

基于改进箱线图和 ISMAHHO-KELM 的变压器油中溶解气体体积分数预测

付文龙^{1,2}, 赵一博¹, 傅雨晨¹, 吴川锋¹

(1. 三峡大学电气与新能源学院, 湖北 宜昌 443000; 2. 三峡大学梯级水电站运行与控制湖北省重点实验室, 湖北 宜昌 443000)

摘要: 油中溶解气体分析能够有效地揭示变压器内部的运行状况, 在评估变压器运行状态和预测潜在故障方面具有重要作用。为此文中提出一种基于改进箱线图和 ISMAHHO-KELM 的变压器油中溶解气体体积分数预测方法。首先, 为提高数据质量, 利用基于中位偏差系数的改进箱线图对原始时间序列进行离群值检测和校正; 然后采用变分模态分解对校正过的时间序列进行分解, 得到多个子序列, 以削弱时间序列的非平稳性; 其次, 通过核极限学习机模型预测子序列, 同时提出改进黏菌-哈里斯鹰算法优化其超参数; 最后, 重构各子序列的预测值, 得到最终预测结果。实例分析表明所提方法具有较高的可靠性和稳定性, 能够为变压器维修决策提供有力支撑。

关键词: 油中溶解气体; 改进箱线图; 变分模态分解; 改进黏菌-哈里斯鹰算法

Viscosity Prediction of Dissolved Gas in Transformer Oil Based on Improved Boxplot and ISMAHHO-KELM

FU Wenlong^{1,2}, ZHAO Yibo¹, FU Yuchen¹, WU Chuanfeng¹

(1. College of Electrical Engineering and New Energy, China Three Gorges University, Hubei Yichang 443000, China;

2. Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Hubei Yichang 443000, China)

Abstract: The analysis of dissolved gases in oil can effectively reveal operational condition inside transformers, which can assess transformer's operational status and predicting potential faults efficiently. Thus, a dissolved gases forecasting framework is proposed based on improved boxplot and ISMAHHO-KELM in this paper. Firstly, to improve data quality, the improved boxplot fused with median deviation coefficient is used to detect and correct outliers. Then, variational modal decomposition is adopted decompose the processed time series into multiple sub-sequences to reduce the non-stationarity of the original time series. Next, kernel extreme learning machine is applied to predict the sub-sequences. Meanwhile, improved slime mold algorithm-harris hawk optimization is proposed to its hyperparameters. Finally, ultimate prediction value is acquired through reconstructing predicted results of each subsequence. Experiments verify the reliability and stability of the proposed framework, thereby providing strong support for maintenance decision-making.

Key words: dissolved gases in oil; improved boxplot; variational modal decomposition; improved slime mold algorithm-harris hawk optimization

0 引言

电力变压器在电力系统中扮演着关键的角色, 其运行状态将直接影响整个系统的稳定性^[1-4]。油中溶解气体分析能有效反映变压器绝缘系统在异常的温度和内部放电情况下变压器油及绝缘纸的异常状态, 可为评估变压器内部状态和早期故障提供重要依据

^[5]。因此, 研究油中溶解气体体积分数预测方法至关重要。

目前, 油中溶解气体的预测方法主要分为物理统计学方法和人工智能预测方法。其中, 常见的物理统计学模型主要有时间序列模型^[6]、灰色关联模型^[7]和差分自回归滑动平均模型^[8]等, 尽管物理统计学模型

擅长处理线性数据,但其预测非线性和非平稳性的油中溶解气体体积分数数据具有局限性。相对地,人工智能预测模型对非线性数据有良好的拟合效果,因此得到了迅速的发展。常用的方法有支持向量机(SVM)^[9]、长短期记忆神经网络(LSTM)^[10]和极限学习机(ELM)^[11]等。文[12]将核熵成分作为 SVM 的输入对油中溶解气体浓度进行预测,验证了 SVM 模型的有效性。文[13]基于奇异谱分析与 LSTM 提出预测模型,取得较好的效果。文[14]利用 ELM 预测变压器油中气体浓度。与 SVR 相比,ELM 模型结构简单,具有学习速度快、非线性处理能力强的特点,但是模型设置输入层权值以及隐含层偏差时伴随着随机性,从而导致模型可能处于不稳定状态。相比之下,核极限学习机(KELM)通过引入核函数,有效提高了模型的稳定性和泛化能力^[15]。文[16]将改进的黏菌优化算法与 KELM 模型进行结合,解决了 KELM 模型不易调参,寻优困难的问题。

以上研究提升了模型预测性能,但仍存在一些问题值得进一步研究和讨论:①变压器原始采样时间序列受到由外部污染、变压器运行状态、设备故障等多种因素产生的离群值^[17],会降低数据质量,从而对预测模型的性能产生负面影响。而离群值检测与校正方法应用较少且传统离群值检测方法分析数据能力有限;②部分研究采用人工和经验方法设置预测模型的固有参数,导致难以达到最优的预测性能。此外,虽然优化算法能够通过参数寻优在一定程度上提升预测性能,并且已有大量学者在原算法的基础上进行改进以期获得更强的寻优能力,但对优化算法进行融合改进的方式仍存在局限性。

综上,文中提出一种基于改进箱线图(BoxplotMC)和 ISMAHHO-KELM 的变压器油中溶解气体体积分数预测方法。首先,为提高数据质量,利用基于中位偏差系数(medcouple)的改进箱线图对原始时间序列进行离群值检测和校正;然后采用变分模态分解(variational model decomposition, VMD)对校正过的时间序列进行分解,得到多个子序列,以削弱时间序列的非平稳性;其次,采用核极限学习机模型预测子序列,同时提出改进黏菌-哈里斯鹰算法(improved slime mold algorithm-harris hawk optimization, ISMAHHO)优化其超参数;最后,重构各子序列的预测值,得到最终预测结果,并通过算例分析验证所提方法的可靠性和稳定性。

