

DOI:10.13296/j.1001-1609.hva.2026.10.007

基于缺失数据填补的变压器油中溶解气体体积分数预测

王振国^{1,2}, 宁 飞¹, 袁 宝¹, 贾 飞¹, 贾毓彦¹

(1. 内蒙古电力(集团)有限责任公司乌海供电分公司, 内蒙古 乌海 016000; 2. 华北电力大学电力工程系, 河北 保定 071003)

摘要: 油中溶解气体体积分数预测对于早期变压器故障诊断具有重要意义。针对当前油中溶解气体体积分数预测算法对数据完整性要求高,泛化能力弱等问题,以油浸式变压器为研究对象,提出一种基于缺失数据填补和改进卡尔曼滤波的油中溶解气体体积分数预测算法。首先以贝叶斯理论和高斯先验假设为基础,构建缺失数据填补的全概率模型,将缺失数据填补转化为求解模型参数最大似然估计问题,利用期望最大(expectation maximization, EM)算法完成模型求解,最终实现缺失数据填补。然后采用卡尔曼滤波进行气体体积分数预测,针对卡尔曼滤波预测结果易发散问题,引入BP神经网络并建立反馈机制,实现对预测结果的动态修正,提升预测精度。基于实际数据的试验结果表明,所提方法在数据缺失率为30%的情况下,填补结果的相关系数达到96.3%,相对于自回归移动平均(ARIMA)和极端随机树(extremely randomized trees, ERT)分别提升11.4%和5.7%。同时,所提方法的预测精度较高,相对于长短时记忆神经网络(long short term memory, LSTM)和长短时记忆—卷积神经网络(long short term memory-convolution neural network, LSTM-CNN)方法预测结果的均方根误差(e_{RMSE})和平均绝对误差(e_{MAPE})分别降低28.5%、33.6%。

关键词: 油浸式变压器; 油中溶解气体; 缺失数据填补; 卡尔曼滤波; 长短时记忆神经网络

Prediction of Dissolved Gas Concentration in Transformer Oil Based on Missing Data Imputation

WANG Zhenguo^{1,2}, NING Fei¹, YUAN Bao¹, JIA Fei¹, JIA Yuyan¹

(1. Wuhai Power Supply Branch of Inner Mongolia Electric Power(Group) Co., Ltd., Inner Mongolia Wuhai 016000, China;

2. Department of Electric Power Engineering, North China Electric Power University, Hebei Baoding 071003, China)

Abstract: The prediction of dissolved gas concentration in oil is of great significance for early fault diagnosis of transformer. To address the issues of high data integrity requirements and weak generalization ability in current prediction algorithms for dissolved gas concentration in oil, in this paper the oil immersed transformers is taken as the research object and a prediction algorithm for dissolved gas concentration in oil based on missing data imputation and an improved Kalman filtering is proposed. First, a full probability model for missing data imputation is constructed based on Bayesian theory and Gaussian prior assumption. The missing data imputation problem is transformed into solving the problem maximum likelihood estimation of model parameters, and the expectation maximization (EM) algorithm is employed to solve the model, ultimately achieving the imputation of missing data. Then, Kalman filtering is used for gas concentration prediction. To address the issue of divergence in Kalman filtering, a BP neural network is introduced and a feedback mechanism is established so to dynamically correct the prediction results and improve prediction accuracy. The experimental results based on actual data show under a data missing rate of 30%, the correlation coefficient of the imputation results achieved by the proposed method reaches 96.3%, which is 11.4% and 5.7% higher than those of the autoregressive integrated moving average (ARIMA) and extremely randomized trees (ERT), respectively. Meanwhile, the proposed method has a higher prediction accuracy, and the root mean square error (e_{RMSE}) and mean absolute error (e_{MAPE}) are reduced by 28.5% and 33.6% respectively compared to long short term memory (LSTM) and long short term memory-convolutional neural network (LSTM-CNN) methods.

Key words: oil immersed transformer; dissolved gas in oil; missing data imputation; Kalman filtering; long short term memory neural network

收稿日期:2026-05-26; 修回日期:2026-08-05

基金项目:国家自然科学基金面上项目(51877084)。

Project Supported by the General Program of the National Natural Science Foundation of China(51877084).

0 引言

油浸式变压器是中国电力系统中广泛使用的一种基础设备,其运行状态直接影响着电力系统的安全性和稳定性。然而,随着运行时间的增加,油浸式变压器不可避免的会出现绝缘劣化等问题,给电力系统带来安全隐患^[1-4]。因此,实时监测变压器的运行状态并提前进行故障预警,对于保证电力系统的安全稳定运行具有重要意义。

油中溶解气体分析(dissolved gas analysis, DGA)能够在不影响变压器正常工作的情况下,通过对气体组成成分或体积分数的变化进行分析从而实现对变压器运行状态监测和潜在故障发现^[5-6],而油中溶解气体体积分数预测是DGA中的关键技术之一^[7-12]。

目前已有的油中溶解气体体积分数预测方法主要有回归分析法,时间序列预测模型,灰色理论预测模型,支持向量机(support vector machine, SVM),长短时神经网络(long short term memory, LSTM),卷积神经网络(convolution neural network, CNN)、组合预测模型等。其中,回归分析法^[13],时间序列预测模型^[14]、灰色理论预测模型^[15]都属于线性预测方法,基本思想是假设气体体积分数在短时间内是线性变化,然后采用最小二乘,ARMA或GM(1, 1)等对线性变化趋势进行量化分析,具有原理简单,实时性高的优点,但是油中溶解气体体积分数是一种典型的非线性非平稳随机过程,上述线性模型不可避免的会出现误差累积,预测性能随时间下降等问题^[16-17]。SVM是基于统计学习理论和结构风险最小化准则建立的一种分类模型,在面对小样本、非线性以及高维模式识别问题时表现出了特有优势,文[18]将SVM应用至油中溶解气体体积分数预测领域,预测精度相对于ARMA提升超过5%,但是SVM的预测性能依赖于模型中参数(核参数,惩罚因子等)的选择。LSTM是针对长时间序列建模发展起来的一种循环神经网络,能够有效捕捉气体体积分数序列中隐藏的时间规律特征,具有较高的预测精度,但是模型的可解释性较差且对训练数据样本要求较高^[19]。单一模型只能从某个特定维度去分析气体体积分数序列,具有局限性,而组合预测模型通过将不同模型相结合,能够从更多维度挖掘气体体积分数序列中隐含的信息,从而进一步提升预测性能。文[20]将LSTM与CNN相结合,将CNN局部特征提取与LSTM时间特征提取相结合,扩充了信息提取维度,有效提升了预测精度,但仍然存在模型的可解释性较差的问题。

