	>0.5	0.3~0.5	0.2~0.3	0.1~0.2	<0.1

表7 油质参数权重表

Table 7 Weighting table for oil quality parameters

油质参数	权重	权重符号
微水含量	0.219 1	W_1
介质损耗因数	0.342 5	W_2
击穿电压	0.219 1	W_3
体积电阻率	0.219 1	W_4

1.4 油中溶解气体健康指数

通过分析油中溶解气体的组分和含量可以对变压器的运行状态进行有效判断。将 H_2 、 CH_4 、 C_2H_4 、 C_2H_6 和 C_2H_2 5 种油中溶解气体的评分和对应权重相乘并累加即可得到油中溶解气体健康指数，如式(7)所示。每种气体的体积分数有相应的评分，具体评分和系数情况分别见表 8 和表 9^[8]。

$$HI_D = \sum_{i=1}^5 S_i \times W_i$$

(7)

式(7)中： HI_D 是油中溶解气体健康指数； i 是各油中溶解气体的编号； S_i 是各油中溶解气体对应的评分； W_i 是各油中溶解气体对应的权重。

表 8 油中溶解气体评分表

Table 8 Scoring table for dissolved gases in oil

评分	0分	2分	6分	8分	10分	16分
$c_{H_2}/(\times 10^{-6})$	<20	20~40	40~100	—	100~200	>200
$c_{CH_4}/(\times 10^{-6})$	<10	10~20	20~40	40~65	65~150	>150
$c_{C_2H_6}/(\times 10^{-6})$	<10	10~20	20~40	40~65	65~150	>150
$c_{C_2H_4}/(\times 10^{-6})$	<10	10~20	20~40	40~65	65~150	>150
$c_{C_2H_2}/(\times 10^{-6})$	<0.5	0.5~1	1~3	3~5	>5	—

表 9 油中溶解气体权重表

Table 9 Weighting table for dissolved gases in oil

气体	权重	权重符号
H_2	0.192 3	W_{H_2}
CH_4	0.115 4	W_{CH_4}
C_2H_6	0.115 4	$W_{C_2H_6}$
C_2H_4	0.115 4	$W_{C_2H_4}$
C_2H_2	0.461 5	$W_{C_2H_2}$

1.5 综合健康指数

基于以上提出的变压器主体健康指数 HI_M 、油质健康指数 HI_Q 和油中溶解气体健康指数 HI_D ，可得到变压器综合健康指数 HI_{com} ，其具体计算方法如式(8)所示^[16]。

$$HI_{com} = HI_M \times f_{com}$$

(8)

当 HI_M 最大时，说明变压器实际运行时间较长，此时分两种情况计算 HI_{com} ，当油质健康指数 HI_Q 和油中溶解气体健康指数 HI_D 均小于 1 时， f_{com} 取值为 1，否则需要结合油质健康指数 HI_Q 和油中溶解气体健康指数 HI_D 的关系来确定 f_{com} 的取值。 f_{com} 的取值规则如表 10 所示。

当 HI_M 不是最大值时，说明变压器的实际运行时间较短，此时则通过式(9)计算变压器的最终健康指数。同样由油质健康指数 HI_Q 和油中溶解气体

表 10 HI_M 最大时 f_{com} 的取值规则

Table 10 f_{com} value rule when HI_M is maximum

组合关系	f_{com}
$\max(HI_Q, HI_D) \leq 2$	1.00
$2 < \max(HI_Q, HI_D) \leq 4$	1.07
$\max(HI_Q, HI_D) > 4$	1.15

健康指数 HI_D 的相对大小关系来确定 f_{com} 的取值，其取值规则如表 11 所示。

$$HI_{com} = \max(HI_M, HI_Q, HI_D) \times f_{com}$$

(9)

表 11 HI_M 非最大时 f_{com} 的取值规则

Table 11 f_{com} value rule when HI_M is not maximum

组合关系	f_{com}
$HI_Q/HI_D \leq 0.3$	1.00
$0.3 < HI_Q/HI_D \leq 0.6$	1.07
$HI_Q/HI_D > 0.6$	1.15

1.6 运行状态和缺陷调节系数

变压器当前的运行状态和缺陷情况也是影响其剩余寿命的重要因素。若变压器处于故障状态，则其绝缘会快速劣化，遇到突发的严重性故障甚至会立刻进入退役报废状态。因此，将变压器运行状态也作为预测剩余寿命的 1 个输入参数，可起到对健康指数的修正作用，运行状态和缺陷调节系数 K 由式(10)计算。参考 DL/T 1685—2017《油浸式变压器(电抗器)状态评价导则》、IEEE 相关标准以及一些研究文献^[10,19-20]，使用模糊综合评判法给出变压器的运行状态调节系数 K_1 和缺陷调节系数 K_2 ，分别如表 12 和表 13 所示。

$$K = K_1 \times K_2$$

(10)

表 12 运行状态调节系数

Table 12 Operating state adjustment factors

故障类型	运行状态调节系数 K_1
正常	0.95
低温过热故障	1.10
局部放电故障	1.20
低能放电故障	1.40
中温过热故障	1.50
高能放电故障	1.70
高温过热故障	1.90