1 改进箱线图

箱线图(boxplot),又称为“盒须图”,是 John Tukey 提出的一种用于数据可视化的方法^[18]。它通过显示数据的最小值(*Min*)、下四分位数(Q_1)、中位数(Q_2)、上四分位数(Q_3)、最大值(*Max*)直观的描述数据的分布情况和性质,见图 1。箱线图通常由一个矩形框(箱体)和两条延伸出去的线段(触须)组成,箱体代表数据的四分位距(inter quartile range, *IQR*, 即 Q_3-Q_1),中间的线表示中位数,箱体上下边界则为上四分位数和下四分位数,分别对应原始气体体积分数时间序列的 50%、25%、75%。触须的长度通常为 1.5 倍 *IQR*,上下界限对应的最大值、最小值的计算式见式(1),超过这个界限的数据点则会被认为是离群值,其将会被极限值代替。

$$\begin{cases} Max = Q_3 + 1.5IQR \\ Min = Q_1 - 1.5IQR \end{cases} \quad (1)$$

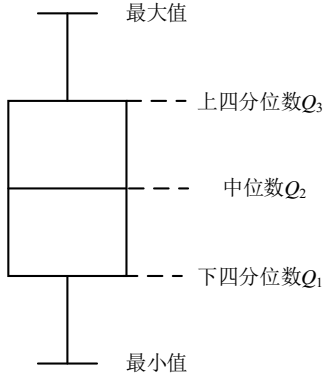


图1 箱体结构图

Fig.1 Structure diagram of boxplot

为衡量数据集对称性和尾部厚重性,引入中位偏差系数用于描述数据分布的对称性和偏斜程度,其计算式见式(2)

$$\begin{cases} x_i \leq Q_2 \leq x_j \\ m(x_i, x_j) = \frac{(x_j - Q_2) - (Q_2 - x_i)}{x_j - x_i} \\ MC = med(m(x_i, x_j)) \end{cases} \quad (2)$$

式(2)中: x_i 和 x_j 分别为小于等于中位数和大于等于中位数的序列的样本。由式(2)可以求得 MC , 然后利用原始序列向上或向下的趋势调整箱体图确定的时间序列的边界。左偏数据表现为 MC 为负的下降趋势, 其中上限 Max 和下限 Min 表示为

$$\begin{cases} Max = Q_3 + 1.5IQR \cdot e^{4MC} \\ Min = Q_1 - 1.5IQR \cdot e^{-3MC} \end{cases} \quad (3)$$

而上升趋势由正 MC 的右偏数据决定, 上限 Max 和下限 Min 为

$$\begin{cases} Max = Q_3 + 1.5IQR \cdot e^{3MC} \\ Min = Q_1 - 1.5IQR \cdot e^{-4MC} \end{cases} \quad (4)$$

2 变分模态分解

VMD 是由 Dragomiretskiy K 等人提出的一种信号变分分解方法^[19-20], 能有效解决 EMD 出现的模态混叠问题。该方法分为变分问题的构造和求解两部分。

对于原始信号 $x(t)$, 可以构造其约束表达式

$$\begin{cases} \min_{\{u_k\}, \{w_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[(\delta(t) + \frac{j}{\pi t}) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum_{k=1}^K u_k(t) = x(t) \end{cases} \quad (5)$$

式(5)中: K 表示模态数量; ∂_t 表示时间 t 的偏微分函数; $\delta(t)$ 表示狄拉克函数; $u_k(t)$ 表示第 k 个模态分量在 t 时刻的值; w_k 表示各分量的中心频率; $x(t)$ 表示原信号; j 表示虚数单位; $*$ 表示卷积运算。

其求解过程如下:

首先将约束问题转化为非约束问题, 得到式(6)

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[(\delta(t) + \frac{j}{\pi t}) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ & + \left\| x(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), x(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \quad (6)$$

式(6)中: λ 表示拉格朗日算子; α 表示惩罚因子。

其次运用傅里叶变换进行时-频域变换, 然后采用交替方向乘子法对 u_k^{n+1} 、 w_k^{n+1} 、 λ^{n+1} 进行迭代更新, 表达式为:

$$u_k^{n+1}(w) = \frac{x(w) - \sum_{i \neq k} u_i^n(w) + \lambda^n(w) / 2}{1 + 2\alpha(w - w_k^n)^2} \quad (7)$$

$$w_k^{n+1} = \frac{\int_0^\infty w |u_k^n(w)|^2 dw}{\int_0^\infty |u_k^n(w)|^2 dw} \quad (8)$$

$$\lambda^{n+1}(w) = \lambda^n(w) + \tau \left[x(w) - \sum_{k=1}^K u_k^{n+1}(w) \right] \quad (9)$$

式(7)-(9)中: n 表示迭代次数; τ 表示保真度系数。

最后, 当 u_k 满足式(10)达到收敛精度或 n 达到最大迭代次数时则停止。

$$\sum_{k=1}^K \frac{\|u_k^{n+1} - u_k^n\|_2^2}{\|u_k^n\|_2^2} < \varepsilon \quad (10)$$

式(10)中: ε 表示收敛精度。

3 基于层次结构的改进黏菌-哈里斯鹰优化算法

3.1 黏菌算法

黏菌算法(slime mold algorithm, SMA)^[21]是通过学习黏菌的自组织行为设计出的一种启发式优化算法, 数学模型表示如下。

首先, 利用以下公式来模拟黏菌接近食物的行为过程:

$$X(t+1) = \begin{cases} X_b(t) + V_b \cdot (w \cdot X_A(t) - X_B(t)), & r < p \\ V_c \cdot X(t), & r \geq p \end{cases} \quad (11)$$

式(11)中: t 表示当前的迭代次数; $X_b(t)$ 表示食物浓度最高的个体位置; V_b 表示一个范围在 $-a$ 到 a 之间的参数; V_c 表示从 1 到 0 线性递减的参数; w 表示黏菌的质量; $X_A(t)$ 、 $X_B(t)$ 表示两个随机选择的黏菌个体位置。此外, 参数 p 的函数表达式为

$$p = \tanh |S(i) - DF|, \quad i \in 1, 2, \dots, M \quad (12)$$

式(12)中: $S(i)$ 表示第 i 个黏菌的适应度; DF 表示最佳适应度; M 表示种群数量。范围参数 a 的函数表达式为

$$a = \arctan h \left[-\left(\frac{t}{\max_t} \right) + 1 \right] \quad (13)$$

式(13)中: \max_t 表示最大迭代次数。黏菌质量 w 的函数表达式为:

$$w(S_r) = \begin{cases} 1 + r \cdot \lg\left(\frac{B_f - S(i)}{B_f - W_f} + 1\right), & \text{condition} \\ 1 - r \cdot \lg\left(\frac{B_f - S(i)}{B_f - W_f} + 1\right), & \text{other} \end{cases} \quad (14)$$

$$S_r = \text{sort}(S(i)) \quad (15)$$

式(14)、(15)中: r 表示 $[0, 1]$ 内的随机值; B_f 、

W_f 分别表示当前迭代期间的最优和最差适应度值; **condition** 表示其 $S(i)$ 位于总体适应度降序排列的前半部分; **other** 表示其 $S(i)$ 位于总体适应度降序排列的后半部分; S_r 表示 $S(i)$ 按升序排序后的适应度值。