当前已有研究大都是假设观测到的油中溶解气体体积分数数据是完整的,满足算法的建模要求。然而实际中,受地理环境、强电磁干扰等因素影响,油中溶解气体观测数据会出现数据缺失等问题,这种情况下,上述方法均会因数据缺失而难以获取预期的性能^[21]。目前针对缺失数据填补的研究相对较少,文[22]提出一种基于自回归移动平均(ARIMA)模型进行缺失数据拟合的方法,实现了输变电设备缺失数据补全,仅适合样本缺失较少的情况(缺失率低于5%);文[23]将极端随机树(extremely randomized trees, ERT)引入油中溶解气体体积分数缺失数据填补领域,当缺失率达到20%时仍然能够获得较高的填补精度,但需要大量统计特性一致的完整数据作为训练样本,这在实际应用中难以满足。

在上述研究的基础上,文中首先基于贝叶斯理论和高斯先验假设建立缺失数据填补模型,利用期望最大(expectation maximization, EM)算法实现模型求解从而完成缺失数据填补。然后提出一种BP神经网络联合卡尔曼滤波(back propagation Kalman filter, BP-KF)方法进行油中溶解气体体积分数预测, BP-KF通过反馈机制实现对预测结果的动态修正,解决KF预测结果易发散问题。试验结果表明,所提方法相对于ARIMA和ERT等方法的数据填补效果更好,相对于LSTM和LSTM-CNN等方法的预测精度更高,且表现出了较强的泛化能力。

1 缺失数据填补算法

1.1 缺失数据填补模型构建

变压器油中溶解气体体积分数可以看作是随时间变化的一维时间序列,记为 $x = [x_1, x_2, \dots, x_T]$, T 为观测时间, $x_i (i = 1, \dots, T)$ 为 t 时刻的观测样本。当观测数据中存在样本缺失时,可以将其转化为利用观测矩阵 Φ 对完整数据进行采样的过程,即

$$y = \Phi x + \varepsilon \quad (1)$$

式(1)中: y 为观测到的不完整数据;观测矩阵 Φ 为对角矩阵形式。对角线上的元素取值为0或者1,当第 t 个对角线元素为0时,表示 x 中 t 时刻的样本 x_t 是缺失的, y 中 t 时刻的样本 $y_t = 0$;反之,当 Φ 第 t 个对角线元素不为0时,表示 x 中 t 时刻的样本 x_t 被采集到, y 中 t 时刻的样本 $y_t = x_t$ 。 ε 为数据采集过程中产生的测量误差,通常认为其服从均值为0的高斯分布,即

$$\varepsilon: N(0, \gamma^{-1}I) \quad (2)$$

式(2)中: γ 为测量误差的方差; I 为单位矩

阵。根据高斯分布的线性性质可知,观测数据 y 的条件概率 $p(y|x)$ 同样服从高斯分布,即

$$p(y|x) = N(\Phi x, \gamma^{-1}I) \quad (3)$$

进一步的,由贝叶斯理论,可以得到未知完整数据 x 的后验概率 $p(x|y)$ 为

$$p(x|y) = \frac{p(x)p(y|x)}{\int p(x)p(y|x)dx} \quad (4)$$

式(4)中, $p(x)$ 为 x 的先验分布。这里首先假设其服从高斯分布,后续对高斯分布假设的合理性进行分析。即

$$x: N(u, D) \quad (5)$$

式(5)中: u 为均值; D 为协方差矩阵。结合式(3)-(5),可以得到 $p(x|y)$ 的解析式为

$$\begin{cases} p(x|y) = N(\eta, \Omega) \\ \eta = \eta + \gamma(D^{-1} + \Phi^T \gamma \Phi)^{-1} \Phi^T (y - \Phi u) \\ \Omega = (D^{-1} + \gamma \Phi^T \Phi)^{-1} \end{cases} \quad (6)$$

式(6)中: η 为给定 y 条件下 x 的后验概率均值; Ω 为对应的协方差矩阵。

1.2 模型分析

式(1)给出的缺失数据填补模型描述的是已知观测数据 y 和观测矩阵 Φ , 求解未知完整数据 x 的问题。假设进行数据填补后得到的完整数据为 $\bar{x}$, 则填补误差可以表示为

$$\sigma = \iint \|\bar{x} - x\|^2 p(x|y) dx \quad (7)$$

根据最小均方误差准则,对式(7)求偏微分可以得到最小均方误差准则下的最优解为

$$\bar{x} = \int x p(x|y) dx = \eta \quad (8)$$

从式(8)可以看出,在最小均方误差准则下,最优解 $\bar{x}$ 为已知 y 条件下 x 的后验分布均值。

1.3 期望最大算法(EM)

根据式(8)可知,在最小均方误差准则下要实现缺失数据的最优填补效果,只需要求解式(6)模型中与 η 相关参数的最大似然估计即可,即将缺失数据填补问题转化为求模型参数的最大似然估计问题。目前常用的求解概率模型中参数最大似然估计的方法为EM算法^[24]。

对于式(1)、(6)给出的模型, y 为观测数据, x 为未知完整数据,即模型中的隐变量,模型中的参数集合为 $\Theta = \{\eta, \Omega, D, \gamma, u\}$ 。EM算法的目标是最大化式(9)给出的对数似然函数^[25]

$$\begin{aligned} \ln p(y|\Theta) &= \int q(x) \ln \left\{ \frac{p(y|\Theta)p(x|y, \Theta)}{p(x|y, \Theta)} \frac{q(x)}{q(x)} \right\} dx = \\ &= \int q(x) \ln \left\{ \frac{p(y, x|\Theta)}{q(x)} \right\} dx + \int -q(x) \ln \left\{ \frac{p(x|y, \Theta)}{q(x)} \right\} dx = \\ &= L(q, \Theta) + KL(q||p) \end{aligned} \quad (9)$$