1.7 剩余寿命计算

得到变压器最终健康指数后，变压器的理想剩余寿命 RUL_1 可以通过式(11)推算得到。

表 13 缺陷调节系数

Table 13 Defect adjustment factors

近五年缺陷次数/次	缺陷调节系数 K_2
0	0.95
1	1.00
2~4	1.05
5~10	1.20
>10	1.50

$$RUL_1 = \frac{\ln(HI^*) - \ln(HI)}{B} \quad (11)$$

式(11)中: B 为老化系数,由式(4)~(5)计算; HI^* 为变压器的报废临界健康指数。一般认为,当健康指数大于6时,变压器的故障率会大幅提高,接近退役报废状态。考虑到一定的裕度,将变压器的报废临界健康指数设为7,得到最终 RUL 的计算公式如式(12)所示。

$$RUL = \frac{\ln(7) - \ln(HI)}{B} \quad (12)$$

2 剩余寿命预测模型

针对变压器剩余寿命预测问题,本文提出了一种利用鹌鹑算法优化随机森林模型的变压器剩余寿命预测模型,以达到利用变压器健康指数体系进行剩余寿命预测的目的。

2.1 随机森林算法

随机森林算法是一种包含多个决策树的集成学习算法,通过随机选择样本数据的子集构建树木,每个决策树之间相互独立,最终结果由投票机制统计所有决策树的结果进行均值输出^[21-22]。由于随机森林算法的采样方法为 Bootstrap 法,可以有效提取出数据集中的数据特征,使过拟合风险得到有效降低,提高模型的泛化能力,适用于求解各种非线性问题,因此适用于运行状况复杂的油浸式变压器剩余寿命预测。随机森林的原理如图1所示。

构建随机森林模型的一般步骤如下^[22]:

(1)采用 Bootstrap 抽样法从原始数据集中随机抽取 M 个样本数相同的数据集,其中每个数据集都含有 N 个特征;

(2)对上述每个数据集均构造 1 棵 CART 决策树。在构造子树的过程中,并不将所有特征用作节点分裂的选择,而是随机选择 K 个特征维度,随后采用 CART 算法分裂而不剪枝;

(3)重复前两步,让决策树充分生长;

(4)建立随机森林后,利用测试样本进入每棵

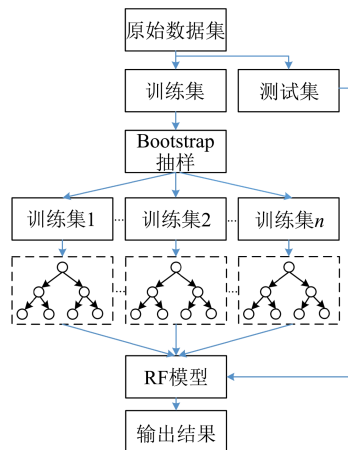


图1 随机森林算法原理

Fig.1 Principle of random forest algorithm

决策树,进行回归输出,并以投票方式输出最终预测值。

2.2 鹌鹑算法(POA)

鹌鹑算法(POA)是由 P TROJOVSKÝ 等^[23]于2022年所提出的一种新型的种群优化算法,其设计思想主要是模拟鹌鹑狩猎过程中的自然行为。将猎物的位置作为算法的最优解,鹌鹑狩猎的行为即算法的寻优过程。算法包含种群初始化阶段、探索阶段和开发阶段,后两个阶段分别对应鹌鹑狩猎时的逼近猎物和水面飞行捕猎这两种行为^[24-25]。不同阶段鹌鹑优化算法的数学模型如式(13)~(15)所示。

(1)种群初始化阶段,鹌鹑优化算法的数学模型计算表达式为式(13)。

$$x_{i,j} = l_j + \text{rand} \cdot (u_j - l_j) \quad (13)$$

式(13)中: $x_{i,j}$ 为第 i 个鹌鹑的第 j 个变量的值; u_j 、 l_j 分别为第 j 个变量的上、下限; rand 为 $[0,1]$ 内的随机数; i 为鹌鹑种群数,即 $i=1, 2, \dots, N$; j 为待求解问题的个数,即 $j=1, 2, \dots, m$ 。

(2)探索阶段,鹌鹑需要确定猎物的位置,并向目标区域移动,此时对应的数学表达式为式(14)。

$$x_{i,j}^{p_1} = \begin{cases} x_{i,j} + \text{rand} \cdot (p_j - I \cdot x_{i,j}), & F_p < F_i \\ x_{i,j} + \text{rand} \cdot (x_{i,j} - p_j), & \text{else} \end{cases} \quad (14)$$

式(14)中: $x_{i,j}^{p_1}$ 为第 i 个鹌鹑在第 j 维度的新位置; p_j 为猎物在第 j 维度上的位置; I 对于每次迭代和每个种群成员来说是随机选择的,其值为1或者2; F_p 和 F_i 分别为猎物、第 i 个鹌鹑的目标函数值。