黏菌在找到当前食物位置后, 仍要探索未知区域去寻找更好的食物位置, 其位置更新的函数表达式为

$$X^* = \begin{cases} \text{rand} \cdot (ub - lb) + lb, & \text{rand} < z \\ X_b(t) + V_b \cdot (w \cdot X_A(t) - X_B(t)), & r < p \\ V_c \cdot X(t), & r \geq p \end{cases} \quad (16)$$

式(16)中: ub 、 lb 分别表示搜索范围的上下限; rand 表示 $[0, 1]$ 内的随机数; z 表示转换概率, 用来调节算法的全局搜索和局部搜索; 文[21]的实验已验证, $z = 0.03$ 时, 寻优效果最好, 故本文设置 $z = 0.03$ 。

3.2 哈里斯鹰算法

哈里斯鹰算法(harris hawk optimization, HHO)^[22]是通过将哈里斯鹰捕捉猎物和猎物逃逸的动态过程融入到优化问题的求解中设计的智能寻优算法, 该算法主要由探索阶段、探索—开发转化阶段和开发阶段组成。实现方式如下。

1)探索阶段, 哈里斯鹰停留在搜索区间的某个位置, 并通过概率相同的两种搜索策略寻找猎物。策略表示为

$$X_{t+1} = \begin{cases} X_{\text{rand}}(t) - r_1 |X_{\text{rand}}(t) - 2r_2 X(t)| & q \geq 0.5 \\ (X_{\text{rabbit}}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)) & q < 0.5 \end{cases} \quad (17)$$

式(17)中: t 表示迭代次数; X_{t+1} 、 $X(t)$ 分别表示种群在当前和下次迭代时的位置; $X_{rand}(t)$ 表示随机位置; $X_{rabbit}(t)$ 表示截止到第 t 次迭代所得的全局最优个体, 也表示猎物位置; r_1 、 r_2 、 r_3 、 r_4 和 q 表示区间 0 到 1 之间的随机数; UB 和 LB 分别表示搜索区域的上下边界; $X_m(t)$ 表示第 t 次迭代后, 个体的平均位置, 表达式为

$$X_m(t) = \frac{1}{N} \sum_{i=1}^N X_i(t) \quad (18)$$

式(18)中: $X_i(t)$ 为第 t 次迭代时第 i 只鹰的位置;

N 为种群的数量。

2)探索—开发转化阶段, 定义猎物的逃逸能量 E , HHO 根据其大小判断是否在全局探索和局部开发之间进行策略转换, 逃逸能量 E 表示为

$$E = 2E_0 \left(1 - \frac{t}{T}\right) \quad (19)$$

式(19)中: T 和 t 分别表示最大迭代次数和当前迭代次数; E_0 为大小在 -1 到 1 之间的随机数, 表示猎物的初始逃逸能量; 当 $|E| \geq 1$ 时进入探索阶段, 否则进入开发阶段。

3)开发阶段, 根据哈里斯鹰捕捉猎物和猎物动态逃逸过程, HHO 设计了多种攻击策略, 具体见表 1。

表 1 局部开发阶段位置更新策略

Table 1 Location update strategy in local development stage

攻击策略	条件	位置更新公式	编号
软围攻	$g \geq 0.5$ $ E \geq 0.5$	$X(t+1) = (X_{rabbit}(t) - X(t)) - E \times J \times X_{rabbit}(t) - X(t) $	(20)
硬围攻	$g \geq 0.5$ $ E < 0.5$	$X(t+1) = (X_{rabbit}(t) - X(t)) - E \times X_{rabbit}(t) - X(t) $	(21)
渐进式快速俯冲软包围	$g < 0.5$ $ E \geq 0.5$	$X(t+1) = X_{rabbit}(t) - E \times JX_{rabbit}(t) - X(t) $ $Z = Y + S \times LF(D)$ $X(t+1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases}$	(22)
渐进式快速俯冲硬包围	$g < 0.5$ $ E < 0.5$	$X(t+1) = X_{rabbit}(t) - E \times JX_{rabbit}(t) - X_m(t) $ $Z = Y + S \times LF(D)$ $X(t+1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases}$	(23)

表 1 中: g 为 0 到 1 之间的随机数; J 表示 0 到 2 之间的随机数; D 表示问题的维数; LF 表示 Levy 飞行函数。

3.3 改进黏菌—哈里斯鹰优化算法

考虑到单一优化算法仍存在陷入局部最优解的问题, 文中提出一种基于层次结构的改进黏菌-哈里斯鹰优化算法, 充分结合 SMA 寻优速度快、算法简单和 HHO 参数少、寻优精度高、鲁棒性强的优点。

其结构见图 2, 其中 SMA 作为上层进行优化搜索, HHO 作为下层进行位置更新。上层包含与下层 M 个群体相对应的 M 个搜索个体, 并且每个下层群体中包含 N 个搜索个体, 新的位置首先由下层 HHO 算法更新得到 M 个最优解, 然后用上层 M 个搜索个体进

行贪婪优化，最后，上层 SMA 算法继续更新位置，以获得最终的最优解。此外，采用 Logistic 映射对其进行混沌初始化，以增强 ISMAHHO 的全局搜索能力，使种群在搜索空间中更加均匀地分布，公式为