式(9)中

$$\begin{cases} L(q, \Theta) = \int q(x) \ln \left\{ \frac{p(y, x|\Theta)}{q(x)} \right\} dx \\ KL(q||p) = \int -q(x) \ln \left\{ \frac{p(x|y, \Theta)}{q(x)} \right\} dx \end{cases} \quad (10)$$

式(10)中: $L(q, \Theta)$ 是关于模型参数 Θ 和隐变量先验分布 $p(x)$ 的函数; $KL(q||p)$ 为 $p(x)$ 和 $p(x|y)$ 之间的KL距离,且 $KL(q||p) \geq 0$ 。EM算法通过迭代实现对模型参数 Θ 的最大似然估计 Θ_{ML} 的求解,迭代的每一步包含求期望(E步)和最大化(M步)两个步骤^[26],在E步中,假设 $\Theta = \Theta^{old}$,通过优化 $p(x)$ 实现最大化 $L(q, \Theta^{old})$ 的值。之后进入M步,固定 $p(x) = p(x|y, \Theta^{old})$,通过最大化 $L(q, \Theta)$ 来更新模型参数的取值从而获得 Θ^{new} 。经过E步和M步的不断迭代, $\ln p(y|\Theta)$ 取值不断增加,模型参数 Θ 逐渐逼近于最大似然估计 Θ_{ML} 。

2 改进卡尔曼滤波预测模型

KF是一种在噪声背景下针对线性高斯系统进行预测和滤波的经典算法。KF采用递归运算,只需要系统当前时刻的观测值和前一时刻的估计值即可实现对下一时刻观测值的预测,具有结构简单,实时性高且噪声鲁棒性好的优点。对于变压器油中溶解气体体积分数时间序列,利用KF建立系统状态方程和观测方程可以表示为^[27-29]

$$\begin{cases} \tilde{x}_t = A\tilde{x}_{t-1} + B\omega_{t-1} \\ x_t = H\tilde{x}_t + \kappa_t \end{cases} \quad (11)$$

式(11)中: $\tilde{x}_t$ 和 $\tilde{x}_{t-1}$ 分别为 t 时刻和 $t-1$ 时刻的气体体积分数估计值; A 为系统状态矩阵; B 为系统控制矩阵; ω_{t-1} 为系统状态噪声; x_t 和 κ_t 分别为 t 时刻的观测值和观测噪声; H 为系统观测矩阵。基于系统状态方程和观测方程, KF对下一时刻观测值进行预测的递归过程可以表示为

$$\begin{cases} P_t = A\tilde{P}_{t-1}A^T + BQ_tB^T \\ K_t = P_t H^T (H_t P_t H^T + O_t)^{-1} \\ \tilde{P}_t = (1 - K_t H) P_t \\ \tilde{x}_{t+1} = \tilde{x}_{t-1} + K_t (x_t - H\tilde{x}_{t-1}) \end{cases} \quad (12)$$

式(12)中: Q_t 和 O_t 分别为 ω_t 和 κ_t 的协方差矩阵; P_t 和 K_t 分别为 t 时刻的预测误差协方差矩阵和 KF 增益; $\tilde{x}_{t-1}$ 为 $t-1$ 误差协方差矩阵的估计值。KF 通过式(12)的不断迭代, 逐次实现对未来时刻气体体积分数的预测。

KF 是一种针对线性系统构建的预测模型, 而变压器油中溶解气体体积分数是一种典型的非线性系统, 这种情况下 KF 的预测结果会发散, 预测性能出现下降。

BP 神经网络是当前应用最广的一种人工神经网络模型, 具备自适应和自学习能力, 适合于对非线性复杂系统进行建模^[30-31], 因此文中将 BP 神经网络与 KF 预测模型相结合, 将 KF 滤波结果的滤波增益 K_t 和信息 $\delta = x_t - x_{t-1}$ 作为 BP 神经网络的输入, 构建典型的三层网络模型, 输出为 BP 神经网络修正后的 KF 预测误差。最后利用修正后的 KF 预测误差对 KF 预测结果进行修正从而获得最终的预测结果, 同时将预测误差反馈至 KF 递归过程以避免后续 KF 滤波的误差累积。BP-KF 预测模型的原理框图见图 1。

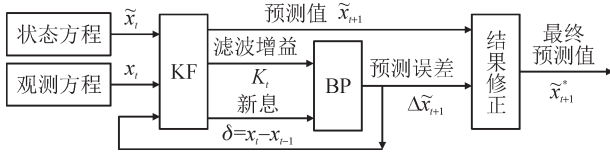


图1 BP-KF 预测模型原理框图

Fig. 1 Principle block diagram of BP-KF prediction model

3 算法流程

根据文中提出的缺失数据填补模型和 BP-KF 预测模型, 建立不完整数据条件下变压器油中溶解气体体积分数预测算法流程见图 2。算法输入为观测到的不完整气体体积分数字列, 即式(1)模型中的 y 。首先需要对观测序列的样本缺失情况进行分析, 具体规则为: 若观测数据 y 中第 t 个采样点取值的绝对值小于 10^{-6} , 则将该数据判决为缺失数据, 对应的位置坐标设置 0, 反之若观测数据 y 中第 t 个采样点取值的绝对值大于 10^{-6} , 则将该数据判决为有效数据, 对应位置坐标设置为 1。完成缺失情况分析后, 根据位置坐标取值为 0 或 1 的情况建立观测矩阵 Φ , 进而完成式(1)模型的构建, 利用 EM 算法对