若第 i 个鹌鹑在本阶段的目标函数值有更优结果,则更新其位置,用 $x_{i,j}^{p_1}$ 替换 $x_{i,j}$, 否则 $x_{i,j}$ 不变。

(3)开发阶段,鹈鹕水面飞行并捕获猎物,对鹈鹕捕猎的行为进行建模,此时其计算表达式为式(15)。

$$x_{i,j}^{p_i} = x_{i,j} + R \cdot \left(1 - \frac{t}{T}\right) \cdot (2 \cdot \text{rand} - 1) \cdot x_{i,j} \quad (15)$$

式(15)中: $x_{i,j}^{p_i}$ 为该阶段第*i*个鹈鹕在第*j*维度上更新后的位置; R 为常数,其值为0.2;系数 $R(1-t/T)$ 为种群成员的邻域半径; t 为当前迭代次数; T 为最大迭代次数。

和探索阶段类似,鹈鹕位置的更新遵循目标函数值是否达到更优。

2.3 基于鹈鹕算法优化随机森林模型的变压器剩余寿命预测模型

在随机森林模型的基础上加入鹈鹕算法对随机森林的决策树数目(tree_num)和最小叶子节点数(MinLeafSize)进行优化,使模型发挥最佳性能,避免模型发生过拟合等不稳定情况。优化后的模型流程如图2所示。

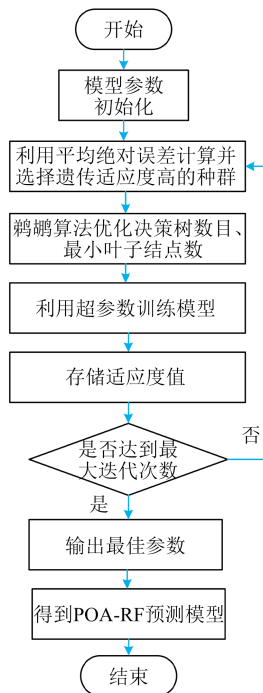


图2 鹈鹕算法优化随机森林模型流程

Fig.2 Flow of the pelican algorithm optimized the random forest model

(1)模型参数初始化,设置鹈鹕优化算法(POA)的种群数量*N*、最大迭代次数*T*、搜索空间的上下限、决策树的数目(tree_num)与最小叶子节点数(MinLeafSize);

(2)选定验证集经标准化后的平均绝对误差计算适应度函数并选择遗传适应度高的种群;

(3)采用鹈鹕算法(POA)开始优化随机森林模型的决策树的数目(tree_num)与最小叶子节点数(MinLeafSize);

(4)利用超参数训练模型并存储适应度值;

(5)判断算法的迭代次数是否达到设定的最大迭代次数,如果满足,则停止迭代,输出最佳参数,否则继续执行步骤(2)直至达到最大迭代次数为止。

基于油浸式变压器的健康指数体系,输入模型的变压器剩余寿命特征参量如表14所示。对于运行状态的特征参量输入,将6种基本故障类型和正常状态以序号1~7进行编码,即1-低温过热故障,2-中温过热故障,3-高温过热故障,4-局部放电故障,5-低能放电故障,6-高能放电故障,7-正常。

表14 变压器剩余寿命特征参量

Table 14 Characterizing parameters for the residual life of transformer

编号	特征量	编号	特征量
1	储存时间	10	$\tan\delta$
2	在运时间	11	体积电阻率
3	平均负载率	12	H_2 体积分数
4	平均油温	13	CH_4 体积分数
5	平均绕组温度	14	C_2H_6 体积分数
6	运行环境	15	C_2H_4 体积分数
7	年最高温度	16	C_2H_2 体积分数
8	微水含量	17	缺陷次数
9	击穿电压	18	故障类型

由于不同变压器特征参量的数值之间差异较大,会对模型的收敛速度和预测能力产生影响,因此使用式(16)将变压器的剩余寿命特征参量进行Z-Score标准化处理。

$$x^* = \frac{x - \bar{x}}{\sigma} \quad (16)$$

式(16)中: x^* 为经过标准化后的特征参量; x 为特征参量原始值; $\bar{x}$ 为特征参量的平均值; σ 为特征参量的标准差。

在训练好的模型对测试集完成预测输出后,需要对模型的效果进行评估。本文选择使用平均绝对误差 E_{MAE} 、均方误差 E_{MSE} 和均方根误差 E_{RMSE} 作为模型的评估参数,其计算公式如式(17)~(19)所示。

$$E_{\text{MAE}} = \frac{1}{n} \sum_i^n |y_i^* - y_i| \cdot 100\% \quad (17)$$

$$E_{\text{MSE}} = \frac{1}{n} \sum (y_i - y_i^*)^2 \tag{18}$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_i^n |y_i^* - y_i|^2} \cdot 100\% \tag{19}$$