$$x_{k+1} = r \cdot x_k (1 - x_k), x_k \in (0,1) \quad (24)$$

式(24)中， r 表示在 0 到 4 之间的控制参数。

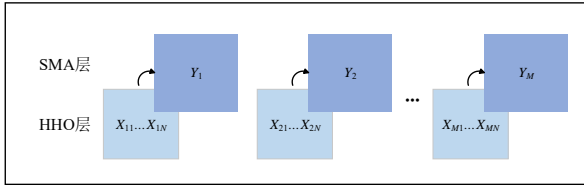


图 2 ISMAHHO 算法结构图

Fig. 2 Structure diagram of ISMAHHO

所提 ISMAHHO 算法的具体步骤如下：

- 1)采用 Logistic 映射策略对种群进行初始化；
- 2)计算每个搜索个体的适应度值；

3)执行迭代过程， $t = t+1$ ；

4)对于下层的搜索个体，根据式(17)-(23)更新它们的位置；

5)根据对应的下层群体找到的最优解更新 X_i ，

其中 X_i 为上层搜索个体。

6)对于上层的搜索个体，根据 SMA 更新其位置。

7)若满足迭代终止条件，则输出最优结果，否则重新返回步骤 3)。

3.4 算法验证

为评估算法的性能优越性，文中选择单峰、多峰和定维多峰 3 种不同的基准测试函数对 ISMAHHO 进行测试，具体函数表达式见表 2。

表 2 基准测试函数

Table 2 Benchmark functions

序号	基准测试函数	维度	搜索范围	最优值
F1	$f_3(x) = \sum_{i=1}^n (\sum_{j=1}^i x_j)^2$	30	[-100, 100]	0.000 0
F2	$f_6(x) = \sum_{i=1}^{n-1} (x_i + 0.5)^2$	30	[-30, 30]	0.000 0
F3	$f_{10}(x) = -20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)) + 20 + e$	30	[-32, 32]	0.000 0
F4	$f_{12}(x) = \frac{\pi}{n} \left\{ 10 \sin(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 [1 + 10 \sin^2(\pi y_{i+1})] + (y_n - 1)^2 \right\} + \sum_{i=1}^n u(x_i, 10, 100, 4)$ $y_i = 1 + \frac{x_i + 1}{4} u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < -a \end{cases}$	30	[-50, 50]	0.000 0
F5	$f_{14}(x) = \left(\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^2 (x_i - a_{ij})^6} \right)^{-1}$	2	[-65, 65]	0.000 0
F6	$f_{15}(x) = \sum_{i=1}^{11} \left[a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4} \right]^2$	4	[-5, 5]	0.000 3

为验证所提 ISMAHHO 算法的有效性,实验选取 SMA, HHO, WOA, SCA, MFO 和 PSO 算法进行比较,取 10 次运行结果的平均值(AVG)和标准差(STD)作为测试指标,且为确保比较的公平性,实验

过程中均设置种群数量 $N=30$, 迭代次数 $T=200$, 算法收敛图见图 3, 实验测试结果对比见表 3, 结果表明,对于不同的测试函数, ISMAHHO 优化算法具有更快的收敛速度和更高的计算精度。