其进行求解从而获得缺失数据填补后的完整数据。

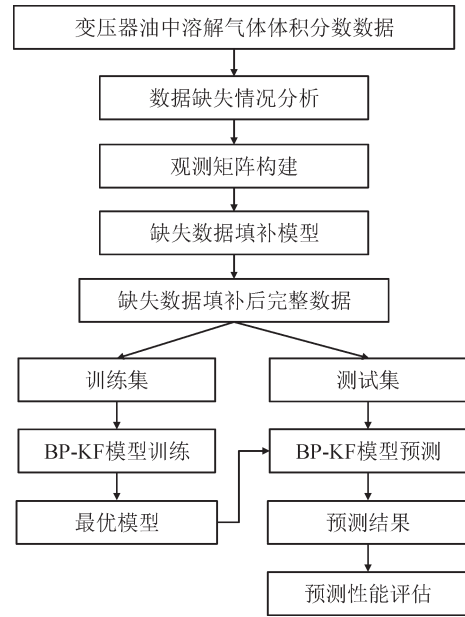


图2 所提方法流程图

Fig. 2 Flow chart of proposed method

然后将数据集划分为训练数据和测试数据, 利用训练数据对所提 BP-KF 模型进行训练, 得到最优预测模型, 并将其应用至测试数据对预测性能进行验证。

最后采用式(13)给出的预测结果均方根误差 (e_{RMSE}) 和平均绝对误差 (e_{MAPE}) 两项指标对预测精度进行定量评估, 并对不同方法的预测性能进行对比分析。

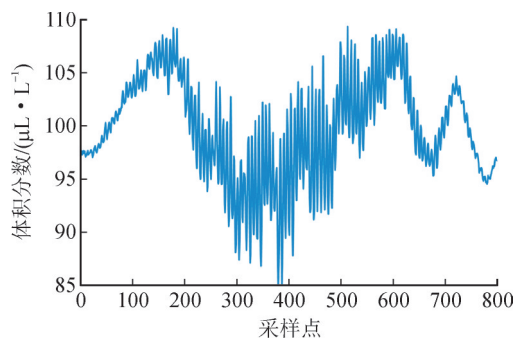
$$\begin{cases} e_{\text{RMSE}} = \sqrt{\frac{1}{M} \sum_{l=1}^M |x_l^* - x_l|^2} \\ e_{\text{MAPE}} = \frac{1}{M} \sum_{l=1}^M |x_l^* - x_l| \end{cases} \quad (13)$$

式(13)中: x_l 和 x_l^* 分别为实际值和预测值; M 为测试样本数。

4 算例分析

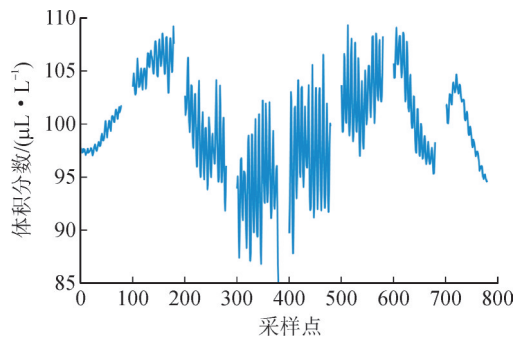
采用某地区一台 500 kV 变电站 4 号 B 相变压器 DGA 数据开展验证试验, 数据采集时间为 2021 年 7 月 31 日—2023 年 10 月 30 日, 采集到 H_2 、 CH_4 、 C_2H_2 、 CO_2 等 7 种气体的体积分数变化数据, 采样周期为 1 天, 每种气体均获得 800 个采样点。 H_2 体积分数变化曲线见图 3。由图 3 可以看出, 变压器中 H_2 体积分数呈现出了非常明显的波动性, 是一种典型的非线性非平稳随机过程。

采用 MATLAB R2020a 作为软件分析平台对所提缺失数据填补算法和 BP-KF 预测算法进行实现, 硬件平台为联想 ThinkPad X1, 处理器为 i7-7200u, 主频为 2.4 G, 操作系统为 64 位 Windows 10 系统。

图3 H₂体积分数变化序列Fig. 3 Concentration change sequence of H₂

4.1 缺失数据填补结果

为了验证所提方法的缺失数据填补性能,采用连续抽取方式构建数据缺失率为30%的样本缺失观测数据,见图4。

图4 缺失率为30%时H₂体积分数变化序列Fig. 4 H₂ concentration change sequence with a missing rate of 30%

连续抽取指的是对每100个采样点抽掉第71~100共30个点作为缺失数据。从图4可以看出,当存在数据缺失时,原始数据的时间连续性被破坏,不利于后续的体积分数预测。

为了建立缺失数据填补模型,首先需要评估原始数据的先验概率 $p(x)$ 是否满足高斯分布,对图3所示数据进行统计得到的归一化直方图见图5。图5中:红色曲线为根据图3所示数据的均值和方差计算得到的归一化高斯概率密度函数曲线。从图5可以看出,油中溶解气体体积分数数据的统计分布与高斯概率密度函数表现出了较强的一致性,表明所提缺失数据填补模型中对 $p(x)$ 先验分布的假设是合理的。

利用所提方法对图4所示数据进行缺失数据填补得到的结果见图6。其中图6(b)为图6(a)的局部放大图。为了对比,利用ARIMA和ERT两种方法进行缺失数据填补得到的结果分别见图7、8。从图7、8可以看出,由于气体体积分数是一种非线性时间序列,线性模型ARIMA的数据填补效果要明显差于ERT和所提方法两种非线性模型,而对比图8(b)和

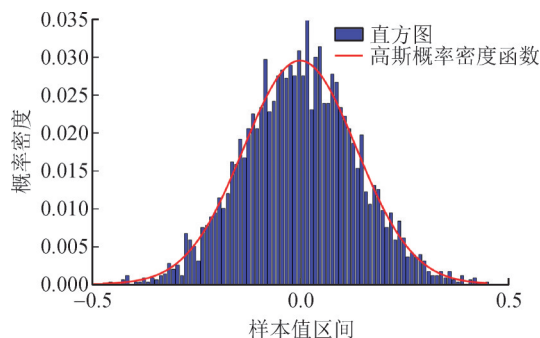
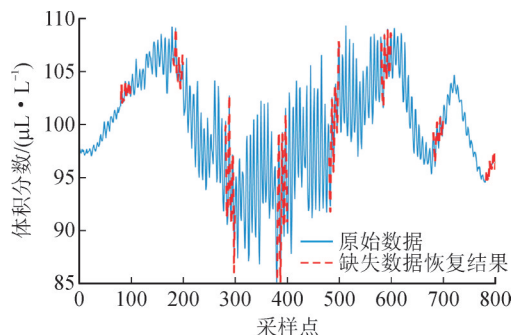