式(17)~(19)中: n 为测试集样本数量; y_i 为剩余寿命计算值; y_i^* 为剩余寿命预测值。

3 算例分析

3.1 模型参数设置

为验证本文提出的基于 POA-RF 和健康指数的油浸式变压器剩余寿命预测模型的有效性,以变压器健康指数体系涉及的特征参量为输入量,基于 Matlab 平台对南方电网某省 110 kV 和 220 kV、运行环境覆盖不同气候和地理环境以及涉及不同生产厂家的主变压器的 368 组实测数据进行建模分析。其中 80 组数据是报废变压器的实测数据,是在变压器报废停止使用之前试验采集得到的。368 组数据按 8:1:1 的比例划分为训练集、验证集和测试集,其中训练集用于预测模型的训练,以找到模型最优参数;测试集和验证集样本数均为 37 组,测试集中包含有 8 组报废变压器数据与 29 组在运变压器数据,用来检验模型的预测能力和应用效果。

为了以尽可能少的总优化次数达到最好的优化效果,本文优化算法的种群数设置为 4,种群进化数设置为 25。随机森林模型的两个优化参数——决策树数目 `tree_num` 与最小叶子节点数 `MinLeafSize` 的初始值分别设置为 100 和 1。

3.2 模型预测验证

由于用于 Matlab 建模分析的 368 组主变压器的剩余寿命是基于油浸式变压器健康指数理论计算得到的,本文对测试集中的 8 组报废变压器数据进行分析,以验证 POA-RF 模型的预测能力。8 组报废样本的实际剩余寿命、健康指数理论计算的剩余寿命以及各模型的预测剩余寿命如表 15 所示,图 3 为 8 组报废样本剩余寿命误差情况分布。

由表 15 和图 3 可以看出,对于测试集中的 8 组报废样本,油浸式变压器健康指数理论计算得到的剩余寿命平均绝对误差为 2.328 5 年,最大绝对误差为 3.593 7 年,最小绝对误差为 0.460 3 年。对于实际剩余寿命越接近 0 的报废样本,健康指数计算值的误差值也越小。究其原因,可能是临近报废的变压器样本通常运行时间较长,其本体健康指数 HI_M 显著高于油质健康指数 HI_Q 和油中溶解气体健康指

表 15 8 组报废样本剩余寿命预测值
Table 15 Residual life prediction value for 8 groups of end-of-life samples

样本编号	实际值 /年	健康指数理论计算值/年	RF 模型预测值/年	POA-RF 模型预测值/年
1	5.112 3	1.518 6	4.659 3	3.962 7
2	3.704 1	2.240 0	5.205 6	2.810 2
3	5.082 2	2.114 1	8.077 8	4.955 1
4	0.460 3	0	2.733 9	2.296 8
5	2.956 2	0	1.317 4	1.357 9
6	2.353 4	0	3.511 3	2.121 9
7	1.934 2	0	3.791 3	4.135 6
8	3.295 9	0.398 2	3.365 3	1.838 4
平均绝对误差	—	2.328 5	1.493 4	1.187 0

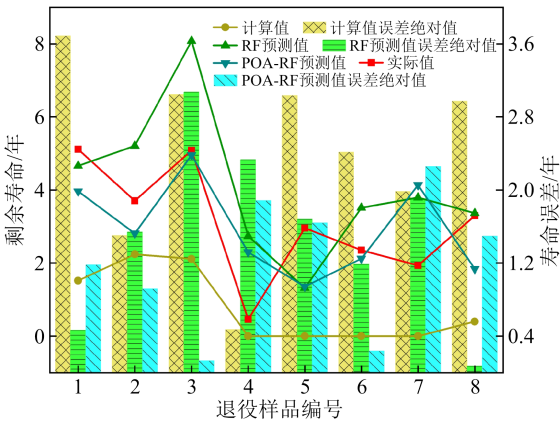


图 3 8 组报废样本剩余寿命误差情况分布图

Fig.3 The error distributions of residual life for 8 groups of end-of-life samples

数 HI_D ,在此类样本中,设备的最终健康指数主要由本体健康指数 HI_M 决定。在本文剩余寿命计算公式中,报废临界健康指数设定为 7。然而,实际运行中,不同变压器的运维条件存在差异,导致其真实的报废临界健康指数发生波动。因此,基于健康指数理论计算的剩余寿命与实际值之间会产生偏差,且越接近报废的变压器样本预测误差越小。相比于 RF 模型的预测值和健康指数理论的计算值,POA-RF 模型输出的剩余寿命预测值最大绝对误差和最小绝对误差均明显更小,其平均绝对误差为 1.187 0 年,较 RF 模型和计算值分别减小了 20.52% 和 49.02%,表明 POA-RF 模型有效减小了健康指数理论计算剩余寿命的误差,能较好地对变压器剩余寿命进行预测。

此外,对 POA-RF 模型预测误差较大的样本进行分析,发现样本 7 的预测误差绝对值最大,超过了

2年,这可能是由于鸚鵡算法的优化目标为验证集经标准化后的平均绝对误差,侧重于提升模型的整体预测性能,然而该优化目标可能导致模型对个别特殊样本(如样本7)的预测精度不足。