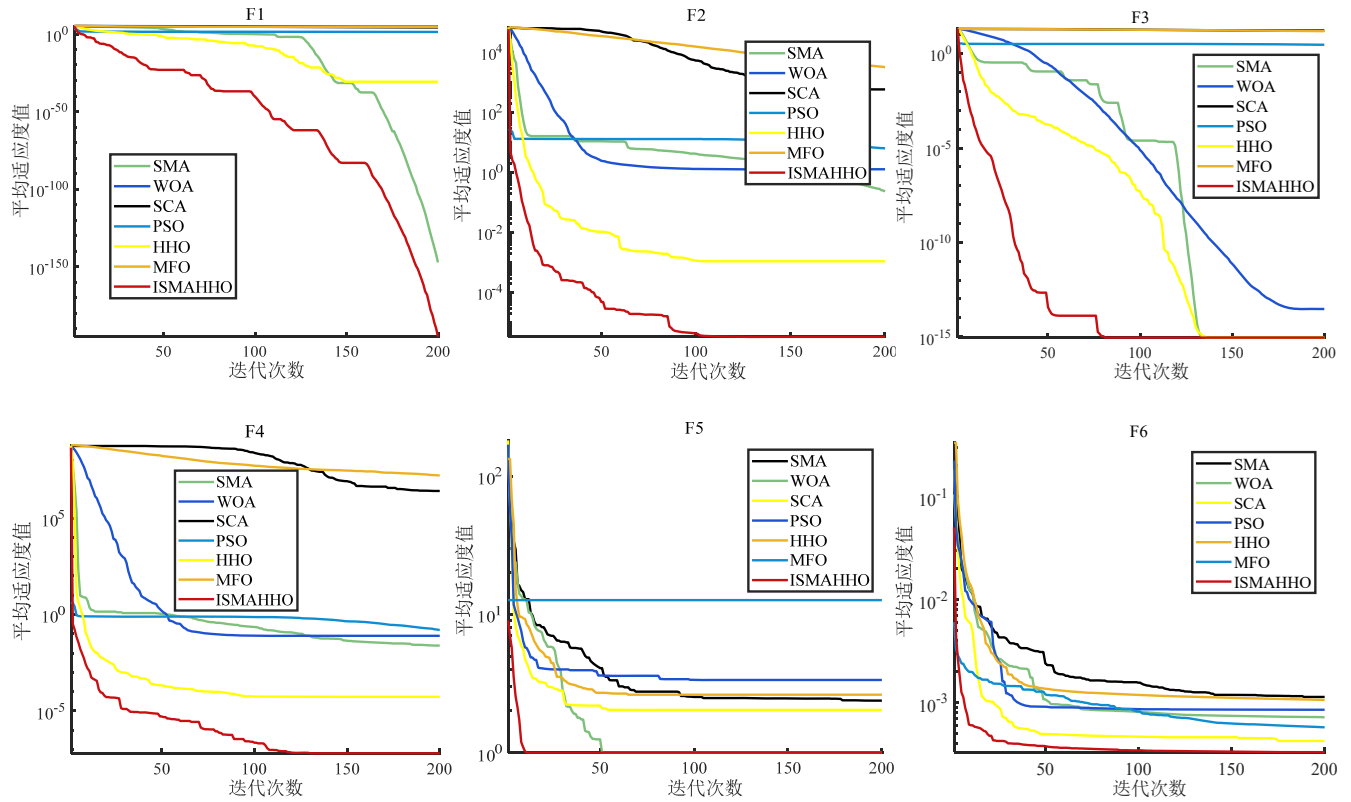


图 3 算法收敛图

Fig. 3 Convergence diagram of the algorithm

表 3 实验测试结果对比

Table 3 Comparison of experimental test results

序号	指标	ISMAHHO	SMA	HHO	WOA	SCA	MFO	PSO
F1	AVG	4.99×10^{-198}	1.64×10^{-168}	7.87×10^{-30}	8.71×10^4	2.14×10^4	2.89×10^4	2.06×10^1
	STD	0.00×10^0	0.00×10^0	3.39×10^{-29}	1.93×10^4	6.40×10^3	8.52×10^3	1.02×10^1
F2	AVG	1.56×10^{-6}	1.98×10^{-1}	0.11×10^{-2}	1.22×10^0	7.71×10^2	3.38×10^3	6.18×10^0
	STD	4.43×10^{-6}	1.54×10^{-1}	0.22×10^{-2}	0.41×10^0	8.44×10^2	4.08×10^3	1.95×10^0
F3	AVG	8.88×10^{-16}	8.88×10^{-16}	8.88×10^{-16}	2.58×10^{-14}	1.56×10^1	1.62×10^1	2.76×10^0
	STD	0.00×10^0	0.00×10^0	0.00×10^0	1.81×10^{-14}	7.22×10^0	4.96×10^0	0.28×10^0
F4	AVG	5.49×10^{-9}	1.43×10^{-2}	2.58×10^{-5}	9.58×10^{-2}	3.67×10^6	3.30×10^5	1.73×10^{-1}
	STD	1.31×10^{-8}	1.42×10^{-2}	3.21×10^{-5}	6.24×10^{-2}	6.47×10^6	8.43×10^5	8.20×10^{-2}
F5	AVG	9.98×10^{-1}	9.98×10^{-1}	1.74×10^0	2.92×10^0	2.12×10^0	1.54×10^0	1.27×10^1
	STD	1.85×10^{-11}	7.34×10^{-12}	1.53×10^0	2.86×10^0	1.03×10^0	1.09×10^0	1.48×10^{-13}
F6	AVG	3.33×10^{-4}	6.37×10^{-4}	4.84×10^{-4}	0.13×10^{-2}	0.11×10^{-2}	0.11×10^{-2}	5.47×10^{-4}
	STD	4.05×10^{-5}	2.75×10^{-4}	3.52×10^{-4}	0.26×10^{-2}	3.57×10^{-4}	3.24×10^{-4}	1.54×10^{-4}

4 基于改进箱线图和 ISMAHHO-KELM 的预测方法

考虑到原始时间序列易受到外部污染、油质量变化、设备故障等多种因素的影响而产生异常值，提出基于中位偏差系数的改进箱线图提高数据质量，从而提高模型预测精度。此外，由于预测模型的固有参数直接影响其预测性能，提出基于层次结构的 ISMAHHO 优化 KELM 模型的参数。在此基础上，本文提出基于改进箱线图和 ISMAHHO-KELM 的变压器油中溶解气体体积分数预测方法，具体的预测流程图见图 4。

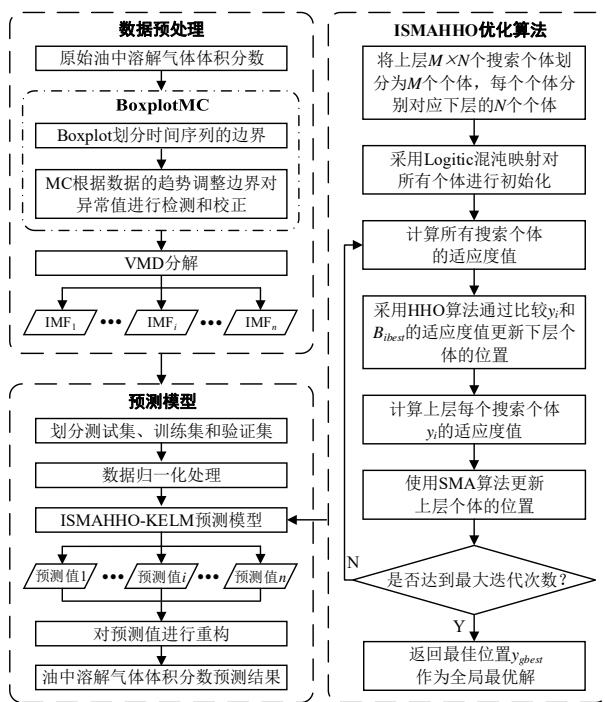


图 4 BoxplotMC-VMD-ISMAHHO-KELM 预测流程图

Fig. 4 BoxplotMC-VMD-ISMAHHO-KELM prediction flow chart

实现步骤如下：

1)利用提出的 BoxplotMC 对原始油中溶解气体时间序列进行异常值检测并校正，提升数据质量；

2)采用 VMD 将校正过的时间序列分解为模态分量，进一步削弱其非平稳性；

3)对模态分量进行划分，并将划分好的数据进行归一化；

4)建立 KELM 模型对各模态分量进行训练，同时通过提出的 ISMAHHO 对 KELM 的超参数进行优化；

5) 采用训练完成的 BoxplotMC-VMD-ISMAHHO-KELM 模型对各模态分量进行预测；

6)对各模态分量的预测值进行反归一化并重构，得到最终预测结果。

5 实例分析

5.1 数据描述

文中选用从 2018 年 2 月 16 日至 2020 年 11 月 20 日某 500 kV 变压器油色谱在线监测系统数据进行分析，以 1 天为采样间隔，共获得 1 008 组数据，在进行实验前，原始采样数据将按照 4 : 1 : 2 被划分为训练集、验证集和测试集，见图 5。限于篇幅，仅以 CH₄ 为例详细分析。

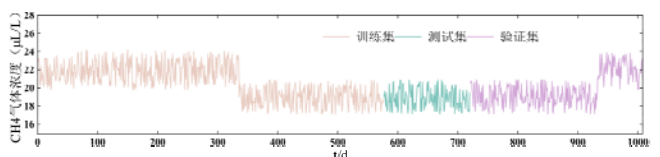


图 5 CH₄ 气体原始数据

Fig. 5 Raw data for CH₄ gas

5.2 数据预处理

从图 5 中可知，CH₄ 的原始数据序列不稳定且波动幅度大，所以导致其预测难度很大。因此，采用 BoxplotMC 减小异常数据对原始数据的影响，然后采用 VMD 对其进行分解，得到 9 组模态分量和 1 组残差分量，见图 6。最后，对各子序列按式(25)进行归一化。

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (25)$$

式(25)中: x 为原始样本数据; x_{\min} 为样本最小

值; x_{\max} 为样本最大值; x_n 为归一化后的样本值。

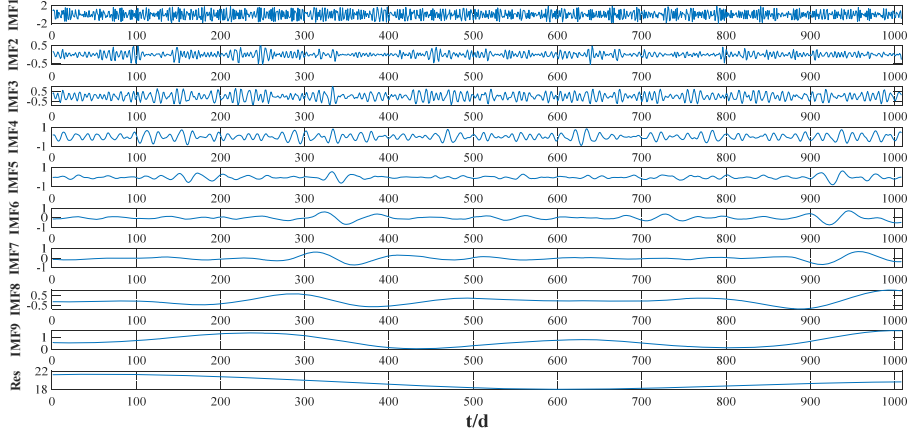


图 6 分解后的子序列

Fig. 6 The decomposed subsequences

5.3 评估指标

文中采用均方根误差(e_{RMSE})、平均绝对误差(e_{MAE})和平均绝对百分比误差(e_{MAPE})对结果进行评估分析^[23], 公式分别为:

$$e_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i - \hat{e})^2} \quad (26)$$

$$e_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |e_i - \hat{e}| \quad (27)$$

$$e_{\text{MAPE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{e_i - \hat{e}}{e_i} \right| \quad (28)$$