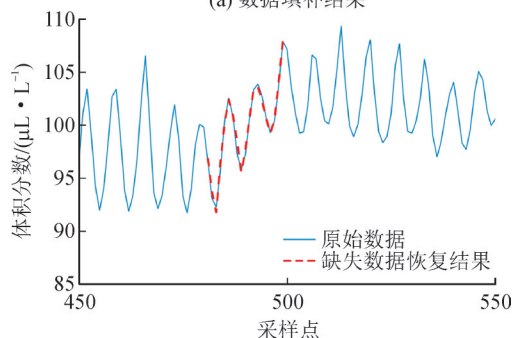
图5 原始数据统计分布

Fig. 5 Statistical distribution of raw data

图6(b)可以看出,ERT方法的填补误差要大于所提方法,这是由于ERT算法的每棵决策树在训练过程中只考虑了缺失数据与之前样本之间的相关信息,并未充分利用缺失数据与之后数据之间的相关信息,而所提方法在进行缺失数据填补过程中,充分利用了全部观测数据的信息,因此性能优于ERT方法。



(a) 数据填补结果



(b) 局部放大图

图6 所提方法的缺失数据填补结果

Fig. 6 The missing data filling results of the proposed method

所提方法、ARIMA、ERT 3种方法进行缺失数据填补后所得数据与完整数据之间的相关系数随缺失率的变化情况见图9。由图9可以看出,所提方法在不同缺失率情况下均能获得最优的数据填补性能。当缺失率为30%时,所提方法获得的相关系数为96.3%,ARIMA为84.9%,ERT为90.6%,填补性能

分别提升11.4%和5.7%。

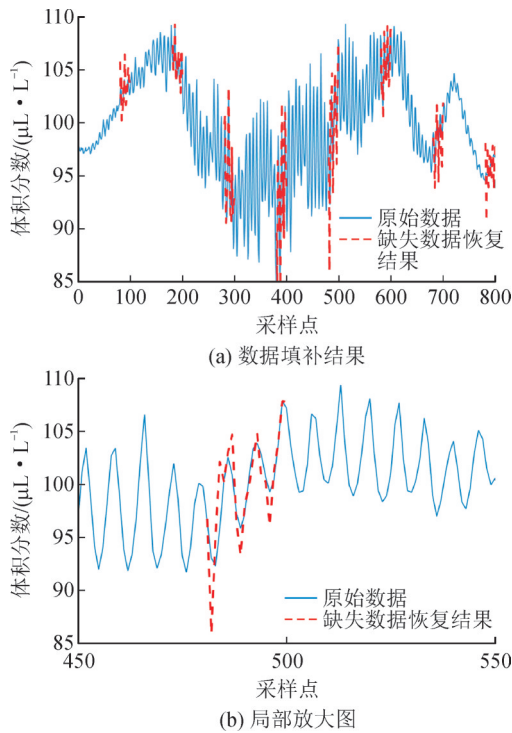


图7 ARIMA方法的缺失数据填补结果

Fig. 7 The missing data filling results of ARIMA

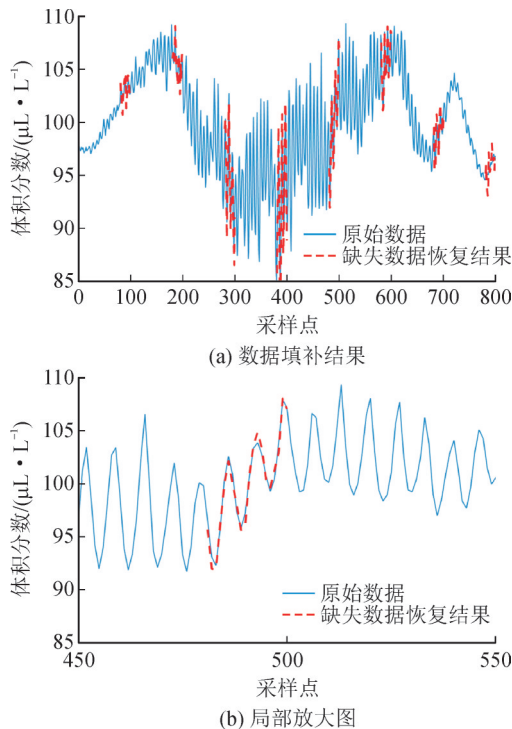


图8 ERT方法的缺失数据填补结果

Fig. 8 The missing data filling results of ERT

缺失数据填补的目的是为了更好地进行气体体积分数预测。为了进一步评估缺失数据填补对气体体积分数预测性能的影响,分别采用图3所示原始数据的前500个采样点,所提方法进行缺失数据填补后数据的前500个采样点,以及未进行缺失数据填补的

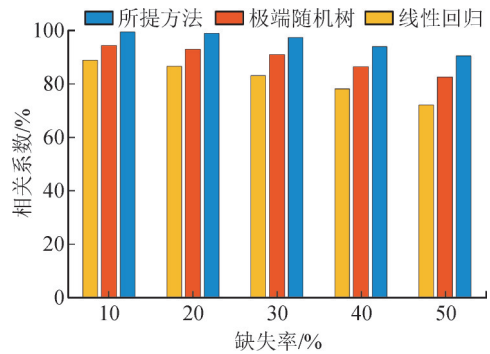


Fig. 9 Comparison of data filling performance under different missing rates

前500个采样点作为训练数据,对后300个采样点进行气体体积分数预测,得到的预测误差见图10。

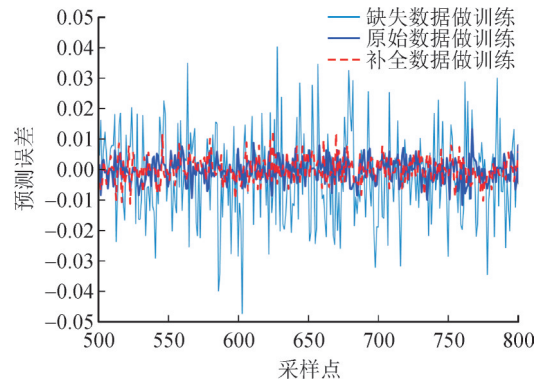


Fig. 10 The impact of missing data filling on predictive performance

从图10可以看出,不进行缺失数据填补直接进行模型训练得到的预测误差最大,而所提方法进行缺失数据填补后进行训练得到的预测误差与利用完整数据进行训练得到的预测误差较为接近,进一步验证了所提缺失数据补全方法的性能。