选取一台报废样本进行具体分析:某变电站一台型号为SFZ9-40000/110的报废主变压器,出厂时间为1996年10月1日,投运时间为1997年5月8日,实际报废时间为2024年5月1日。设备运行环境为户外,于2021年1月份测得油中微水含量为3.5 mg/L,击穿电压为75.5 kV,油介质损耗因数为0.010 3,体积电阻率为 $572\times10^{10}\ \Omega\cdot\text{m}$,油色谱实验测得 H_2 、 CH_4 、 C_2H_4 、 C_2H_6 、 C_2H_2 的体积分数分别为 8.03×10^{-6} 、 4.98×10^{-6} 、 1.64×10^{-6} 、 18.86×10^{-6} 、0,其近5年出现缺陷次数为2次,运行状态正常,平均负载率为46.66%,平均油温为47.52℃,平均绕组温度为45.81℃。将上述指标作为模型输入量,经POA-RF模型输出该变压器的剩余寿命预测值为1.838 4年,而实际剩余寿命为3.295 9年。模型对该台变压器的预测误差为1.457 5年,证明了本文所提POA-RF模型预测变压器剩余寿命的准确性。

3.3 模型预测应用

在验证了模型的剩余寿命预测能力后,对模型在所有测试样本上的应用效果进行研究,图4和图5分别为基于RF模型和POA-RF模型的测试集样本的剩余寿命预测结果和相关评价参数。

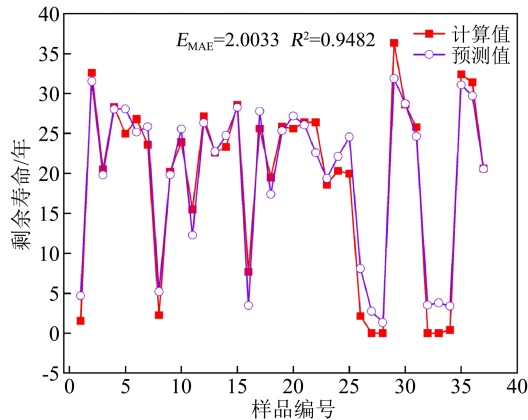


图4 测试样本RF模型预测结果

Fig.4 Prediction results of test samples based on RF model

分析图4和图5可知,POA优化的RF模型对测试集样本的平均绝对误差 E_{MAE} 和决定系数 R^2 分别为1.690 5和0.961 4,平均绝对误差 E_{MAE} 比优化前降低了15.61%,决定系数 R^2 比优化前提高了1.39%,表明POA-RF模型对变压器的剩余寿命预测效果明

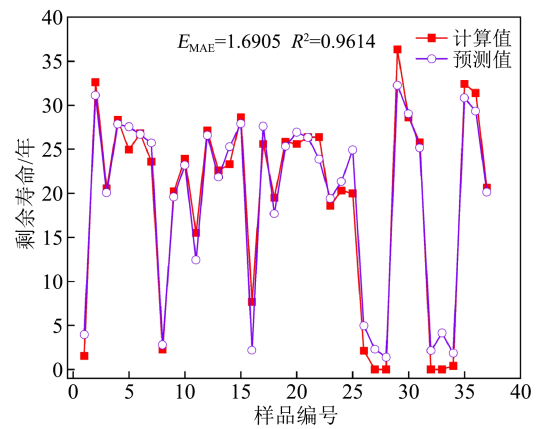


图5 测试样本POA-RF模型预测结果

Fig.5 Prediction results of test samples based on POA-RF model

显优于RF模型,准确性得到提高。

为了验证POA算法的优化性能,本文选择常用的粒子群算法(PSO)、模拟退火算法(SA)、灰狼优化算法(GWO)和麻雀搜索算法(SSA)等优化算法进行参数寻优,图6为所有优化算法的适应度值曲线。

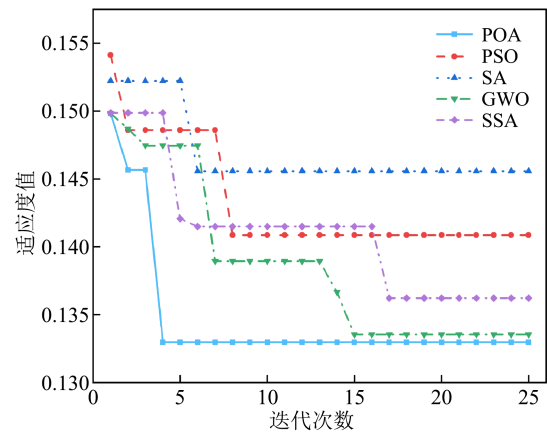


图6 适应度值对比曲线

Fig. 6 Comparison curves of adaptation value

由图6可知,PSO算法在第8次迭代就达到收敛,适应度值为0.140 87;SA算法的收敛速度比PSO算法更快,在第6次迭代即收敛于0.145 58;而在所有算法中,GWO算法的收敛速度较慢,在第15次迭代时才收敛,但是其适应度值低于PSO、SA和SSA算法;SSA算法则在第17次迭代时收敛于0.136 22。POA算法在第2次迭代时适应度值就下降到0.145 67,随后在第4次迭代时,适应度值再次迅速下降并稳定至0.132 97,表明POA算法在逼近最优解的速度和收敛精度上均明显优于其他优化算法,具有更好的局部参数寻优能力。