式(26)-(28)中: e_i 表示实际值; \hat{e} 表示模型预测

值; N 表示预测子序列的长度。

5.4 结果分析

将经过 Boxplot-VMD 处理的数据输入到 ISMAHHO-KELM 中进行训练, 在其训练完成后进行预测, 本次实验中, 由所提的 ISMAHHO 确定 KELM 的超参数, 其中设置 ISMAHHO 的种群数为 30, 最大迭代次数为 50; KELM 的正则化系数 C 和核函数 γ 搜索空间均为 [0.01, 1 000], 子序列预测对比结果见图 7。

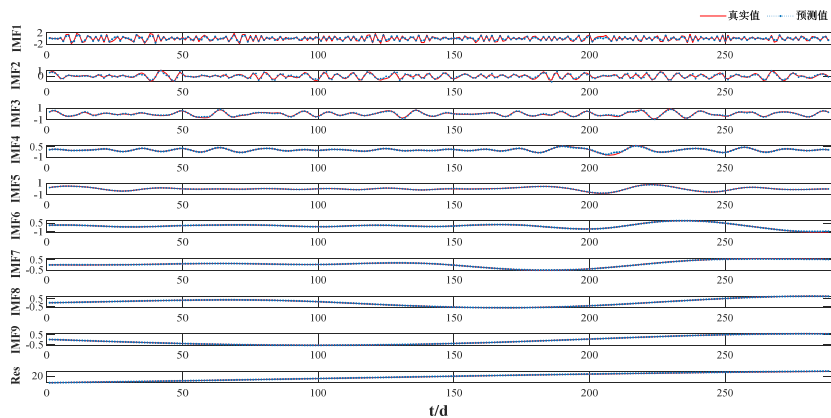
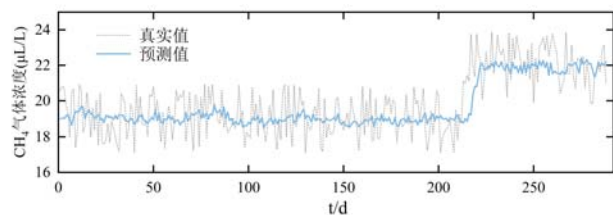


图 7 子序列对比结果

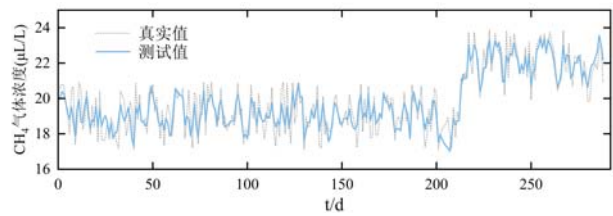
Fig. 7 Comparison results of subsequence

从图 7 中可知,文中所构造的 ISMAHHO-KELM 模型对经过 BoxplotMC-VMD 处理后得到的各子序列预测结果较为理想。证明 BoxplotMC 能够提高原始序列数据质量,以及 VMD 能够有效削弱其非线性,为后续预测提供有力保障。

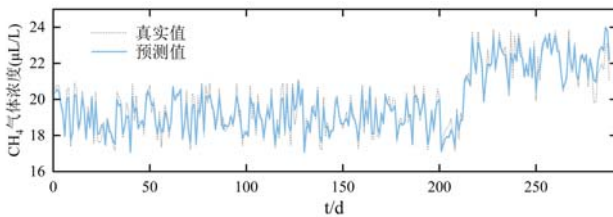
为验证文中所提出预测模型的有效性,分别与 KELM、VMD-KELM、BoxplotMC-VMD-KELM 模型进行对比分析,使用同一变压器 CH₄ 气体体积分数作为各模型的数据输入,且对比模型中的参数都采用网格搜索法进行选取,各预测模型的对比结果和评估指标分别见图 8 和表 4。



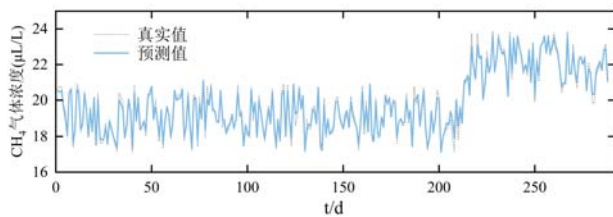
(a) KELM



(b) VMD-KELM



(c) BoxplotMC-VMD-KELM



(d) BoxplotMC-VMD-ISMAHHO-KELM

图 8 预测结果对比图

Fig. 8 Comparison chart of predicted results

表 4 评估指标

Table 4 Assessment of indicators

模型类型	评估指标类型		
	e_{RMSE}	e_{MAE}	e_{MAPE}
BoxplotMC-VMD-ISMAHHO-KELM	0.379 8	0.310 1	1.582 3
BoxplotMC-VMD-KELM	0.523 4	0.437 2	2.194 9
VMD-KELM	0.715 5	0.589 3	3.002 1
KELM	1.217 2	0.990 0	4.970 9

从图 8 和表 4 中可知,KELM 模型的 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 分别为 1.217 2、0.990 0、4.970 9, VMD-KELM 模型的 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 分别为 0.715 5、0.589 3、3.002 1; 与 KELM 模型相比,VMD-KELM 模型对应的评估指标分别下降了 0.501 7、0.400 7、1.968 8, 对应的下降率分别为 41.22%、40.48%、39.61%, 表明 KELM 模型对非线性的原始序列的预测效果较差,但原始序列经过 VMD 进行分解后,预测序列与真实序列的拟合效果有明显的提升,说明 VMD 分解可以有效降低原始数据的非平稳性,以进一步提高预测精度。与 VMD-KELM 模型相比,BoxplotMC-VMD-KELM 模型对应的评价指标有明显的下降,BoxplotMC-VMD-KELM 的 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 分别为 0.523 4、0.437 2、2.194 9,其分别下降了 0.192 1、0.152 1、0.807 2,对应的下降率分别为 26.85%、25.81%、26.89%,表明信号分解前,采用基于改进箱线图的离群值检测方法能够有效分析数据及其变化趋势,从而提升数据质量和减小预测模型的预测误差。与 BoxplotMC-VMD-KELM 模型相比,BoxplotMC-VMD-ISMAHHO-KELM 模型对应的的评估指标分别为 0.379 8、0.310 1、1.582 3,其对应指标分别下降了 0.143 6、0.127 1、0.612 6,下降率分别为 27.44%、29.07%、27.91%,表明与网格搜索法选取参数相比,文中所提的 ISMAHHO 选择超参数具