4.2 预测结果及对比分析

根据图2所示算法流程,完成缺失数据填补后,将数据前500个采样点作为训练数据,实现对所提BP-KF模型训练,获得最优预测模型,将剩余300个点作为测试数据,对最优预测模型的预测性能进行验证。

利用所提BP-KF方法、传统KF方法、LSTM方法、LSTM-CNN组合方法4种不同方法得到的预测结果见图11。4种方法预测结果的 e_{RMSE} 和 e_{MAPE} 指标见表1。从图11可以看出,传统KF方法在初期的预测精度较高,而随着预测时间的增长,KF预测结果逐渐发散,预测误差随时间明显变大,这是由于短时间内气体体积分数可以看作是线性的,而随着时间增长,气体体积分数的非线性特征逐渐明显;LSTM方法得到的预测精度较低,这是由于气体体

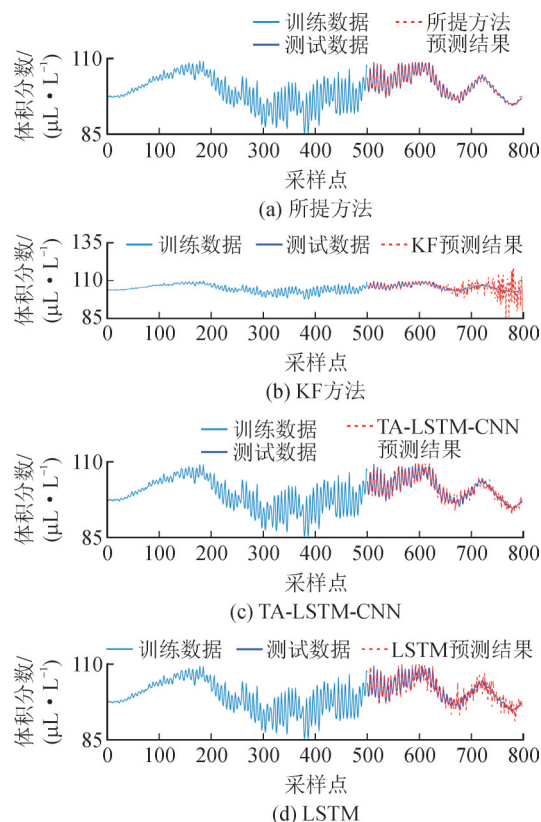


图11 不同方法预测结果

Fig. 11 Prediction results of different method

表1 不同方法预测结果

Table 1 Prediction results of different methods

预测方法	e_{RMSE}	e_{MAPE}
KF	0.265	0.178
LSTM	0.213	0.141
LSTM-CNN	0.202	0.138
BP-KF	0.157	0.103

积分数序列波动较大且波动频率较高,单一LSTM难以充分挖掘序列之间的相关性信息;LSTM-CNN的预测精度略优于LSTM单一模型,但是在部分采样点处的预测结果偏离真实值较大;所提方法由于融合了KF对线性系统的预测能力和BP对非线性系统的预测能力,因此获得了最优的预测性能,且对于300个测试样本点的预测精度较为接近,表现出了较高的稳定性。

表1所示结果与图11一致,所提方法的RMSE和MAPE两项指标均优于3种对比方法,相对于传统KF分别提升68.8%和72.5%,相对于LSTM分别提升35.6%和37.2%,相对于LSTM-CNN分别提升28.5%和33.6%,充分验证所提方法的有效性和优越性。

4.3 对不同气体体积分数预测的泛化能力对比分析

为了进一步评估不同方法的泛化能力,在相同条件下分别采用4种方法对CH₄、C₂H₂、C₂H₄、C₂H₆、CO、CO₂6种气体体积分数进行预测,对应的预测结

果见表2。从表2可以看出,CH₄、C₂H₂、C₂H₆、CO₂4种气体体积分数预测结果与表1较为一致,所提方法的预测精度最高,LSTM-CNN预测性能优于LSTM,KF方法的预测性能最差。而对于C₂H₄和CO两种气体体积分数预测结果,LSTM方法的预测精度略优于LSTM-CNN,表明LSTM-CNN方法的泛化能力不强,所提方法仍然获得了最优的预测结果。上述结果表明,所提方法相对于3种对比方法表现出了更强的泛化能力。

表2 不同方法对其他气体的预测结果

Table 2 Prediction results of different models for other gases

气体类型	预测方法	e_{RMSE}	e_{MAPE}
CH ₄	KF	0.379	0.280
	LSTM	0.294	0.182
	LSTM-CNN	0.257	0.169
	BP-KF	0.165	0.121
C ₂ H ₂	KF	0.259	0.590
	LSTM	0.191	0.125
	LSTM-CNN	0.176	0.107
	BP-KF	0.123	0.087
C ₂ H ₄	KF	0.294	0.356
	LSTM	0.202	0.155
	LSTM-CNN	0.238	0.178
	BP-KF	0.135	0.104
C ₂ H ₆	KF	0.535	0.603
	LSTM	0.343	0.421
	LSTM-CNN	0.306	0.387
	BP-KF	0.213	0.189
CO	KF	0.336	0.394
	LSTM	0.163	0.210
	LSTM-CNN	0.167	0.218
	BP-KF	0.115	0.127
CO ₂	KF	0.247	0.255
	LSTM	0.126	0.116
	LSTM-CNN	0.109	0.098
	BP-KF	0.093	0.075