为进一步检验模型的预测能力,本文将随机森林(RF)模型、决策树(DT)、支持向量机(SVM)、POA优化的支持向量机(POA-SVM)、POA优化的决策树(POA-DT)、粒子群算法优化的随机森林(PSO-RF)模型、模拟退火算法优化的随机森林(SA-RF)模型、灰狼算法优化随机森林(GWO-RF)模型、麻雀搜索算法优化的随机森林(SSA-RF)模型和所提出的POA-RF模型进行性能对比。其中,优化算法优化SVM的参数为核尺度(KernelScale)、正则化参数(c)以及不敏感参数(ε);优化算法优化DT的参数为最小叶子节点数(MinLeafSize)。两者的初始值设置如表16所示。粒子群(PSO)、模拟退火(SA)、灰狼算法(GWO)和麻雀搜索(SSA)等优化算法的具体参数设置如表17所示。

表16 SVM和DT的参数设置

Table 16 Parameter settings for SVM and DT

模型	参数
SVM	KernelScale=1.00, $c=0.8$, $\varepsilon=0.080$
DT	MinLeafSize=7

表17 优化算法 PSO、SA、GWO 和 SSA 的参数设置

Table 17 Parameter settings for optimization algorithms of PSO, SA, GWO, and SSA

优化算法	参数
PSO	$V_{\max}=6$; $w_{\max}=0.9$; $w_{\min}=0.6$; $c_1=2$; $c_2=2$;
SA	优化变化容忍度 $T=10\times10^{-10}$
GWO	—
SSA	生产者比例 $P_{\text{percent}}=0.2$, 危险麻雀比例为0.2, 生产者更新策略切换阈值为0.8

上述所有模型均基于收集到的368组变压器样本数据进行分析,表18给出了各模型在测试集上的平均绝对误差 E_{MAE} 、均方误差 E_{MSE} 和均方根误差 E_{RMSE} 。由表18可知,相较于其他模型,POA-RF模型的 E_{MAE} 、 E_{MSE} 、 E_{RMSE} 均最低。此外随机森林(RF)模型的预测精度明显高于支持向量机(SVM)和决策树(DT),POA算法对3种算法模型均具有较为明显的优化作用。而对比PSO、SA、GWO和SSA等优化算法,POA优化的RF各项评估参数均达到最优,验证了POA算法对RF模型的优化能力。

进一步选取一台在运测试样本进行分析:某变电站一台型号为SFSZ11-180000/220的主变压器,出厂时间为2011年10月24日,投运时间为2012年12月5日,设备运行环境良好,投运11年后即2023

表18 模型评估参数对比表

Table 18 Comparison of model evaluation parameters

模型	$E_{\text{MAE}}/\text{年}$	E_{MSE}	E_{RMSE}
RF	2.003 3	6.193 3	2.488 6
SVM	8.202 8	121.437 8	11.019 9
DT	2.682 7	19.990 5	4.471 1
POA-SVM	2.572 3	11.875 0	3.446 0
POA-DT	2.502 3	14.547 0	3.814 1
PSO-RF	1.881 2	5.490 3	2.343 1
SA-RF	2.089 8	6.503 2	2.550 1
GWO-RF	1.916 6	5.479 2	2.340 8
SSA-RF	2.348 6	8.496 0	2.914 8
POA-RF	1.690 5	4.618 0	2.148 9

年11月测得油中微水含量为10.7 mg/L,击穿电压为60.2 kV,油介质损耗因数为0.074,体积电阻率为 $169\times10^{10}\ \Omega\cdot\text{m}$,离线油色谱实验测得 H_2 、 CH_4 、 C_2H_4 、 C_2H_6 、 C_2H_2 的体积分数分别为 3.513×10^{-6} 、 11.165×10^{-6} 、 1.386×10^{-6} 、 1.154×10^{-6} 、0,其近5年出现缺陷次数为4次,运行状态正常,平均负载率为26.84%,平均油温为43.33℃,平均绕组温度为43.04℃。将上述指标作为模型输入量,POA-RF模型预测的剩余寿命为24.51年。判断表明,该变压器的运行工况较好,且在运期间未发生明显事故,已运行11年,设计预期寿命为30年,由于之前考虑退役报废健康指数存有一定裕度,可认为预测的剩余寿命24.51年与实际相差不大,证明了POA-RF模型预测变压器剩余寿命的准确性。

4 结论

本文提出了一种基于鹌鹑算法优化随机森林模型和健康指数的油浸式变压器剩余寿命预测方法,主要得到如下结论:

(1)将变压器运行状态以调节因子的形式加入到剩余寿命预测模型,以POA优化的随机森林模型进行实验验证,结果证明,对于测试样本中的8组报废样本,POA-RF模型的平均绝对误差为1.187 0年,比RF模型的预测值和健康指数理论的计算值分别减小了20.52%和49.02%,能有效预测变压器的剩余寿命。而对于所有测试样本,和未优化的RF模型相比,POA-RF模型预测剩余寿命的平均绝对误差降低了15.61%,优化效果明显。

(2)RF模型预测精度明显高于支持向量机(SVM)和决策树(DT),POA算法对3种算法模型均

具有较为明显的优化作用。对比 PSO、SA、GWO 和 SSA 等优化算法, POA 优化的 RF 模型各项评估参数均达到最优, 验证了 POA 算法对 RF 模型的优化能力。在所有预测模型中, POA-RF 模型的评价指标均达到最优。

参考文献 References

- [1] 袁婉玲, 崔子轩, 禹洪波, 等. 基于粗糙集-G1 法组合赋权的变压器健康状态评估[J]. 电力建设, 2022, 43(3): 50-57.
YUAN Wanling, CUI Zixuan, YU Hongbo, et al. Transformer health status evaluation based on rough set G1 combined weighting[J]. Electric Power Construction, 2022, 43(3): 50-57.
- [2] 刘光伟, 关宇, 董明, 等. 变压器匝间故障特性与检测研究进展[J]. 绝缘材料, 2023, 56(11): 7-15.
LIU Guangwei, GUAN Yu, DONG Ming, et al. Research progress in characteristics and detection of transformer interturn fault [J]. Insulating Materials, 2023, 56(11): 7-15.
- [3] 耿江海, 祝建昊, 姚嘉欣, 等. 基于组合赋权和改进 TOPSIS 的绝缘纸老化状态评估[J]. 绝缘材料, 2025, 58(2): 129-138.
GENG Jianghai, ZHU Jianhao, YAO Jiaxin, et al. Ageing state evaluation of insulating paper based on combined assignment and improved TOPSIS[J]. Insulating Materials, 2025, 58(2): 129-138.
- [4] 赵跃, 冷威, 林琴, 等. 基于三维荧光的变压器绝缘油绝缘状态监测技术研究[J]. 绝缘材料, 2022, 55(12): 87-91
ZHAO Yue, LENG Wei, LIN Qin, et al. Study on insulation states monitoring technology of transformer insulating oils based on three-dimensional fluorescence spectroscopy[J]. Insulating Materials, 2022, 55(12): 87-91.
- [5] Da SILVA D G T, Da SILVA H J B, MARAFÃO F P, et al. Enhanced health index for power transformers diagnosis[J]. Engineering Failure Analysis, 2021, 126: 105427.
- [6] GUO H, GUO L. Health index for power transformer condition assessment based on operation history and test data[J]. Energy Reports, 2022, 8: 9038-9045.
- [7] 齐金平, 李鸿伟, 何鹏, 等. 牵引变压器油纸绝缘老化机理与寿命预测方法综述[J]. 绝缘材料, 2023, 56(9): 14-22.
QI Jinping, LI Hongwei, HE Peng, et al. Review on ageing mechanism and life prediction method of oil-paper insulation for traction transformer[J]. Insulating Materials, 2023, 56(9): 14-22.
- [8] 梁博渊, 刘伟, 杨欣桐. 变压器健康状态评估与剩余寿命预测[J]. 电网与清洁能源, 2010, 26(11): 37-43.
LIANG Boyuan, LIU Wei, YANG Xintong, et al. Transformer condition assessment and residual life prediction[J]. Power System and Clean Energy, 2010, 26(11): 37-43.
- [9] 李喜桂, 常燕, 罗运柏, 等. 基于健康指数的变压器剩余寿命评估[J]. 高压电器, 2012, 48(12): 80-85.
LI Xigui, CHANG Yan, LUO Yunbai, et al. Remnant life estimation of power transformer based on health index[J]. High Voltage Apparatus, 2012, 48(12): 80-85.
- [10] 李振柱, 谢志成, 熊卫红, 等. 考虑绝缘剩余寿命的变压器健康状态评估方法[J]. 电力自动化设备, 2016, 36(8): 137-142, 169.
LI Zhenzhu, XIE Zhicheng, XIONG Weihong, et al. Transformer health condition assessment considering insulation remaining life[J]. Electric Power Automation Equipment, 2016, 36(8): 137-142, 169.
- [11] 席禹, 赵继光, 林冬, 等. 基于 DBN 与健康指数的变压器评估与寿命预测方法[J]. 电子器件, 2022, 45(4): 878-882.
XI Yu, ZHAO Jiguang, LIN Dong, et al. Transformer evaluation and life prediction method based on DBN and health index[J]. Chinese Journal of Electron Devices, 2022, 45(4): 878-882.
- [12] 李以策, 郭江, 张珂斐, 等. 基于改进灰色预测与熵权融合体系的变压器寿命预测研究[J]. 电测与仪表, 2017, 54(9): 58-65.
LI Yice, GUO Jiang, ZHANG Kefei, et al. Transformer life prediction model based on the improved grey forecasting and entropy weight fusion system[J]. Electrical Measurement & Instrumentation, 2017, 54(9): 58-65.
- [13] 胡碧伟, 邓祥力, 贾声昊. 基于 ANFIS 的变压器寿命预测和状态评估[J]. 电测与仪表, 2022, 59(1): 61-68.
HU Biwei, DENG Xiangli, JIA Shenghao. Transformer life estimation and state assessment based on ANFIS[J]. Electrical Measurement & Instrumentation, 2022, 59(1): 61-68.
- [14] 刘咏鑫, 宋斌, 王力农, 等. 基于混沌序列优化的 BP 网络油纸绝缘变压器寿命预测[J]. 电测与仪表, 2022, 59(4): 137-143.
LIU Yongxin, SONG Bin, WANG Linong, et al. Life prediction of oil-paper insulated transformer based on chaotic sequence optimization BP neural network[J]. Electrical Measurement & Instrumentation, 2022, 59(4): 137-143.
- [15] SONI R, MEHTA B. Evaluation of power transformer health analysis by internal fault criticalities to prevent premature failure using statistical data analytics approach[J]. Engineering Failure Analysis, 2022, 136: 106213.
- [16] 凌静. 110~500 kV 电力变压器健康指数模型研究与应用[D]. 广州: 华南理工大学, 2017.
LING J. Research and application of health index modeling for 110-500 kV power transformers[D]. Guangzhou: South China University of Technology, 2017.
- [17] TAMMA W R, PRASOJO R A, SUWARNO. High voltage power transformer condition assessment considering the health index value and its decreasing rate[J]. High Voltage, 2021, 6(2): 314-327.
- [18] 江涛, 杨传凯, 王荆, 等. 基于健康指数的电力变压器绝缘状态评估方法研究[J]. 高压电器, 2024, 60(5): 124-129.
JIANG Tao, YANG Chuankai, WANG Jing, et al. Research on insulation state assessment method of power transformer based on health index[J]. High Voltage Apparatus, 2024, 60(5): 124-129.
- [19] 郑文光, 谢红玲, 李燕青, 等. 基于 WPHM 模型和模糊综合评估的变压器健康状态评价[J]. 电力科学与工程, 2021, 37(2): 32-41.
ZHENG Wenguang, XIE Hongling, LI Yanqing, et al. Health condition evaluation of transformer based on WPHM model and Fuzzy comprehensive evaluation[J]. Electric Power Science and Engineering, 2021, 37(2): 32-41.
- [20] 应永跃, 唐荣, 王世山, 等. 基于模糊综合评判的变压器固体绝缘状态评估[J]. 自动化与仪器仪表, 2022(7): 162-165.