有一定的优越性，也进一步说明所提预测方法的有效性。

5.5 其他气体分析结果

选取此变压器与 CH₄ 采样数据相同时段内油中溶解的其他气体体积分数数据进行预测分析，以进一

步验证所提模型的有效性，其预测结果相关评估指标见表 5。从表 5 中可知，文中所提模型对其他气体的误差评估指标均是最优。

表 5 不同模型对其他气体的评估指标

Table 5 Evaluation metric for different models on other gases

气体类型	评估指标	模型类型			
		BoxplotMC-VMD-ISMAHHO-KELM	BoxplotMC-VMD-KELM	VMD-KELM	KELM
C ₂ H ₄	e_{RMSE}	0.235 6	0.301 5	0.425 0	0.725 8
	e_{MAE}	0.186 1	0.246 1	0.344 9	0.587 6
	e_{MAPE}	2.110 2	2.456 9	3.558 0	6.192 7
C ₂ H ₆	e_{RMSE}	0.105 7	0.118 5	0.153 4	0.222 1
	e_{MAE}	0.088 8	0.097 3	0.127 5	0.188 1
	e_{MAPE}	2.230 4	2.260 6	3.200 5	4.740 5
H ₂	e_{RMSE}	1.333 4	2.204 0	2.661 4	3.923 1
	e_{MAE}	1.085 3	1.827 7	2.100 5	3.324 0
	e_{MAPE}	1.784 0	2.938 1	3.378 1	5.432 4
总烃	e_{RMSE}	0.382 1	0.404 5	0.664 7	1.295 0
	e_{MAE}	0.304 2	0.322 2	0.525 9	1.046 6
	e_{MAPE}	1.215 5	1.279 8	2.091 6	4.151 1

6 结论

为提升数据质量，以及增强模型预测性能，文中提出基于改进箱线图和 ISMAHHO-KELM 的变压器油中溶解气体体积分数的预测方法，结论如下：

1)利用基于中位偏差系数的改进箱线图对原始时间序列进行离群值检测和校正，相较于未采取离群值检测和校正的模型，基于改进箱线图的混合模型误差评估指标分别下降了 26.85%、25.81%、26.89%，证明基于中位偏差系数的改进箱线图能够提高数据质量，进而提高模型的预测精度。

2)文中所提 ISMAHHO 算法可以有效解决单一优化算法陷入局部最优解的问题，进而优化模型参数，提高模型预测性能，相较于未采取 ISMAHHO 的

模型，基于 ISMAHHO 的所提模型误差评估指标分别下降了 27.44%、29.07%、27.91%。

3) 文中提出的 BoxplotMC-VMD-ISMAHHO-KELM 模型相较于其他预测模型，其相应误差评估指标均是最优，说明所提模型具有更好的数据拟合能力和更高的预测性能，可为变压器运维决策提供一定支持。

参考文献:

[1] 杨海晶, 孙运全, 朱 伟, 等. 基于 CEEMD-TCN 模型的变压器油中溶解气体浓度预测方法[J]. 电子器件, 2021, 44(4): 887-892.
 YANG Haijing, SUN Yunquan, ZHU Wei, et al. Prediction method of dissolved gas concentration in transformer oil based on CEEMD-TCN model[J]. Chinese Journal of Electron Devices, 2021, 44(4): 887-892.