5 结语

针对存在缺失数据情况下的变压器油中溶解气体体积分数预测问题开展研究,首先提出一种基于贝叶斯理论和高斯先验假设的缺失数据填补模型,将数据补全问题转化为求解模型参数的最大似然估计问题,并利用EM算法实现模型求解。构建不同数据缺失率试验,结果表明,在数据缺失率达到30%情况下,所提方法的填补结果与原始数据之间的相关系数仍能达到96.3%,相对于ARIMA和ERT分别提升11.4%和5.7%。之后针对气体体积分

数预测问题,以缺失数据补全后的数据作为输入,建立BP-KF预测模型,融合了KF对线性系统的预测能力和BP对非线性系统的预测能力,从而提升预测性能。基于实际数据的试验结果表明,BP-KF预测结果相对于传统KF,LSTM和LSTM-CNN的均方根误差(e_{RMSE})和平均绝对误差(e_{MAPE})分别降低28.5%和33.6%,并且具有更强的泛化能力。所提方法能够有效降低实际应用过程中对试验数据完整性的要求,能够为相关从业人员进行变压器故障诊断和检修决策提供支撑。

参考文献:

- [1] 邹阳,俞豪奕,金涛.融合模糊K近邻及证据理论的变压器油纸绝缘状态评估方法[J].电力系统保护与控制,2023,51(14):55-63.
Zou Yang, Yu Haoyi, Jin Tao. Evaluation method of the oil-paper insulation condition of a transformer based on fuzzy K nearest neighbor and evidence theory[J]. Power System Protection and Control, 2023, 51(14): 55-63.
- [2] 林智勇,苏若杭,黄国泰,等.基于等值老化容抗的变压器油纸绝缘状态评估[J].电力系统保护与控制,2025,53(3):140-147.
Lin Zhiyong, Su Ruohang, Huang Guotai, et al. Evaluation of the oil-paper insulation state of a transformer based on equivalent aging capacitance[J]. Power System Protection and Control, 2025, 53(3): 140-147.
- [3] 奚广宇,徐建军,朱景鹏.基于容性功率转移原理的高压大容量直流变压器设计[J].浙江电力,2025,44(11):125-136.
Xi Guangyu, Xu Jianjun, Zhu Jingpeng. Design of high voltage and large capacity DC transformers based on capacitive power transfer principle[J]. Zhejiang Electric Power, 2025, 44(11): 125-136.
- [4] 段舒帆,成立,罗豪,等.基于虚实融合的变压器油中气体输运过程数字孪生方法研究[J].智慧电力,2025,53(6):68-77.
Duan Shufan, Cheng Li, Luo Hao, et al. Research on a digital twin method for gas transport process in transformer oil based on cyber-physical integration[J]. Smart Power, 2025, 53(6): 68-77.
- [5] Wang Lin, Littler T, Liu Xueqin. Gaussian process multi-class classification for transformer fault diagnosis using dissolved gas analysis[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2021, 28(5): 1703-1712.
- [6] Lin Jun, Sheng Gehao, Yan Yang, et al. Prediction of dissolved gas concentrations in transformer oil based on the KPCA-FFOA-LSTM model[J]. Energies, 2018, 11(1): 225.
- [7] 汤建,侯慧娟,陈洪岗,等.基于BI-GRU改进的Seq2Seq网络的变压器溶解气体体积分数预测方法[J].电力自动化设备,2022,42(3):196-202.
Tang Jian, Hou Huijuan, Chen Honggang, et al. Concentration prediction method based on Seq2Seq network improved by BI-GRU for dissolved gas in transformer oil[J]. Electric Power Automation Equipment, 2022, 42(3): 196-202.
- [8] 徐鹏,吴鹏,甘强,等.500 kV变压器纯瓷套管漏油事故调查及防范措施[J].电磁避雷器,2022,3(307):186-190.
Xu Peng, Wu Peng, Gan Qiang, et al. Investigation on oil leakage accident of pure porcelain bushing of 500 kV transformer and preventive measures[J]. Insulators and Surge Arresters, 2022, 3(307): 186-190.
- [9] 王杰,李永鑫,张军亮,等.基于SBOA-TVFEMD的变压器油中溶解气体浓度预测[J].变压器,2025,62(5):23-31.
Wang Jie, Li Yongxin, Zhang Junliang, et al. Prediction of dissolved gas concentration in transformer oil based on SBOA-TVFEMD[J]. Transformer, 2025, 62(5): 23-31.
- [10] 黄丰,郭淳,邱冰兵,等.基于电磁-结构场的油浸式变压器振动特性及信号反演研究[J].电力科学与技术学报,2025,40(2):276-286.
Huang Feng, Guo Chun, Qiu Bingbing, et al. Vibration characteristics and signal inversion of oil-immersed transformer based on electromagnetic-structural field[J]. Journal of Electric Power Science and Technology, 2025, 40(2): 276-286.
- [11] 年泓昌.变压器绝缘油中含气量超标分析处理与防范措施[J].变压器,2024,61(8):73-75.
Nian Hongchang. Analysis treatment and preventive measures for excessive gas content in transformer insulation oil[J]. Transformer, 2024, 61(8): 73-75.
- [12] 陈铁,陈一夫,李咸善,等.基于SDS-SSA-LSTM的变压器油中溶解气体体积分数预测[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.
- [13] Jiang Jun, Chen Ruyi, Zhang Chaohai, et al. Dynamic fault prediction of power transformers based on lasso regression and change point detection by dissolved gas analysis[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2020, 27(6): 2130-2137.
- [14] 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.
- [15] 张鹏,齐波,李成榕,等.电力变压器油中溶解气体特性影响因素的量化分析[J].中国电机工程学报,2021,41(10):3620-3631.
Zhang Peng, Qi Bo, Li Chengrong, et al. Quantitative analysis of influence factors of dissolved gas characteristics in power transformer oil[J]. Proceedings of the CSEE, 2021, 41(10): 3620-3631.
- [16] 安国庆,史哲文,马世峰,等.基于RF特征优选的WOA-SVM变压器故障诊断[J].高压电器,2022,58(2):171-178.
An Guoqing, Shi Zhewen, Ma Shifeng, et al. Fault diagnosis of WOA-SVM transformer based on RF feature optimization[J]. High Voltage Apparatus, 2022, 58(2): 171-178.
- [17] 陈铁,陈卫东,李咸善,等.基于油中溶解气体分析的变压器故障预测[J].电子测量技术,2021,44(22):25-31.
Chen Tie, Chen Weidong, Li Xianshan, et al. Transformer fault prediction based on analysis of dissolved gas in oil[J]. Electronic Measurement Technology, 2021, 44(22): 25-31.
- [18] Zeng Bo, Guo Jun, Zhang Fan, et al. Prediction model for dissolved gas concentration in transformer oil based on modified grey wolf