- YING Yongyue, TANG Rong, WANG Shishan, et al. State evaluation of transformer solid insulation based on fuzzy comprehensive evaluation[J]. Automation & Instrumentation, 2022(7): 162-165.
- [21] 俞华,刘宏,王璇,等.基于多频超声和GWO-RF算法的绝缘油击穿电压预测研究[J].绝缘材料,2025,58(1):130-136.
- YU Hua, LIU Hong, WANG Xuan, et al. Research on prediction of breakdown voltage of insulating oil based on multi-frequency ultrasound and GWO-RF algorithm[J]. Insulating Materials, 2025,58(1):130-136.
- [22] 姚海成,李豹,黄兆琴.基于随机森林算法的电力系统小干扰稳定评估校正[J].广东电力,2023,36(7):60-69.
- YAO Haicheng, LI Bao, HUANG Zhaoshen. Small signal stability assessment and correction control of power system based on random forest algorithm[J]. Guangdong Electric Power, 2023,36(7):60-69.
- [23] TROJOVSKÝ P, DEHGHANI M. Pelican optimization algorithm: a novel nature-inspired algorithm for engineering applications[J]. Sensors, 2022,22(3):855.
- [24] 朱佳淼,王召斌,李久鑫.基于POA优化支持向量回归模型的航天电磁继电器贮存寿命预测[J].电器与能效管理技术,2023(8):61-67.
- ZHU Jiamiao, WANG Zhaobin, LI Jiuxin. Prediction of storage life of aerospace electromagnetic relay based on POA optimization support vector regression model[J]. Appliance and Energy Efficiency Management Technology, 2023(8):61-67.
- [25] 陈晓华,吴杰康,张勋祥,等.基于鹈鹕优化算法优化广义回归神经网络的电动汽车充电负荷短期预测[J].山东电力技术, 2024,51(7):1-9.
- CHEN Xiaohua, WU Jiekang, ZHANG Xunxiang, et al. Electric vehicle charging load short-term prediction based on generalized regression neural network optimized by pelican optimization algorithm[J]. Shandong Electric Power, 2024,51(7):1-9.

收稿日期:2025-04-22;修回日期:2025-07-14。

作者简介:

温爱辉(1983—),男(汉族),广东梅州人,高级工程师,主要从事电力系统及其自动化方向的研究;

通信作者:高凌志(1999—),男(汉族),安徽淮南人,硕士生,主要从事变压器故障诊断和寿命预测的研究。