[2] 苏 磊, 陈 璐, 徐 鹏, 等. 基于 GRNN 和 KPCA 组

- 合模型的变压器油中气体体积分数短期预测[J]. 高压电器, 2021, 57(1): 82-88.
- SU Lei, CHEN Lu, XU Peng, et al. Short-term prediction of gases dissolved in transformer oil based on GRNN and KPCA model[J]. High Voltage Apparatus, 2021, 57(1): 82-88.
- [3] 李小平, 白超, 石森. 基于 CNN-BiLSTM 模型的机车变压器油中溶解气体浓度预测方法[J]. 铁道学报, 2022, 44(5): 42-48.
- LI Xiaoping, BAI Chao, SHI Sen. Prediction method of dissolved gas concentration in transformer oil based on CNN-BiLSTM mode[J]. Journal of the China Railway Society, 2022, 44(5): 42-48.
- [4] 张玲, 卞建鹏, 郝培旭, 等. 基于 CEEMD 和 ABC-LSTM 的电力变压器油中溶解气体预测[J]. 兰州交通大学学报, 2021, 40(5): 58-64.
- ZHANG Ling, BIAN Jianpeng, HAO Peixu, et al. Prediction of dissolved gases in power transformer oil based on CEEMD and ABC-LSTM[J]. Journal of Lanzhou Jiaotong University, 2021, 40(5): 58-64.
- [5] 曹正江, 付文龙, 文斌, 等. 基于 TVFEMD 和多模型融合的变压器油中溶解气体体积分数预测方法[J/OL]. 高压电器: 1-13[2024-03-29].<http://kns.cnki.net/kcms/detail/61.1127.TM.20221202.1602.001.html>.
- CAO Zhengjiang, FU Wenlong, WEN Bin, et al. Concentration prediction of dissolved gas in transformer oil based on TVFEMD and multi-model fusion[J/OL]. High Voltage Apparatus: 1-13[2024-03-29].<http://kns.cnki.net/kcms/detail/61.1127.TM.20221202.1602.001.html>.
- [6] 刘慧鑫, 张江龙, 连鸿松, 等. 基于时间序列模型的变压器油中溶解气体预测[J]. 高压电器, 2019, 55(12): 193-199.
- LIU Huixin, ZHANG Jianglong, LIAN Hongsong, et al. Prediction of the gases dissolved in transformer oil by sequential learning[J]. High Voltage Apparatus, 2019, 55(12): 193-199.
- [7] 吴汶捷, 何怡刚, 段嘉珺, 等. 油中溶解气体的灰色模型残差修正融合算法[J]. 电子测量与仪器学报, 2018, 32(10): 87-94.
- WU Wenjie, HE Yigang, DUAN Jiajun, et al. Residual-modified fusion algorithm of dissolved gas in transformer oil based on grey model[J]. Journal of Electronic Measurement and Instrument, 2018, 32(10): 87-94.
- [8] 贾茹宾, 高金峰. 基于 ARIMA 模型的变压器油中溶解气体含量时间序列预测方法[J]. 郑州大学学报(工学版), 2020, 41(2): 67-72.
- JIA Rubin, GAO Jinfeng. Time series prediction method of dissolved gas content in transformer oil based on ARIMA model[J]. Journal of Zhengzhou University(Engineering Science Edition), 2020, 41(2): 67-72.
- [9] 唐勇波, 丰娟. KTA-SVM 的变压器油中溶解气体浓度预测[J]. 控制工程, 2017, 24(11): 2263-2267.
- TANG Yongbo, FENG Juan. A prediction method of dissolved gas content in transformer oil based on KTA-SVM[J]. Control Engineering of China, 2017, 24(11): 2263-2267.
- [10] 陈铁, 陈一夫, 李咸善, 等. 基于 CEEMDAN-SG-BiLSTM 的变压器油中溶解气体体积分数预测[J]. 高压电器, 2023, 59(12): 168-175.
- CHEN Tie, CHEN Yifu, LI Xianshan, et al. Prediction for dissolved gas concentration in power transformer oil based on CEEMDAN-SG-BiLSTM[J]. High Voltage Apparatus, 2023, 59(12): 168-175.
- [11] 陈龙, 张菁, 张昊立, 等. 基于 VMD 和射箭算法优化改进 ELM 的短期光伏发电预测[J]. 太阳能学报, 2023, 44(10): 135-141.
- CHEN Long, ZHANG Jing, ZHANG Haoli, et al. Short-term photovoltaic power generation forecast based on VMD-IAA-IHEKLM model[J]. Acta Energetica Sinica, 2023, 44(10): 135-141.
- [12] 江风云, 唐勇波. 基于核熵成分分析的油中溶解气体浓度预测[J]. 控制工程, 2020, 27(8): 1419-1424.
- JIANG Fengyun, TANG Yongbo. Concentration prediction of dissolved gas in oil based on kernel entropy component analysis[J]. Control Engineering of China, 2020, 27(8): 1419-1424.
- [13] 陈铁, 陈一夫, 李咸善, 等. 基于 SDS-SSA-LSTM 的变压器油中溶解气体浓度预测[J]. 电子测量技术, 2022, 45(12): 6-11.
- CHEN Tie, CHEN Yifu, LI Xianshan, et al. Prediction of dissolved gas concentration in transformer oil based on

- SDS-SSA-LSTM[J]. *Electronic Measurement Technology*, 2022, 45(12): 6-11.
- [14] 杨敏. 基于极限学习机的变压器油中气体浓度预测模型[J]. *变压器*, 2023, 60(11): 10-17.
- YANG Min. Prediction model of gas concentration in transformer oil based on extreme learning machine[J]. *Transformer*, 2023, 60(11): 10-17.
- [15] HUANG Guangbin, ZHOU Hongming, DING Xiaojian, et al. Extreme learning machine for regression and multiclass classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics : A Publication of the IEEE Systems, Man, and Cybernetics Society*, 2012, 42(2): 513-529.
- [16] 傅雨晨, 陈星, 付文龙, 等. 基于多级特征提取和 IHHO-KELM 的变压器油中溶解气体体积分数预测[J/OL]. *高压电器*: 1-14[2024-03-29].<http://kns.cnki.net/kcms/detail/61.1127.TM.20230713.1803.004.html>.
- FU Yuchen, CHEN Xing, FU Wenlong, et al. Prediction of dissolved gas concentration in transformer oil based on multistage feature extraction and IHHO-KELM[J/OL]. *High Voltage Apparatus*: 1-14[2024-03-29].<http://kns.cnki.net/kcms/detail/61.1127.TM.20230713.1803.004.html>.
- [17] 江军, 张文乾, 李波, 等. 电力变压器油中溶解气体离群值识别和数据重构[J/OL]. *电工技术学报*: 1-13[2024-03-29].<https://doi.org/10.19595/j.cnki.1000-6753.tces.231033>.
- JIANG Jun, ZHANG Wenqian, LI Bo, et al. Outlier detection and data Reconstruction of dissolved gas in oil for power transformers[J/OL]. *Transactions of China Electrotechnical Society*: 1-13[2024-03-29].<https://doi.org/10.19595/j.cnki.1000-6753.tces.231033>.
- [18] 赵颖祺, 陈玉虎, 张晰, 等. 一种基于箱线图的 SAR 图像舰船检测算法研究[J]. *中国海洋大学学报(自然科学版)*, 2021, 51(10): 130-140.
- ZHAO Yingqi, CHEN Yuhu, ZHANG Xi, et al. A ship detection algorithm for SAR image based on box-plot[J]. *Periodical of Ocean University of China*, 2021, 51(10): 130-140.
- [19] 谢乐, 仇炜, 李振伟, 等. 基于变分模态分解和门控循环单元神经网络的变压器油中溶解气体预测模型[J]. *高电压技术*, 2022, 48(2): 653-660.
- XIE Le, QIU Wei, LI Zhenwei, et al. Prediction model of dissolved gas in transformer oil based on variational modal decomposition and recurrent neural network with gated recurrent unit[J]. *High Voltage Engineering*, 2022, 48(2): 653-660.
- [20] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. *IEEE Transactions on Signal Processing*, 2014, 62(3): 531-544.
- [21] LI Shimin, CHEN Huiling, WANG Mingjing, et al. Slime mould algorithm: A new method for stochastic optimization[J]. *Future Generation Computer Systems*, 2020(111): 300-323.
- [22] HEIDARI A A, MIRJALILI S, FARIS H, et al. Harris hawks optimization: Algorithm and applications[J]. *Future Generation Computer Systems*, 2019(97): 849-872.
- [23] 付文龙, 章轩瑞, 张海荣, 等. 多尺度特征提取与非线性融合的综合能源系统多元负荷短期预测[J]. *电力系统及其自动化学报*, 2023, 35(12): 89-99.
- FU Wenlong, ZHANG Xuanrui, ZHANG Hairong, et al. Short-term multivariate load forecasting of integrated energy system based on multiscale feature extraction and nonlinear fusion[J]. *Proceedings of the CSU-EPSA*, 2023, 35(12): 89-99.
- 付文龙(19—), 男, 博士, 副教授, 主要从事电气设备状态监测与诊断研究(通信作者)(E-mail: ctgu_fuwenlong@126.com)。