- optimizer and LSSVM with grey relational analysis and empirical mode decomposition[J]. *Energies*, 2020, 13(2):422.
- [19] 栗磊, 王廷涛, 殷浩然, 等. 基于GWO-LSTM与NKDE的变压器油中溶解气体体积分数点-区间联合预测方法[J]. *高压电器*, 2022, 58(11):88-97.
- Su Lei, Wang Tingtao, Yin Haoran, et al. Point - interval joint prediction method of dissolved gas volume fraction in transformer oil based on GWO-LSTM and NKDE[J]. *High Voltage Apparatus*, 2022, 58(11):88-97.
- [20] 陈铁, 张治藩, 李咸善, 等. 基于混合模态分解和LSTM-CNN的变压器油中溶解气体体积分数预测[J]. *中国电力*, 2023, 56(1):132-141.
- Chen Tie, Zhang Zhifan, Li Xianshan, et al. Prediction of dissolved gas concentration in transformer oil based on hybrid mode decomposition and LSTM-CNN[J]. *Electric Power*, 2023, 56(1):132-141.
- [21] 周远翔, 林孟龙, 陈健宁, 等. 基于自注意力生成对抗网络的电力设备在线监测缺失数据填补[J]. *高电压技术*, 2023, 49(5):1795-1809.
- Zhou Yuanxiang, Lin Menglong, Chen Jianning, et al. Missing data imputation for online monitoring of power equipment based on self-attention generative adversarial networks[J]. *High Voltage Engineering*, 2023, 49(5):1795-1809.
- [22] 严英杰, 盛戈峰, 陈玉峰, 等. 基于时间序列分析的输变电设备状态大数据清洗方法[J]. *电力系统自动化*, 2015, 39(7):138-144.
- Yan Yingjie, Sheng Gehao, Chen Yufeng, et al. Cleaning method for big data of power transmission and transformation equipment state based on time sequence analysis[J]. *Automation of Electric Power Systems*, 2015, 39(7):138-144.
- [23] 廖才波, 杨金鑫, 邱志斌, 等. 基于缺失数据填补的油浸式变压器故障诊断[J]. *高电压技术*, 2024, 50(9):4091-4100.
- Liao Caibo, Yang Jinxin, Qiu Zhibin, et al. Fault diagnosis for oil-immersed transformer based on missing data imputation[J]. *High Voltage Engineering*, 2024, 50(9):4091-4100.
- [24] 程道, 李平, 郭凌旭, 等. 基于量测数据贝叶斯概率矩阵分解的变压器运行状态监测方法[J]. *电力系统及其自动化学报*, 2022, 34(1):100-107.
- Cheng Xiao, Li Ping, Guo Lingxu, et al. Transformer operation status monitoring method based on Bayesian probabilistic matrix factorization of measured data[J]. *Proceedings of the CSU-EPSCA*, 2022, 34(1):100-107.
- [25] Jolliffe I T. Principal component analysis[M]. New York: Springer, 1986.
- [26] 肖怀硕, 李清泉, 施亚林, 等. 灰色理论—变分模态分解和NSGA-II优化的支持向量机在变压器油中气体预测中的应用[J]. *中国电机工程学报*, 2017, 37(12):3643-3653.
- Xiao Huaishuo, Li Qingquan, Shi Yalin, et al. Prediction of dissolved gases in oil for transformer based on grey theory - variational mode decomposition and support vector machine improved by NSGA-II [J]. *Proceedings of the CSEE*, 2017, 37(12):3643-3653.
- [27] Nyamundanda G, Brennan L, Gormley I C. Probabilistic principal component analysis for metabolomic data[J]. *BMC Bioinformatics*, 2010(11):571-581.
- [28] 黄靖, 陈斌艺, 黄晓生, 等. 基于卡尔曼滤波便捷式接地线接地电阻检测装置[J]. *电磁避雷器*, 2023(1):65-73.
- Huang Jing, Chen Binyi, Huang Xiaosheng, et al. Portable grounding resistance detection device of grounding wire based on kalman filter[J]. *Insulators and Surge Arresters*, 2023(1):65-73.
- [29] 张佳伟, 张江南, 吴坡, 等. 基于融合容积卡尔曼滤波的直流微电网虚假数据检测[J]. *智慧电力*, 2024, 52(3):87-93.
- Zhang Jiawei, Zhang Jiangnan, Wu Po, et al. False data detection in DC microgrids based on fusion of cubature kalman filters[J]. *Smart Power*, 2024, 52(3):87-93.
- [30] Wang Zheng, Wang Bo, Liu Chun, et al. Improved BP neural network algorithm to wind power forecast[J]. *The Journal of Engineering*, 2017(13):940-943.
- [31] 左宏基, 陈宏军, 徐峰亮, 等. 基于BP神经网络建模的并联型有源电力滤波器谐波综合治理控制算法[J]. *电力电容器与无功补偿*, 2022, 43(1):50-56.
- Zuo Hongji, Chen Hongjun, Xu Fengliang, et al. Integrated control algorithm on harmonics of shunt active power filter based on BP neural network modeling[J]. *Power Capacitor & Reactive Power Compensation*, 2022, 43(1):50-56.
- 王振国(1973—), 男, 高级工程师, 研究方向为变电设备在线监测与故障诊断(E-mail:554355184@qq.com)。
- 宁飞(1989—), 男, 高级工程师, 研究方向为高电压绝缘(通信作者)(E-mail:lzlq2602@163.com)。